UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE

Marjan Cugmas

Vpliv lokalnih mehanizmov na razvoj bločnih modelov
Local mechanisms affecting the evolution of blockmodels

Doktorska disertacija

Ljubljana, 2019

UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE

Marjan Cugmas

Mentor: izr. prof. dr. Aleš Žiberna
Somentorica: prof. dr. Anuška Ferligoj

Vpliv lokalnih mehanizmov na razvoj bločnih modelov
Local mechanisms affecting the evolution of blockmodels

Doktorska disertacija

Ljubljana, 2019

Tone Pavček
Popotnik

Ko hodiš,
pojdi zmeraj do konca.

Spomladi do rožne cvetice,
poleti do zrele pšenice,
jeseni do polne police,
pozimi do snežne kraljice,
v knjigi do zadnje vrstice,
v življenju do prave resnice,
v sebi do rdečice čez eno in drugo lice.

A če ne prideš ne prvič, ne drugič
do krova in pravega kova
poskusi: vnovič in zopet in znova.

Zahvala Posebno zahvalo namenjam mentorjema izr. prof. dr. Alešu Žiberni in prof. dr. Anuški Ferligoj, ki sta me pri nastajanju doktorske naloge navdihovala ne le z neprecenljivimi nasveti in spodbudami, pač pa tudi z zgledom. ● I would like to thank to the members of the commission, Assoc. Prof. Anja Žnidaršič and Prof. Patrick Doreian, for very insightful and constructive comments, suggestions and questions that helped me to take a look from the other perspective. ● Hvala tudi Luki, ki me je zvlekel v vse to ter Sari in Špeli za navdihujoče razprave o znanosti in tegobah doktorskih študentov. ● Hvala Anji za prijazne nasvete, potrpežljivost in pomoč pri urejanju službenih poti in naročilnic. ● Posebno zahvalo dolgujem svoji družini in svojim najbližjim, ki so mi stali ob strani ter rasli z mano tudi v trenutkih, ki niso bili lahki niti zame, niti zanje. ● Hvala tudi vsem ostalim, ki niso omenjeni zgoraj, a so pomembno prispevali k nastanku disertacije. ● Hvala davkoplačevalcem, ki so financirali moje izobraževanje ter raziskavo.

Univerza
v Ljubljani
Fakulteta
za družbene vede

Kardeljeva ploščad 5
1000 Ljubljana, Slovenija
telefon 01 58 05 120

IZJAVA O AVTORSTVU
doktorske disertacije

Podpisani/-a Marjan Cugmas, z vpisno številko 74150901, sem avtor/-ica doktorske disertacije z naslovom: Vpliv lokalnih mehanizmov na razvoj bločnih modelov.

S svojim podpisom zagotavljam, da:

- je predložena doktorska disertacija izključno rezultat mojega lastnega raziskovalnega dela;
- sem poskrbel/-a, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem delu, navedena oz. citirana v skladu s fakultetnimi navodili;
- sem poskrbel/-a, da so vsa dela in mnenja drugih avtorjev oz. avtoric navedena v seznamu virov, ki je sestavni element predloženega dela in je zapisan v skladu s fakultetnimi navodili;
- sem pridobil/-a vsa dovoljenja za uporabo avtorskih del, ki so v celoti prenesena v predloženo delo in sem to tudi jasno zapisal/-a v predloženem delu;
- se zavedam, da je plagiatorstvo – predstavljanje tujih del, bodisi v obliki citata bodisi v obliki skoraj dobesednega parafraziranja bodisi v grafični obliki, s katerim so tuje misli oz. ideje predstavljene kot moje lastne – kaznivo po zakonu (Zakon o avtorski in sorodnih pravicah (UL RS, št. 16/07-UPB3, 68/08, 85/10 Skl.US: U-I-191/09-7, Up-916/09-16)), prekršek pa podleže tudi ukrepom Fakultete za družbene vede v skladu z njenimi pravili;
- se zavedam posledic, ki jih dokazano plagiatorstvo lahko predstavlja za predloženo delo in za moj status na Fakulteti za družbene vede;
- je elektronska oblika identična s tiskano obliko doktorske disertacije ter soglašam z objavo doktorske disertacije v zbirki »Dela FDV«.

V Ljubljani, dne 22. 10. 2019 Podpis avtorja/-ice: _____

Local mechanisms affecting the evolution of blockmodels

Social scientists often seek to understand the relationship between micro social mechanisms and macro social output. In the context of social networks, different micro social mechanisms are usually operationalized by local network mechanisms, while macro social outputs are operationalized by global network structures (Stadtfeld, 2018). Therefore, the aim of this dissertation is to study the relationship between local network mechanisms and global network structures. Not only is the emergence of the selected global network structures addressed, but so too is the transition from one global network structure to another.

Moreno was one of the earliest social network scientists to study global structures of observed networks (Moreno, 1934). By considering the nodes' attributes and using structured interviews, he explained the social mechanisms responsible for the emergence of the observed global network structures. Later, researchers (Cartwright & Harary, 1956; Davis, 1967; Davis & Leinhardt, 1967; Heider, 1946; Johnsen, 1985) proposed several models for global network structures. Perhaps one of the most popular is the balance model (Cartwright & Harary, 1956), which consists of two clusters of nodes which are internally linked with positive ties and the nodes from different clusters are not linked or they are linked by negative ties. By considering the appearance of different triad types, Davis & Leinhardt (1967) proposed an approach for testing the existence of a selected global network structure in an empirical network.

In this study, a blockmodel is used to define a global network structure. A blockmodel is defined as a network in which the nodes represent clusters of equivalent nodes (according to the structure of their links) from the studied network, while the links in a blockmodel represent the relationships between and within the clusters. The term "block" refers to a submatrix in an adjacency matrix that shows the relationships between nodes from two different clusters or between nodes from the same cluster (Doreian, Batagelj, & Ferligoj, 2005). In this way, a blockmodel can be a very exact representation of a chosen type of global network structure and is used widely across many scientific fields. The blockmodel types considered in this dissertation are the most commonly studied blockmodel types: cohesive blockmodel, (symmetric and asymmetric) core-periphery blockmodel, hierarchical blockmodel, hierarchical-cohesive blockmodel, transitivity blockmodel and transitive-cohesive blockmodel.

Moreover, the social (network) mechanisms can be defined in different ways. Common to the various definitions of social (network) mechanisms is the claim that social mechanisms hold a very important explanatory role (Hedström & Swedberg, 1998). Hedström & Swedberg (1998, p. 7) summarized Schelling (1998) when saying that a "social mechanism can be seen as a systematic set of statements that provide a plausible account on how input and output are linked to one another". Therefore, social mechanisms may be seen as "models of interaction among individuals that generate the particular social structures" (Gambetta, 1998, p. 102). According to Hedström (2005, p. 25), a social mechanism "describes a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome". In the context of social networks, the (macro) social outcome can be operationalized by the global network structure, while the local network mechanisms can be operationalized in different ways, according to their type. Stadtfeld (2018) and Hedström & Swedberg (1998) defined, based on Coleman's macro-micro-macro model (Coleman (1986)), three types of mechanisms: situational mechanisms (related to the global network structure's impact on, e.g. the beliefs, desires and opportunities of

an individual, action-formation mechanisms (associated with the impact of individuals' beliefs, desires and opportunities on their actions/behaviour) and transformational mechanisms (related to the impact of individuals' actions on the global network structure). In this study, the main focus is given to the last two types of local network mechanisms. When Stochastic Actor Oriented Models (SAOM) are used, these types of local network mechanisms are defined through the selected local network statistics, whereas in the case of algorithms from the family of the Network Evolution Models (NEM) (Toivonen et al., 2009) the local network mechanisms are often defined through a set of "if-then" rules.

Even though many studies use the blockmodeling approach to describe the global network structures of the networks observed and many studies rely on Exponential Random Graph Models (ERGM) or SAOM to explain the dynamics and underling mechanisms of the network dynamics, there is no systematic study focusing on the relationship between selected local network mechanisms and selected global network structures, as operationalized by blockmodels. Providing the framework for studying this phenomenon is one aim of this dissertation.

The dissertation consists of two parts. The ability to generate networks with the selected blockmodel types, by considering only the triad types, is addressed in the first part. This research question is especially important because (although the correlation between the number of different triad types and the presence of a given global network structure is well known and generally used) there is no known systematic study addressing a relationship between different triad types and blockmodels as the operationalization of global network structures. Whether the selected blockmodel types can be generated by considering only the triad types without any nodes' attributes shows that these blockmodels can emerge as a consequence of local network mechanisms such as popularity, assortativity, different transitivity-related mechanisms and others.

To study the mentioned research question, different triad types are classified in the set of allowed and the set of forbidden triad types for each blockmodel type that is considered. A given triad type is called 'allowed' if its frequency in a given ideal blockmodel (without any inconsistency) is higher than 0; otherwise, it is called 'forbidden'. Based on the A-measure (defined as a ratio between the number of triads of the selected type in a network with a given blockmodel and the expected number of triads in a random network with the same density), the sets of allowed and forbidden triad types are further reduced to a set of selected allowed triad types and to a set of selected forbidden triad types.

Two algorithms are used to generate the networks. The first is the proposed Relocating Links algorithm (RL algorithm) while the second is the Markov Chain Monte Carlo algorithm (MCMC algorithm) as implemented in the "ergm" package (Hunter, Handcock, Butts, Goodreau, & Morris, 2008) for the R programming language (Team, 2000). The two different algorithms are used to reduce the possibility of being unable to generate networks with a given blockmodel type due to the characteristics of the algorithm. The RL algorithm is more deterministic than the MCMC algorithm and the RL algorithm requires the exact distribution of triad types for the selected blockmodel whereas the coefficients in MCMC algorithms are arbitrarily set to 2 (for allowed triad types) and -2 (for forbidden triad types).

Several networks are generated by considering each set of triads and each blockmodel type. The level of inconsistencies is evaluated by the proposed Mean Improvement Value (MIV), which

allows the fit of a selected (ideal) blockmodel type to be compared against blockmodels of different generated networks.

In general, most studied blockmodels can be generated by only considering different triad types. This shows that some global network structures can emerge by virtue of the local network mechanisms that does not include the nodes' attributes. Blockmodels generated by considering only the list of forbidden triad types do not have a much higher amount of inconsistencies than networks generated by considering all triad types. This is important to note because the frequencies of the allowed triad types (which is taken into account when the RL algorithm is used to generate networks) contain information on the number of clusters, their size, and the size of the network. However, this is not the case with the forbidden triad types where a researcher must provide only the information regarding which triad types should not appear in the network rather than the frequency of each triad type.

While there is a small number of inconsistencies in the networks generated with most blockmodel types compared to the ideal blockmodel type, it is harder to generate networks with a hierarchical blockmodel. Considering some other local network structures, such as paths of length three, considerably reduces the number of inconsistencies in the blockmodels generated.

The second part of the dissertation considers local network mechanisms, instead of local network structures, in the context of different blockmodel types. While local network structures are represented by different types of subgraphs, the local network mechanisms are processes that drive the specific actions of the nodes in the network, as described above. Different local network mechanisms are operationalized using different network statistics, which are considered by the nodes, when they obtain an opportunity to change the status of their links. This is done by different proposed algorithms from the NEM family, which rely on the following logic: at each iteration of the NEM algorithm, a node is randomly selected. Then, by considering selected node i and all other nodes, different local network statistics are calculated by considering the selected local network mechanisms. These statistics are weighted to enable the different levels of importance of the selected local network mechanisms to be considered. Based on these weighted local network statistics, the selected node creates a link, dissolves or confirms an already existing link. The proposed NEM algorithms mainly differ with respect to how the symmetric links are considered, the way in which the duration of links is considered, and whether newcomers and outgoers are present.

In this study, the mechanisms' weights are randomly generated. For each set of weights, several networks are generated by using the proposed NEM algorithm while the global network structures are evaluated by the number of inconsistent blocks (Žnidaršič, Ferligoj, & Doreian, 2012, 2017, 2018) and the value of the proposed Relative Fit function (RF), which quantifies the level of inconsistencies in a generated blockmodel according to the ideal blockmodel type.

Given that there are many possible blockmodel types and possible local network mechanisms, the social context of the study is taken into account to select the most relevant blockmodel types and corresponding local network mechanisms. Two such social contexts considered in this dissertation are: (i) friendships and likings among pre-schoolers; and (ii) the flow of knowledge among employees of an international, knowledge-based company. Based on these two social contexts, two blockmodel types are proposed: an (symmetric and asymmetric) core-cohesive blockmodel, and a hierarchical-cohesive blockmodel with last non-cohesive group.

The first blockmodel type consists of at least three groups of nodes, where one group is called the core group and the other groups are called cohesive groups. In networks without inconsistencies, the nodes within all groups are internally all linked to each other. In both the symmetric and asymmetric case, all the nodes from cohesive groups are linked to all the nodes from the core group while only in the symmetric case the core nodes are also linked to the cohesive ones. The latter, a hierarchical-cohesive blockmodel with the last non-cohesive group, is similar to the well-known hierarchical-cohesive blockmodel where the clusters are hierarchically ordered and the nodes within all clusters are linked, but with the proposed blockmodel the nodes from the cluster on the lowest hierarchical level are not linked to each other. It is shown that the symmetric core-cohesive blockmodel type and the hierarchical-cohesive blockmodel with the last non-cohesive group are appropriate to be considered in the social context relating to a kindergarten and a company.

The results of the Monte Carlo simulations show that the symmetric and asymmetric core-cohesive blockmodel types can emerge due to the mutuality, popularity, assortativity (of in-degree) and transitivity-related local network mechanisms when the initial global network structure is an empty network, a network with a cohesive blockmodel, or a network with an asymmetric core-periphery blockmodel. Observations have revealed that some intermediate blockmodel types can emerge during the evolutionary process of generating the chosen blockmodel type.

It was also shown (based on empirical data collected within a larger longitudinal study in the USA and analysed by several researchers, e.g. Schaefer et al. (2010) and DeLay et al. (2016)) that the symmetric core-cohesive blockmodel type appears in interactional networks among pre-schoolers. The fact this blockmodel type can be generated by the selected local network mechanisms does not imply that the global network structures of the empirical networks emerged due to the studied local network mechanisms. However, the appearance of this blockmodel type in the empirical data raises some very important developmental questions, which should be answered by considering the nodes' attributes. Some of these questions include whether differences exist in some psychological and other types of characteristics (e.g. gender, level of extroversion) between children from the core group and those from the cohesive groups, and whether such a global network structure should be encouraged among pre-schoolers or not.

The same methodology was used to generate the other selected blockmodel types. By considering the selected mechanisms, cohesive, (symmetric and asymmetric) core-periphery and transitive blockmodels can also be generated, but not a hierarchical blockmodel, hierarchical-cohesive blockmodel or transitive-cohesive blockmodel.

A hierarchical-cohesive blockmodel with the last non-cohesive group can emerge as a result of so-called value-related mechanisms (i.e. hierarchical position of the alter, tenure of the alter, popularity level of the alter, the number of partners shared by the ego and the alter) and cost-related mechanisms (i.e. difference in hierarchical position between the ego and the alter, difference in tenure between the ego and the alter, distance between the ego and the alter, the number of partners shared by the ego and the alter). Value and cost are defined through the ego's perception of the costs of obtaining the alter's knowledge and the value of the knowledge so obtained (Nebus, 2006). This blockmodel type can also emerge when newcomers and outgoers are considered. The ability to generate the global network structure, with local network mechanisms that do not consider the nodes' attributes (except tenure), indicates that a company can develop policies that lead a knowledge flow towards the desired global structure (if it has one).

The most important contribution of this dissertation is the observation that the most common blockmodel types can be generated by the basic local network mechanisms, without taking the attributes of the nodes into account. However, it is necessary to consider the social context and corresponding constraints on the nodes' characteristics and their behaviour (Doreian & Conti, 2012) while analysing evolution of the global network in real networks.

Keywords: social network analysis, network evolution, local network mechanisms, global network structures, blockmodel, Exponential Random Graph Modelling, Stochastic Actor Oriented Models, Network Evolution Models, Relative Fit function, interactional networks, knowledge-flow networks, preschool environment, organizational environment.

Vpliv lokalnih mehanizmov na razvoj bločnih modelov

Raziskovalci s področja družboslovja želijo pogosto razumeti odnos med družbenimi mikromehanizmi in družbenim makroizidom. V okviru analize omrežij so različni družbeni mikromehanizmi navadno operacionalizirani z lokalnimi omrežnimi mehanizmi, družbeni izidi na makroravni pa so operacionalizirani z različnimi globalnimi zgradbami omrežja (Stadtfeld, 2018). Tako je namen pričujoče disertacije proučiti odnos med lokalnimi omrežnimi mehanizmi in globalnimi omrežnimi zgradbami. Poleg nastanka izbranih vrst globalnih omrežnih zgradb je naslovljen tudi prehod iz ene v drugo globalno zgradbo omrežja.

Moreno je eden izmed najzgodnejših raziskovalcev s področja analize omrežij, ki je proučeval globalno zgradbo empiričnih omrežij (Moreno, 1934). Z upoštevanjem lastnosti vozlišč in z uporabo strukturiranih intervjujev je pojasnil družbene mehanizme, ki so vplivali na nastanek opažene globalne zgradbe v empiričnih omrežjih. Poznejši raziskovalci (Cartwright in Harary, 1956; Davis, 1967; Davis in Leinhardt, 1967; Heider, 1946; Johnsen, 1985) so predlagali več modelov globalnih zgradb omrežij. Eden izmed najbolj znanih je verjetno ravnotežnostni model (Cartwright in Harary, 1956), ki je sestavljen iz dveh skupin vozlišč. Vozlišča znotraj skupin so dobro povezana, med vozlišči iz različnih skupin pa ni povezav ali pa obstajajo zgolj negativne povezave. Davis in Leinhardt (1967) sta predlagala pristop za preverjanje obstoja omenjene globalne omrežne zgradbe ter tudi nekaterih drugih omrežnih zgradb, ki temelji na upoštevanju različnih vrst triad.

V tej raziskavi so različne vrste globalnih zgradb omrežij opredeljene z različnimi vrstami bločnih modelov. Bločni model je definiran kot omrežje, v katerem so vozlišča skupine enakovrednih (glede na zgradbo povezav) vozlišč proučevanega omrežja, povezave pa so povezave med skupinami in znotraj skupin. Izraz »blok« se nanaša na matriko povezav, ki prikazuje povezave med vozlišči iz dveh različnih skupin ali med vozlišči znotraj ene skupine (Doreian in drugi, 2005). V tem pogledu so lahko bločni modeli zelo natančen opis izbranih vrst globalnih zgradb omrežij in so uporabljeni v različnih znanstvenih disciplinah. V pričujoči disertaciji so upoštevane najpogostejše vrste bločnih modelov: koheziven bločni model, (simetričen in asimetričen) središčno-periferen bločni model, hierarhičen bločni model, hierarhično-koheziven bločni model, tranzitiven bločni model ter tranzitivno-koheziven bločni model.

Tako kot različne globalne zgradbe omrežij je tudi družbene mehanizme mogoče opredeliti na različne načine. Skupna mnogim opredelitvam je trditev, da ima upoštevanje družbenih mehanizmov zelo pomembno vlogo pri pojasnjevanju družbenih pojavov (Hedström in Swedberg, 1998). Hedström in Swedberg (1998, str. 7) povzemata Schellinga (1998), ki navaja, da je mogoče družbene mehanizme razumeti kot urejeno množico trditev, ki pojasnjujejo povezavo med vhodom in izhodom. Tako je mogoče na družbene mehanizme gledati kot na modele interakcij med posamezniki, ki vplivajo na nastanek določenih družbenih zgradb (Gambetta, 1998, str. 102). Podobno Hedstöm (2005, str. 25) družbene mehanizme opisuje kot skupek entitet in aktivnosti, ki so urejene tako, da privedejo to določenega izida. V okviru družbenih omrežij je mogoče izid na makroravni operacionalizirati z različnimi globalnimi omrežnimi zgradbami, operacionalizacija lokalnih omrežnih mehanizmov pa je odvisna od vrste mehanizmov. Stadtfeld (2018) in Hedström in Swedberg (1998) so na podlagi Colemanovega makro-mikro-makro modela (Coleman, 1986) opredelili tri vrste mehanizmov: situacijske mehanizme (angl. *situational mechanisms*) (nanašajo se na vpliv globalne omrežne zgradbe na lastnosti posameznika, na primer na njegove želje,

prepričanja in možnosti), vedenjske mehanizme (angl. *action-formation mechanisms*) (nanašajo se na to, kako lastnosti posameznika vplivajo na njegovo vedenje) in pretvorbene mehanizme (angl. *transformational mechanisms*) (nanašajo se na to, kako vedenje posameznika vpliva na globalno omrežno zgradbo). V tej disertaciji so naslovljeni vedenjski ter pretvorbeni mehanizmi, ki jih je mogoče operacionalizirati prek množice pravil o vzpostavljanju, vzdrževanju in prekinjanju povezav. Verjetnostni modeli na ravni posameznika (angl. *Stochastic Actor-Oriented Models*; SAOM) tovrstne lokalne omrežne mehanizme opredeljujejo z izbranimi omrežnimi statistikami, modeli iz družine modelov razvoja omrežij (angl. *Network Evolution Models*; NEM) pa lokalne omrežne mehanizme pogosto opredeljujejo z množico pravil »če – potem«.

Čeprav mnogo raziskav naslavlja vprašanje globalne zgradbe omrežij (z uporabo bločnega modeliranja) ter dinamike v omrežjih z upoštevanjem pripadajočih lokalnih omrežnih mehanizmov (z uporabo SAOM ali eksponentih slučajnih grafov (angl. *Exponential Random Graph Models*; ERGM)), pa ni raziskave, ki bi sistematično naslovila odnos med izbranimi lokalnimi omrežnimi mehanizmi in med izbranimi globalnimi omrežnimi zgradbami, opredeljenimi z različnimi vrstami bločnih modelov. Eden izmed ciljev te disertacije je predstaviti okvir za proučevanje omenjenega odnosa.

Disertacija je urejena v dveh delih. Zmožnost generiranja omrežij z izbrano vrsto bločnega modela, z upoštevanjem različnih vrst tirad, je naslovljena v prvem delu disertacije. To raziskovalno vprašanje je pomembno zlasti zato, ker še ni raziskave, ki bi sistematično naslovila povezavo med različnimi vrstami triad in različnimi vrstami bločnih modelov, kljub znani in pogosto upoštevani povezavi med različnimi vrstami triad in globalnih omrežnih zgradb. Zmožnost generiranja omrežij z izbranimi vrstami bločnih modelov, brez upoštevanja lastnosti vozlišč, kaže, da lahko izbrane vrste bločnih modelov nastanejo kot rezultat izbranih lokalnih mehanizmov, kot so mehanizem popularnosti, mehanizem podobnosti stopenj, mehanizmi, povezani s tranzitivnostjo, in preostali.

Za namene prvega raziskovalnega vprašanja so različne vrste triad razvrščene v množico dovoljenih ali v množico prepovedanih vrst triad. Takšna klasifikacija je narejena za vsako obravnavano vrsto bločnih modelov. Dana vrsta triade je dovoljena, če je njena frekvenca v omrežju z izbranim bločnim modelom brez napak večja od 0, sicer pa je prepovedana. Na osnovi A-mere (definirane kot razmerje med številom izbrane vrste triad v empiričnem omrežju z danim bločnim modelom in pričakovanim številom triad v slučajnem omrežju z enako gostoto) so množice dovoljenih in prepovedanih vrst triad nadalje zmanjšane na množice izbranih dovoljenih vrst triad in množice izbranih prepovedanih vrst triad.

Dva različna algoritma sta uporabljena za generiranje omrežij z upoštevanjem prepovedanih in/ali dovoljenih vrst triad. Prvi je predlagan algoritem prestavljanja povezav (algoritem RL), drugi pa algoritem MCMC (angl. *Markov Chain Monte Carlo algorithm*), ki je implementiran v paketu »ergm« (Hunter in drugi, 2008) za programski jezik R (Team, 2000). Dva različna algoritma sta uporabljena z namenom zmanjšanja verjetnosti nezmožnosti generiranja omrežij z izbrano vrsto bločnega modela kot posledice lastnosti uporabljenega algoritma. V primerjavi z algoritmom MCMC je algoritem RL bolj determinističen, a (kot vhodni parameter) zahteva natančno porazdelitev upoštevanih vrst triad za izbrano vrsto bločnega modela, medtem ko so parametri algoritma MCMC poljubno nastavljeni na 2 (za dovoljene vrste triad) oziroma -2 (za prepovedane vrste triad).

Z upoštevanjem različnih vrst triad in različnih vrst bločnih modelov je generiranih mnogo omrežij. Raven napak v generiranih omrežjih je ocenjena s povprečno vrednostjo izboljšanja (angl. *Mean Improvement Value*; MIV), ki omogoča primerjavo prileganja globalne zgradbe različnih generiranih omrežij z različnimi vrstami bločnih modelov.

Z upoštevanjem različnih vrst triad je mogoče generirati večino analiziranih vrst bločnih modelov. To nakazuje, da se različne vrste bločnih modelov lahko pojavijo kot posledica lokalnih omrežnih mehanizmov, ki so neodvisni od lastnosti vozlišč. Nadalje, omrežja, ki so generirana z upoštevanjem množice prepovedanih vrst triad, ne vsebujejo bistveno višje ravni napak, kakor omrežja, ki so generirana z upoštevanjem vseh vrst triad. To je pomembna ugotovitev zato, ker frekvence dovoljenih vrst triad (ki so upoštevane pri uporabi algoritma RL) vsebujejo informacijo o številu skupin, velikosti skupin ter velikosti omrežja. To pa ne drži v primeru, ko so omrežja generirana z upoštevanjem prepovedanih vrst triad. V takem primeru je treba algoritmu RL posredovati zgolj informacijo o tem, katere vrste triad se ne smejo pojaviti v omrežju, ne pa tudi o frekvenci pojavljanja različnih vrst triad v omrežju z izbranim idealnim bločnim modelom (brez napak).

Medtem ko omrežja za večino analiziranih bločnih modelov vsebujejo relativno nizko raven napak, pa je težje generirati omrežja s hierarhičnim bločnim modelom. Upoštevanje nekaterih drugih lokalnih omrežnih zgradb, kot so poti dolžine tri, bistveno zmanjša raven napak v generiranih hierarhičnih bločnih modelih.

Drugi del disertacije obravnava lokalne omrežne mehanizme namesto lokalnih omrežnih zgradb, v okviru različnih vrst bločnih modelov. Medtem ko so lokalne omrežne zgradbe opredeljene z različnimi vrstami podomrežij, so lokalni omrežni mehanizmi procesi, ki vplivajo na konkretna dejanja vozlišč v omrežju, kot je to opisano v prejšnjih odstavkih. Različni lokalni omrežni mehanizmi so opredeljeni z različnimi omrežnimi statistikami, kar je upoštevano v algoritmih NEM, ki so definirani na naslednji način: v vsaki iteraciji algoritma je po naključju izbrano eno vozlišče. Nato so, upoštevajoč izbrano vozlišče in vsa druga vozlišča, izračunane različne omrežne statistike, opredeljene z izbranimi lokalnimi omrežnimi mehanizmi. Te statistike so utežene, kar omogoča upoštevanje različnih moči oziroma pomembnosti izbranih lokalnih omrežnih mehanizmov. Na osnovi uteženih omrežnih statistik izbrana enota vzpostavi novo povezavo ter prekine ali potrdi že obstoječo povezavo. Uporabljene različice algoritmov NEM se razlikujejo predvsem po načinu obravnave simetričnih povezav, trajanja povezav ter po tem, ali je upoštevan prihod novih vozlišč ter osip vozlišč.

V tej raziskavi so uteži lokalnih omrežnih mehanizmov generirane naključno. Za vsako množico uteži je generiranih več omrežij, globalne omrežne zgradbe pa so preverjane s številom neskladnih blokov (Žnidaršič in drugi, 2012, 2017, 2018) in vrednostjo predlagane funkcije relativnega prileganja, ki meri stopnjo napak v generiranih bločnih modelih, glede na izbrane vrste idealnih bločnih modelov.

Ker obstaja veliko vrst bločnih modelov in lokalnih omrežnih mehanizmov, je za izbiro lokalnih omrežnih mehanizmov in pripadajočih vrst bločnih modelov dobro upoštevati izbrane družbene kontekste. V tej raziskavi sta upoštevana naslednja družbena konteksta: (i) prijateljstva in naklonjenosti med predšolskimi otroki; ter (ii) pretok znanja med zaposlenimi v mednarodnem podjetju, ki temelji na znanju. Na osnovi navedenih družbenih kontekstov sta predlagani dve vrsti

bločnih modelov: (simetričen ter asimetričen) središčno-koheziven bločni model ter hierarhično-koheziven bločni model z nekohezivno zadnjo skupino.

Omrežje s prvim, središčno-kohezivnim bločnim modelom je sestavljeno iz vsaj treh skupin vozlišč, kjer je ena skupina vozlišč imenovana »središča skupina«, preostale skupine vozlišč pa so poimenovane kot »kohezivne skupine«. V bločnih modelih brez napak so vsa vozlišča znotraj skupin neposredno povezana med seboj. V omrežjih z eno ali drugo vrsto (simetrično ali asimetrično) središčno-kohezivnega bločnega modela so vozlišča iz kohezivnih skupin neposredno povezana z vozlišči iz središčnih skupin, a so samo v simetričnem središčno-kohezivnem bločnem modelu tudi vozlišča iz središčne skupine povezana z vozlišči iz kohezivnih skupin. Drugi, hierarhičen bločni model z nekohezivno zadnjo skupino, je podoben hierarhično-kohezivnemu bločnemu modelu, le da v primeru predlaganega bločnega modela vozlišča iz skupine na najnižji hierarhični ravni niso povezana med seboj. V disertaciji je pokazana smiselnost obravnave obeh vrst bločnih modelov v okviru družbenih kontekstov, ki so povezani z vrtci ali s podjetji.

Rezultati Monte Carlo simulacij kažejo, da lahko (simetričen in asimetričen) središčno-kohezivni bločni model nastane kot posledica mehanizmov vzajemnosti, popularnosti, podobnosti stopenj in mehanizmov, povezanih s tranzitivnostjo. To velja za vse obravnavane začetne zgradbe omrežij: prazno omrežje, omrežje s kohezivnim bločnim modelom ter omrežje z asimetričnim središčno-perifernim bločnim modelom. Analize generiranih omrežij kažejo, da lahko med razvojem globalne zgradbe omrežja nastanejo nekatere prehodne globalne zgradbe.

V disertaciji je pokazano (na osnovi že obstoječih podatkov, zbranih v ZDA), da se simetrična središčno-kohezivna vrsta bločnega modela pojavlja v omrežjih interakcij med predšolskimi otroki. To, da je mogoče omrežja s tako vrsto bločnega modela generirati z upoštevanjem navedenih lokalnih omrežnih mehanizmov, še ne pomeni, da so globalne zgradbe v empiričnih omrežjih nastale kot posledica analiziranih (v simulacijski študiji) lokalnih omrežnih mehanizmov. Ne glede na to, pojav take globalne zgradbe v omrežju prinaša nekatera pomembna vprašanja, povezana z (psihološkim) razvojem otrok, na katera je mogoče odgovoriti z upoštevanjem globalne zgradbe omrežja ter lastnosti otrok. Na primer, ali obstajajo razlike v določenih psiholoških značilnostih med otroki iz centralne skupine ter drugimi ter ali je koristno spodbujati nastanek take globalne omrežne zgradbe med predšolskimi otroki ali ne.

Zgoraj opisana metodologija je uporabljena tudi za generiranja omrežij z drugimi izbranimi vrstami bločnih modelov. Z upoštevanjem izbranih lokalnih omrežnih mehanizmov je mogoče generirati omrežja s kohezivnim, (simetričnim in asimetričnim) središčno-perifernim ter tranzitivnim bločnim modelom, ni pa mogoče generirati omrežij s hierarhičnim bločnim modelom, hierarhično-kohezivnim bločnim modelom ter tranzitivno-kohezivnim bločnim modelom.

Hierarhični bločni model z nekohezivno zadnjo skupino lahko nastane kot rezultat mehanizmov, povezanih s stroški (hierarhični položaj alterja, staž alterja, stopnja popularnosti alterja, število skupnih partnerjev ega in alterja), in mehanizmov, povezanih z vrednostjo (razlika v hierarhičnem položaju ega in alterja, razlika v stažu med egom in alterjem, število skupnih partnerjev ega in alterja). V tem primeru se stroški in vrednost navezujejo na dojemanje alterjevega znanja s strani ega (Nebus, 2006). Omenjena vrsta bločnega modela lahko nastane tudi v primeru prisotnosti novih enot ter osipa enot. Zmožnost generiranja globalnih omrežnih zgradb znotraj tega družbenega konteksta, z upoštevanjem lokalnih omrežnih mehanizmov, ki ne upoštevajo lastnosti

vozlišč (z izjemo staža), nakazujejo na to, da je mogoče oblikovati take politike podjetja, ki spodbujajo nastanek želenega vzorca pretoka znanja (če tak želen vzorec pretoka znanja v podjetju obstaja).

Osrednji prispevek disertacije je spoznanje, da lahko najbolj znane vrste bločnih modelov nastanejo kot rezultat zelo osnovnih lokalnih omrežnih mehanizmov, brez upoštevanja lastnosti vozlišč. Pri analizi razvoja bločnih modelov v empiričnih omrežjih je nujno upoštevati družbeni kontekst nastanka empiričnih omrežij ter vpliv družbenega konteksta na posameznikovo vedenje (Doreian in Conti, 2012).

Ključne besede: analiza družbenih omrežij, razvoj omrežja, lokalni omrežni mehanizmi, globalne omrežne zgradbe, bločni model, modeli eksponentih slučajnih grafov, verjetnostno modeliranja na ravni posameznika, modeli razvoja omrežij, funkcija relativnega prileganja, omrežja interakcij, omrežja pretoka znanja, predšolsko okolje, organizacijska omrežja.

Table of contents

List of tables

List of figures

List of abbreviations and symbols

NEM	*Network Evolution Model*	model razvoja omrežja
LE	*levels of errors*	raven napak
RF	*Relative Fit*	relativno prileganje
CF	*Criterion Function*	kriterijska funkcija
MIV	*Mean Improvement Value*	povprečna vrednost izboljšanja
SAOM	*Stochastic Actor-Oriented Models*	verjetnostni modeli na ravni posameznika
ERGM	*Exponential Random Graph Models*	modeli eksponentnih slučajnih grafov
IB	*Inconsistent Blocks*	neskladni bloki
NAM	*Nodal Attribute Models*	modeli na osnovi lastnosti enot
MCMC	*Markov Chain Monte Carlo*	Monte Carlo Markovske verige
OTP	*Outgoing Two-Paths*	izhodne poti dolžine dve
OSP	*Outgoing Shared Partners*	skupna vozlišča izhodnih povezav
ITP	*Incoming Two-Paths*	vhodne poti dolžine dve
ISP	*Incoming Shared Partners*	skupna vozlišča vhodnih povezav
MLE	*Maximum Likelihood Estimator*	ocena na podlagi največjega verjetja

1 Introduction

A network is defined by the set of nodes (also called vertices, units or actors) and by the set of links which represents ties between the nodes. These two sets determine a graph which describes the network's structure. Additional data can be assigned to the nodes and links to describe their properties (also called attributes) (Batagelj, Doreian, Ferligoj, & Kejzar, 2014, pp. 18–19).

Since the nodes and links can be defined very broadly and because the networks so obtained can be a consequence of different phenomena and processes, network analysis is applied widely across many scientific disciplines from both the natural and social sciences (Borgatti, Mehra, Brass, & Labianca, 2009; Hidalgo, 2016). It is worth mentioning that the field of (social) network analysis has grown significantly in the last few decades. This is reflected in the number of publications and amount of scientific disciplines in which researchers are using the network analysis approach (Maltseva & Batagelj, in press).

The main attention in this dissertation is given to the networks that are usually analyzed by social scientists, yet many research results can be generalized to other types of networks as well. When social networks are analyzed, the nodes are typically individuals and the links between them are operationalizations of the relationships among them (Handcock, Robins, Snijders, Moody, & Besag, 2003).

The most fundamental concepts, such as social mechanisms, local network structures and global network structures, are presented in this introductory chapter. These are necessary to understand the two general research questions proposed in subsection 1.4. Both relate to the main goal of this dissertation of studying the relationship between local network mechanisms (social mechanisms) and global network structures.

1.1 Social mechanisms

Since the study of social mechanisms is very broad and forms its own field of research, the aim of this section is to establish the underlying grounds for the discussions on social mechanisms that follow.

In the past, thinking about explicitly formulated mechanism-based theories was more common in other scientific fields (e.g. economics) than in sociology. Among the sociological classics, Robert Merton introduced the idea of mechanisms, yet never developed a very clear definition of the mechanism concept (Hedström & Wennberg, 2017). He rejected all attempts to develop general systems of sociological theory and instead asserted that sociological theory should deal with social mechanisms. He thought about social mechanisms as "social processes having designated consequences for designated parts of the social structure" (Merton, 1949, p. 451). He viewed mechanisms as the building blocks of middle-range theories (Boudon, 1991). Among non-social scientists, the biologist Francis Crick preferred to think in terms of mechanisms instead of laws, arguing that the notion of laws is generally reserved for physicists, who are the only ones able to produce explanations based on powerful and often counterintuitive laws with no significant exceptions (Crick (1989, p. 138)[1] in Hedstörm & Swedberg (1998)).

Even though various authors addressed the question of social mechanisms and provided a wide range of definitions of the term (Hedström & Ylikoski, 2010), it is still worthwhile to find a general definition of social mechanism that captures the concept's essence (Hedström & Swedberg, 1996). Gambetta (1998, p. 102) defines a mechanism as "a hypothetical causal models that make sense of individual behaviour. They have form, 'Given certain conditions, an agent will do x because of [mechanism] M with probability p'". Some other common definitions of the mechanisms are collected by Hedström & Ylikoski (2010).

As summarized by Hedström & Swedberg (1998), Harre (1970) wrote that one of the most important functions of the mechanism is when it performs in an exploratory account. For example, let us assume we observed the association between I (input) and O (output). In this case, we are interested in finding mechanism M that would explain the observed association. This means we are not satisfied with just the fact that the association between I and O exists – instead, we are interested in *why* the observed association exists. This may help distinguish between genuine causality and coincidental association, thus increasing understanding of why we observe what we observe. The mechanisms can be hard to recognize and identifying such underlying mechanisms is sometimes the hardest part of the scientific work entailed (Hedström & Wennberg, 2017).

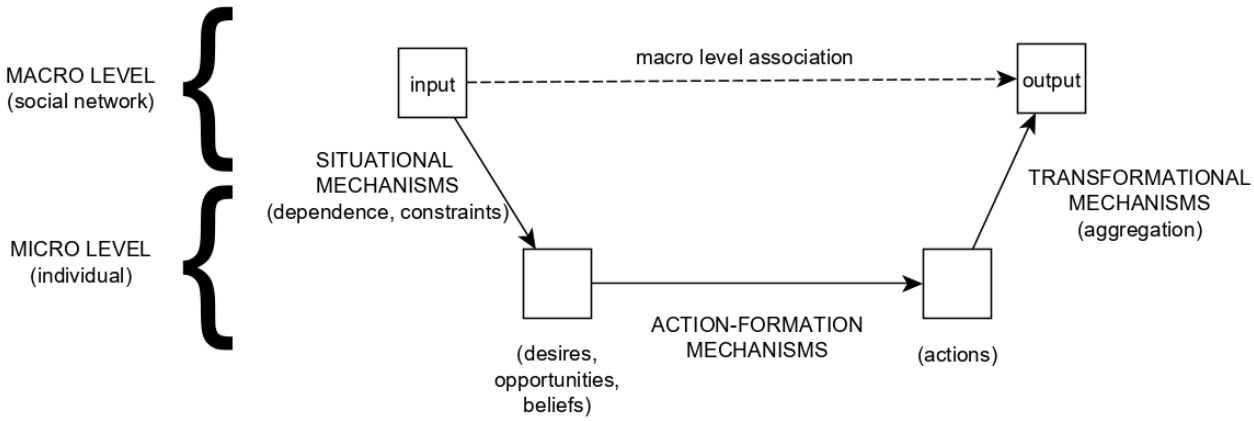
[1] Crick, F. (1989). *What mad pursuit: A personal view of scientific discovery*. London: Penguin Books.

Hedström (2005, p. 25) wrote that "a social mechanism /…/ describes a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome. We explain an observed phenomenon by referring to the social mechanism by which such phenomena are regularly brought about".

When searching for such mechanisms, it is not enough to explain the relations between macro properties, but one must provide a detailed description of the mechanisms on the micro level and, from here, explain the relations on the macro level (Hedström & Ylikoski, 2010; Ylikoski, 2012). Here, the micro can refer to, e.g. individuals in the context of macro organizations, or the micro can refer to e.g. organizations in the macro context of inter-organizational networks.

In the setting of the emergence of a global network structure (as summarized by Stadtfeld (2018)), studying how the global network structure changes on the global level is not sufficient. Instead, one must understand how a current global network structure affects individuals' characteristics (e.g. their desires, opportunities, beliefs) and how a current global network structure affects the set of possible actions of given individuals. Then, one has to understand how the individuals' characteristics and possibilities are reflected in actual changes in the links on the individual level. One is then able to study how the changes in the links of many individuals affect the global network structure (Coleman, 1994) (Figure 1.1).

Figure 1.1: A social network perspective on Coleman's macro-micro-macro model



Sources: Stadtfeld (2018), Hedström & Swedberg (1998) and Coleman (1986).

Following Coleman's macro-micro-macro model (Coleman, 1986), the mechanisms can be placed in three classes (Hedström & Swedberg, 1998): (i) situational mechanisms (how the macro

environment shapes actors' opportunities, goals, beliefs etc.); (ii) action-formation mechanisms (how these opportunities, goals, beliefs etc. influence the actors' behaviour); and (iii) transformational mechanisms (how the behaviour of the individuals affects the macro output). As noted (Hedström & Ylikoski, 2010), the observed macro phenomenon can only be explained by considering all the described mechanism types, as also illustrated by Doreain & Conti (2012) who showed how contextual features ('input' in Coleman's macro-micro-macro model; which can be represented by a current global network structure in this dissertation) affect the formation of ties in a network.

Agent-based simulations (Bianchi & Squazzoni, 2015; Hedström & Manzo, 2015; Macal & North, 2010), whose development started in the 1960s, are often used to study social mechanisms. They are not necessarily relied on to explain any particular empirical fact directly (yet may be used to integrate theoretical ideas with the results of empirical research), but to provide a general understanding of "how things could work" (Hedström & Ylikoski, 2010). These models assume that very complex social dynamics (or at least the basic cogs and wheels of these social processes) can be explained with quite simple models (Hedström & Ylikoski, 2011, p. 396). Goldthrope (1996; 2001) proposed a three-step based methodology for studying the causation between macro-level input and output by considering the micro-level actions within agent-based simulations:

> (i) establishing the phenomena that form the *explanada*;
> (ii) hypothesizing generative processes at the level of social action; and
> (iii) testing the hypotheses.

As he explained, the phases of the three-step methodology are unlikely to be that separable in applied sociological studies. This is also the case with a more detailed five-step structure of a generative research strategy (Epstein, 2006; Hedström & Bearman, 2009; Hedström & Ylikoski, 2010), namely:

1. Start with a clearly delineated social fact that is to be explained.
2. Formulate different hypotheses about relevant micro-level mechanisms.
3. Translate the theoretical hypothesis into computational models.
4. Simulate the models to derive the type of social facts that each micro-level mechanism brings about.
5. Compare the social facts generated by each model with the actually observed outcomes.

The 4th and 5th steps depend on the nature of the simulations and the social phenomenon being studied. Therefore, the social fact to be compared can be reflected by a given global network structure (as in this dissertation) or differently. For example, Steglich et al. (2019) studied the impact of bilingual education on segregation. In his study, the social output (segregation) was measured by the number of components, the number of isolates, a fragmentation index, geodesic distance, a segmentation index, and many other local and global network statistics. He studied the impact of the selected local (network) mechanisms on different indicators of segregation by generating many networks with different strengths of the local network mechanisms under study. Something similar was done by (Schaefer, Adams, & Haas, 2013), who studied the relationship between peer influence and the smoking alter popularity on smoking outcome.

Agent-based simulations can "look at the dynamic nature of social facts better than most other scientific methods" (Bianchi & Squazzoni, 2015, p. 285). Bianchi & Squazzoni (2015) noted the main advantages are the possibility of accounting for the irreducible heterogeneity of social behaviour, the possibility to consider out-of-equilibrium social dynamics and the possibility to deal with micro-generative processes (Bianchi & Squazzoni, 2015). Manzo (2007) is another author who recognizes (actor-oriented) simulation methods are an important tool in analytical sociology:

> If we recall the problems mentioned earlier, it can be claimed that these techniques reinforce that type of sociology precisely where it is weakest. Given that variable sociology tends to underestimate the role of theory, simulation works to strengthen theoretical models. Given that the language of variables underestimates the plurality of levels specific to sociological analysis and favors linear relations, simulation methods represent a powerful technical support for handling the micro-macro problem, and by directly modeling structures of interdependence among agents, they favor a "configurational", non-linear view of causation. Given that variable analysis requires generative mechanism reasoning to explain the empirical regularities it brings to light, simulation constitutes a tool for formally studying the mechanisms of phenomena production (N. Gilbert & Troitzsch, 1999).

Agent-based models have already been used to study various kinds of social phenomena, such as reciprocity, commitment, reputation, trust and signalling, trust and partner selection, segregation, opinion dynamics, collective behaviour, social stratification and others (Bianchi & Squazzoni, 2015).

Stochastic actor-oriented models (see subsection 2.3) may be seen as a special type of agent-based models (Snijders & Steglich, 2015). Here, the model simulates the actors' probabilistic choices of

ties, which are modelled as being dependent on mechanisms based on actors' attributes and positions in the network. Compared to other actor-based simulations, these models allow the relative importance of each considered mechanism to be estimated. A significant contribution to the development of actor-oriented models was made by Tom Snijders and colleagues (Snijders, Van de Bunt, & Steglich, 2010).

There are incentives to link the two general approaches which combine computational and statistical models to create something called "empirically calibrated computation network simulation models" (Hedström, 2005; Stadtfeld, 2018).

Stadtfeld (2018) highlights the importance of combining statistical and computational models. In his work, he used SAOM on empirical data in the first step, while in the second step the obtained model is used as an agent-based simulation model. By varying the share of nodes with homophily preferences, he shows that even actors with no preference for homophily are in the same places in the network where the majority of their network neighbours are.

The study by Stadtfeld (2018) was partly inspired by the work of Hedström & Aberg (2005) who used an "empirically calibrated actor-based model" to study social influence mechanisms in relation to youth unemployment rates. In the agent-based simulation model, they considered some socio-demographic features of the neighbourhood clusters in the network and the transition probabilities of leaving unemployment, which were estimated based on empirically observed data on youth unemployment in Stockholm between 1993 and 1999.

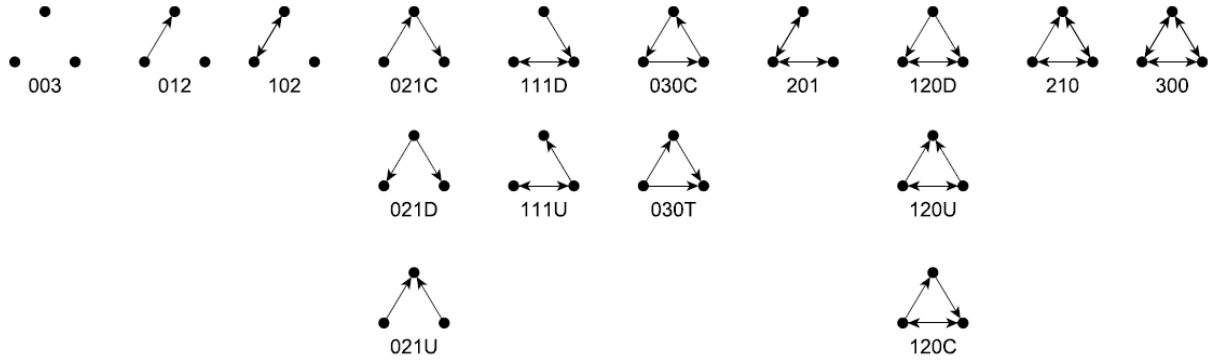
Very different social mechanisms can be considered when using agent-based models because there are no a priori constraints on how they are defined, except they have to be action-related (Hedström & Ylikoski, 2011, p. 396). The most frequently discussed social mechanisms in the social network analysis are mutuality (also known as reciprocity), popularity (also known as the Matthew Effect or preferential attachment), transitivity, and assortativity[2] (also known as assortative mixing or homophily) (Stadtfeld, Takács, & Vörös, 2018). They are discussed in more detail in later chapters.

[2] The mechanism assortativity may be considered through the assortativity of in-degree (the in-degree is often used as an indicator of popularity, although one must be aware of the difference between "friend nomination" and "popular student nomination") or other nodes' attributes, such as gender (Maccoby & Jacklin, 1987; Vaughn, Colvin, Azria,

1.2 Local network structures

When studying empirical networks, social mechanisms are often operationalized by rules (called local network mechanisms) governing how the links between studied nodes are created, maintained and dissolved. These rules can be represented in various ways (e.g. as 'if-then' rules or through the local network statistics), depending on the type of (statistical) model (e.g. Exponential Random Graph Models, Stochastic Actor-Oriented Modes, Network Evolution Models etc.). But, in some cases, different local network structures can be used to study possible social mechanisms that lead a network towards an observed global network structure or to study an observed global network structure.

Figure 1.2: The collection of all triad types (triad census)



Source: Davis and Leinhardt (1967).

Such local network structures can be different graphs of size three (triads). The classification (called MAN classification) of all possible triad types was proposed by Davis & Leinhardt (1967). It is visualized in Figure 1.1. in which all the triad types are labelled. The labels consist of three digits: the first digit denotes the number of mutual links (↔), the second stands for the number of arcs (→) while the third denotes the number of non-links (or negative links) between two nodes. Some types of triads with the same distribution of links are further differentiated (see columns) and labelled with a letter (C stands for cycle, T for transitivity, U for up and D for down).

In the past, different triad types were used for different purposes. The best known is to test for the existence of a specific global structures (see Section 1.3) in a given network using the method of

Caya, & Krzysik, 2001). When the assortativity of in-degree is used, the assortativity mechanism is sometimes referred to as »popularity-based homophily«.

counting the number of allowed/forbidden triad types in the network. However, application of this methodology is limited when analysing an empirical network with a certain level of errors. Holland & Leinhardt (1970), therefore, emphasized the importance of a probabilistic approach to test for the existence of a certain global network structure by counting the number of allowed/forbidden types of triads. Thus, they carried out the distribution of different types of triads (the expected number and the variance) in the case of a random network and defined the test statistics as

$$\tau = \frac{T - \mu_T}{\sigma_T} \tag{1.1}$$

where T is the number of a certain triad type in an empirical network, μ_T is the expected number of a certain triad type in a random network, and σ_T stands for the standard deviation of the number of the same type of triads in a random network. Holland & Leinhardt (1970) assumed that the distribution of the test statistics is asymptotic normal, which was used when testing for the frequency of the selected triad types in empirical networks. Nowadays, the term motif is often used. A motif is defined as pattern of interconnections occurring in complex networks at numbers significantly higher than those in randomized networks (Milo et al., 2002), which also include other types of subgraphs of any size (but usually up to four nodes), not only different types of triads. Artzy-Randrup et al. (2004) warn that the randomization of networks, which forms the basis for testing the null hypothesis, can be defined in different ways. However, motifs were successfully used to e.g. describe the hierarchical structure of the human brain (Yu & Gerstein, 2006) or cluster empirical networks (Milo et al., 2004).

Considering the number of chosen subgraph types can also be used to characterize the global network structure when the networks are too large to use e.g. the blockmodeling approach (see subsection 2.4) (Doreian & Mrvar, 2016).

1.3 Global network structures

Moreno studied how an individual's involvement in social relations affects their psychological well-being and was one of the first to apply the network methodology to analyse a certain psychological phenomenon. He called a network a sociogram. Here, the nodes were students and the links among them represented the different kinds of relationships they had (e.g. willingness to

share a school table or friendships). Based on his sociograms, he was able to show that the links among the nodes form a certain structure which is not random and is changing over the years (Moreno, 1934).

Based on many observed empirical networks and various psychological and sociological theories, subsequent researchers have proposed different models of global network structures. A well-known model is the balance model (Cartwright & Harary, 1956), which provides a generalization of balance theory (Heider, 1946). The balance model assumes exactly two clusters of nodes (called cliques[3]) where the nodes within each cluster are linked with positive links and the nodes from different clusters are linked with negative links[4] (non-links can also exist). Johnsen (1985) summarized the model in such a way that two cliques are assumed, where only positive links are present within cliques while the nodes from different cliques are linked by a negative or null (the absence of links) links. In this context, the next, clustering model (Davis, 1967) can be seen as a generalization of the balance model with an arbitrary number of clusters[5].

Different cliques[6] can form a single hierarchy, which is assumed by the ranked clusters of the M-cliques model (Davis & Leinhardt, 1967). There are no links (or only negative links exist) among the nodes from the cliques which are on the same level, while asymmetric links exist between the cliques, which are on different levels. The transitivity model (Holland & Leinhardt, 1971) is

[3] In this context, the term "clique" is not defined as a group of nodes which are all linked to each other, but some non-links can also exist. To be more precise (Davis & Leinhardt, 1967, p. 7), "/…/ cliques are subsets of individuals with higher rates of positive relationships among themselves than with outsiders".

[4] Cartwright & Harary (1956, p. 290) highlighted several ambiguities and limitations of Heider's theory of balance. They relate to the treatment of non-symmetrical relations, the generalization to systems containing more than three entities, the distinction between the complement (e.g. link vs. no-link) and the opposite (e.g. positive vs. negative link) of a relation, the simultaneous existence of relationships of different types and the applicability of the concept of balance to empirical systems other than cognitive ones. Based on empirical evidence obtained by Jordan (1953), Cartwright & Harary (1956, pp. 290–291) assumed that some types of relations (e.g. "has no sort of bond or relationship with") should be considered as the absence of a link.

[5] Doreian and Mrvar (2009) proposed the relaxed structural balance model (for signed networks) which allows for negative blocks to appear anywhere in the blockmodel, not only on the off-diagonal blocks (see subsection 1.3.1 for definitions of blockmodel and block). Also, positive blocks can appear anywhere in the blockmodel.

[6] Cliques are defined here in the same way as for the models described previously.

defined similarly as the ranked clusters of the M-cliques model, but here different cliques can form several hierarchies (a graph may consist of several components if only positive and null links are possible).

Davis & Leinhardt (1967) proved the existence of seven specific types of triads (see subsection 1.2) is a necessary and sufficient condition to form a global structure consisting of several hierarchically ordered cliques. Similar classifications of allowed/forbidden triad types were made for all other mentioned global structures (see Johnsen (1985)). Since one triad type was systematically more common in many empirical networks, as would be expected on the assumption of a given global network structure, Johnsen (1985) proposed a hierarchical model with M cliques, where the nodes within the cliques form a hierarchical structure.

1.3.1 Blockmodels

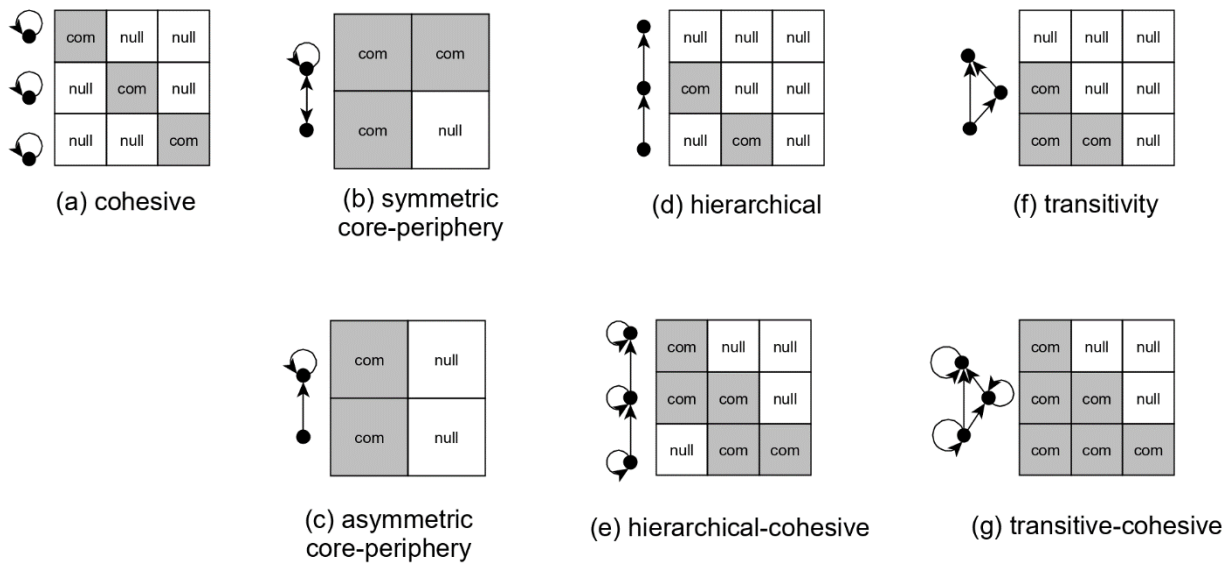
Today, considerable attention is being paid to the global network structures that can be described by blockmodels (Doreian et al., 2005) in both social network analysis and other scientific fields. The development of blockmodeling was initially based on social theories in the last century (Homans, 1950; Lewin, 1936; Nadel, 1957). Later, blockmodeling was used in various scientific fields (Alderson & Beckfield, 2004; Barnett & Danowski, 1992; Glückler & Panitz, 2016; Kronegger, Ferligoj, & Doreian, 2011; L. Prota & Doreian, 2016; Laura Prota, D'Esposito, De Stefano, Giordano, & Vitale, 2013; Žiberna & Lazega, 2016).

The term "blockmodel" reflects the fact that if a network is represented by a matrix, which is then split according to a partition (a set of non-overlapping clusters), blocks (submatrices) are formed in the matrix. The term "block" refers to a submatrix showing the links between nodes from the same or different clusters. Two selected nodes are structurally equivalent (also see subsection 2.4.1) if they have links to the same set of other nodes (Batagelj, Ferligoj, & Doreian, 1992; Lorrain & White, 1971). This is not the only definition of equivalence. Structural equivalence (Lorrain & White, 1971) and its generalization – regular equivalence (White & Reitz, 1983) – are the most common equivalencies. When structural equivalence is used, only null and complete blocks are possible. In ideal complete blocks, all possible links are present while there are no links in ideal null blocks.

An assumption often made with blockmodeling that similar (or equivalent) nodes are of the same type and therefore share the same rules on social behaviour (Lorrain & White, 1971; Michaelson & Contractor, 1992). The decision on whether to use one type of equivalence or another in the blockmodeling context depends on the topic under study and on whether the blockmodeling is being used as a data-reduction technique or serves by way of operationalization of a social role (Borgatti & Everett, 1992). In the latter case, structural equivalency is often criticized for being too restrictive (Borgatti & Everett, 1992) yet it is also worth mentioning that actors can have the same social role without being equivalently linked to the same others (Mizruchi, 1993).

Several blockmodels are well-known (Doreian et al., 2005, p. 236) and studied. They are visualized in Figure 1.3 and described in more detail in the following paragraphs.

Figure 1.3: Different blockmodel types represented by a graph and image matrix



Cohesive blockmodel

According to Doreian, Batagelj & Ferligoj (2005), a cohesive blockmodel is defined by several clusters of nodes which are internally highly linked, but where no links exist between the nodes from different clusters.

This very basic global network structure was studied e.g. in the context of the structural organization of the brain (Shen, Hutchison, Bezgin, Everling, & McIntosh, 2015). This is

(approximately, in some cases) the global structure also found by community-detection methods (see Lancichinetti & Fortunato (2009)).

Symmetric and asymmetric core-periphery blockmodel

Even though the core-periphery structure is one of the most typical and most often analysed, it is defined differently by various authors (Borgatti & Everett, 2000; Nordlund, 2018). Borgatti & Everett (2000) summarize the three intuitive views of the core-periphery structure. The first assumes the existence of one cluster to which the nodes belong to a greater or lesser extent. Similar to the first definition is the definition whereby links between the core nodes exist and where are no links between the peripheral nodes. The links between the core and peripheral nodes can either exist or not. The third definition is spatially defined: the core nodes which are positioned in the centre of the Euclidean space are close to all nodes in the network while those on the outskirts are close only to the centre. These definitions all consider one core and one periphery, whereas definitions referring to several cores also can be found (Cugmas, Ferligoj, & Kronegger, 2016; Kronegger, Mali, Ferligoj, & Doreian, 2012). Rombach et al. (2017) considered the situation where several core-periphery structures (as defined by Doreian, Batagelj & Ferligoj (2005)) appear in a single network.

In this study, the definition of core-periphery blockmodel comes from Doreian, Batagelj & Ferligoj (2005). In their definition, a core-periphery blockmodel consists of one internally well-linked cluster of core nodes and one cluster of peripheral nodes, which are not linked to each other. There are links between the core and the periphery. The core-periphery blockmodel is called symmetric in this study[7] when the links between the peripheral and core nodes are mutual (Figure 1.3b), and asymmetric when only the peripheral nodes are linked to the core ones (Figure 1.3c). Another version of the asymmetric core-periphery blockmodel is where only the core nodes are linked to the peripheral ones (Figure 1.3h)[8]. The definition used in this study is today one of the most common definitions of the core-periphery model (Nordlund, 2018).

[7] Doreian, Batagelj & Ferligoj (2005) call the asymmetric versions of the core-periphery blockmodel type a "centralized blockmodel".

[8] In this dissertation, the first version of the asymmetric core-periphery blockmodel is referred to by the name "asymmetric core-periphery blockmodel".

The core-periphery structure lies in the middle of several extreme properties, e.g. clique vs. star configurations, network assortativity vs. network disassortativity, hierarchy vs. non-hierarchy etc. (Csermely, London, Wu, & Uzzi, 2013).

A clear core-periphery blockmodel was found among high school students, where a relation was defined by one student asking another student to borrow their study notes (Batagelj, Mrvar, Ferligoj, & Doreian, 2004). This blockmodel was also found when studying individual creative performances in the Hollywood film industry (Cattani & Ferriani, 2008), in the analysis of metabolic networks (Da Silva, Ma, & Zeng, 2008), and in many studies of scientific co-authorships (Chinchilla-Rodríguez, Ferligoj, Miguel, Kronegger, & de Moya-Anegón, 2012; Cugmas et al., 2016; Hu & Racherla, 2008).

Hierarchical blockmodel and hierarchical-cohesive blockmodel

When a hierarchical blockmodel consists of three clusters, then a cluster of internally non-linked nodes exists which are all linked to the second cluster of internally non-linked nodes and the nodes from the second cluster are all linked to the third cluster of nodes which are also not internally linked to each other.

A hierarchical-cohesive blockmodel is characterized by complete blocks on the diagonal of an image matrix, which means that the nodes belonging to a certain cluster are internally highly linked. Also, in the case of a hierarchical blockmodel, the clusters of nodes can be hierarchically ordered.

A hierarchical structure is often associated with companies' organizational structure. Oberg & Walgenbach (2008) analysed employee communications in a given company. Even though the company's policies encourage the principles of non-hierarchical functioning (including ways of communicating among the employees), they confirmed that their day-to-day communication on the intranet indicates the existence of a hierarchy within the organization.

Transitivity blockmodel and transitive-cohesive blockmodel

A transitivity blockmodel is similar to a hierarchical blockmodel, except that in the case of a transitivity blockmodels links also exist from the clusters on the lowest level to all clusters on the highest levels (or vice versa). In the literature, both hierarchical and transitive global network structures are often called hierarchical.

A transitive-cohesive blockmodel is similar to a transitivity blockmodel, but with the former one there are links between nodes from the same cluster.

The definitions of the transitivity, transitive-cohesive, hierarchical and hierarchical-cohesive blockmodel types, used in this dissertation, consider the links from the nodes on a lower hierarchical level to those on a higher hierarchical level. Well-known are also definitions, where the links goes from the nodes from a higher hierarchical position to the nodes from the lower hierarchical level.

1.4 Mechanisms and blockmodels

The emergence of a given global network structure is a consequence of certain (social) processes. Different processes and mechanisms can cause different global network structures. For example, in the context of social capital, different habituses form within different groups of individuals that restrict the influx of newcomers. Accordingly, the boundaries between different groups are maintained, resulting in a cohesive structure (Lutter, 2013).

The core-periphery model can be the outcome of many social mechanisms (e.g. cooperation, where the defectors are pushed out to the periphery (Sohn, Choi, & Ahn, 2019)), but it is often associated with the existence of elites. An elite group is a small core of nodes that are all linked to each other. Compared to the peripheral nodes, the core nodes have greater prestige, defined by a higher number of incoming ties. The existence of elites may be due to the competitive interactions among the nodes (Csermely et al., 2013). After studying the bank market, In't Veld et al. (2016) confirmed that for big banks, which have more trading opportunities, it appears to be more economically beneficial to establish links to other big banks, which produces a core-periphery structure.

For studying hierarchical structures, Chase (1982) and Fararo & Skvoretz (1986) proposed a probabilistic model with two effects: the ‘victim effect’ (when a node attacks another node, a tie between them is established) and the ‘bystander effect’ (the bystander tries to attack the victim and protect himself from the attacker, where the attacker is trying to dominate the bystander). They report that the bystander effect is required for a hierarchical structure to be formed.

A hierarchical structure was also found among adolescents. Eder (1985) described the emergence of a hierarchy among middle-school girls based on qualitative observations. A group of girls

achieved a high level of popularity through cheerleading elections. This group was liked highly by other girls and, therefore, they endeavour to make friendships with the highly popular girls. The less popular girls also recognized that making friendships with the more popular girls can affect their own popularity status. Not only due to the natural limit on the number of friendships one can maintain, but also to avoid losing high status by having relationships with lower-status girls, the most popular girls ignored the affiliative attempts made by many girls, which led to the fact that the most popular ones became increasingly disliked. Similarly, Dijkstra, Cillessen & Borch (2013) confirmed (based on analyses of temporal friendship networks collected among middle-school students) that popularity increases the receipt of best-friend nominations, but decreases upon giving them and that higher-status adolescents strive to maintain their status by keeping lower-status adolescents at a distance.

Later, Rubineau, Lim & Neblo (2019) studied the relationship between negative ties and social status among university students. To this end, they analyzed temporal network data collected between 2008 and 2012. The key dependent variable was a negative tie between two students, while a key explanatory variable was social status as operationalized by different network statistics, obtained on networks with positive ties. Their results show that negative ties are much more likely to go from higher-status individuals to lower-status individuals than vice versa.

After studying the evolution of new scientific disciplines by applying social network methodology, Bettencourt et al. (2009) reported that new scientific disciplines emerge through the linking of small unlinked groups of researchers until one big component is formed. During further development of a scientific discipline, the reverse process is launched. In this process, a big component starts to split into several smaller and less connected groups of researchers. Similarly, the global structures of networks among preschool children in kindergartens are strongly determined by the popularity mechanism in early stages while, later on, the transitive closure mechanism emerges (Leinhardt, 1973; Schaefer, Light, Fabes, Hanish, & Martin, 2010). In the blockmodel type context, this may indicate the transition from the core-periphery blockmodel to the cohesive blockmodel with one or several popular clusters of individuals.

1.5 Models that generate networks by considering local network mechanisms

Many statistical models seek to explain how local network mechanisms affect a certain global network structure (Toivonen et al., 2009). In the social network context, these models usually aim to explain the characteristics of a given global network structure through selected local network mechanisms (or to study the impact of a certain local network mechanism on a global network structure). Most of such models attempt to recreate/simulate different network characteristics such as specific density, distribution of degree or value of a clustering coefficient (Kejžar, 2007) rather than networks with a particular global structure (blockmodel) as described in the previous sections.

Two models for generating networks with a specific expected density are $G(n, p)$ proposed by Gilbert (1959) and $G(n, m)$ proposed by Erdős & Rényi (1959). While these models seek to generate random networks, they can be modified to create networks with a pre-specified blockmodel structure in such a way that each block of the assumed blockmodel is generated separately but with different parameters (the density is controlled for each block separately) regarding the type of block (e.g. null vs. complete). This approach is in line with mixture models for random graphs (Daudin, Picard, & Robin, 2008).

Because these models do not enable the impact of different mechanisms on the global network structure to be studied, the main focus is placed on more complex models, which require some additional assumptions. These include Network Evolution Models (NEMs) (Toivonen et al., 2009) which are used to test hypotheses that specific local network mechanisms lead the network towards specific global structures and other hypotheses related to the network's evolution (see section 2.1). For example, Kumpula et al. (2007) tested a hypothesis about the emergence of a cohesive network structure. They considered two local network mechanisms. The first one, called "local attachment", is similarly defined as a transitivity mechanism (in their study, the values on the links are considered), while the second one, called "global attachment", represents a situation where a given node creates a link to a randomly selected node (each link is created with equal probability). By considering these two mechanisms (and an additional one for removing and adding nodes), they are able to generate networks with a global network structure similar to the cohesive global network structure and which possess many global properties common to social networks.

In general, when using a NEM to generate networks one has to define the mechanisms through the rules on how the ties are established or dissolved, which are then used in an iterative procedure. Unfortunately, NEMs do not enable the statistical significance of some defined mechanisms to be estimated on empirical data (Snijders, Van de Bunt, et al., 2010) and only a limited number of mechanisms can be considered at once (Stadtfeld, 2018).

This limitation can be overcome by using either Exponential Random Graph Models (ERGM) (Hunter et al., 2008; Koskinen & Snijders, 2013; Morris, Handcock, & Hunter, 2008) or Stochastic Actor-Oriented Models (SAOM) (Block, Stadtfeld, & Snijders, 2016; Handcock et al., 2003; Snijders, Van de Bunt, et al., 2010). In both cases, the researcher must specify one or several terms/mechanisms/effects (which usually operationalize one or several local network mechanisms) and corresponding parameters (related to the strength of these mechanisms and their effect on the global network structure). These values are usually estimated based on empirical networks. Once the model is stated and/or estimated, random networks can be generated based on the model.

There are fundamental differences between the ERGM and SAOM, which are described in more detail in section 2.2 and section 2.3. However, the biggest conceptual difference is that ERGM is a tie-oriented model (used to check the extent to which the global network structure can be explained considering the structure of the links and/or characteristics of the nodes) while SAOM is an actor-oriented model (used to test hypotheses about processes where the nodes have control over the changing of links).

1.6 General research questions

Based on the above theoretical foundations, two general research questions are stated.

RQ-1: Is it possible to generate networks with a given blockmodel type when considering only different types of triads?

In response to this research question, the triad types that cannot occur in a network with a given blockmodel without errors are defined. Based on this, Monte Carlo simulations are used to test whether one can generate networks with a given blockmodel type when considering only different

types of triads. The possibility of generating networks with a given blockmodel when considering only a subset of all possible types of triads is also examined.

Reducing the set of triad types for generating networks with a mentioned structure is especially important for a researcher wishing to generate such networks. The ambition is to reduce the amount of information needed to generate the network with a given blockmodel, which means the researcher does not need to perform an extensive pre-analysis. Therefore, it would be most beneficial for a researcher if they could generate networks with a selected blockmodel by considering only the forbidden triad types.

Other local network properties are considered when it turns out that networks with the mentioned blockmodel structures cannot be generated by considering only different triad types. Here, the question of generating random networks with an exactly defined global structure (without errors) as well as the question of how to generate networks with a given blockmodel type but with a certain amount of errors is addressed. To do this, the relationship between different types of triads (and other local network characteristics) and different levels of errors has to be examined.

This topic is particularly important since the ability to generate networks with certain blockmodel types considering only micro-substructures indicates that the global network structures being considered may be a consequence of only local network mechanisms and thus do not necessarily depend on the characteristics of the nodes.

RQ-2: Which mechanisms (or combination of several mechanisms) affect the change in blockmodel type?

The second research question entails a study of the mechanisms which affect the change in blockmodel type. The relationship between local network mechanisms and global network structure has been extensively studied based on either empirical networks or extensive agent-based simulation models (Bianchi & Squazzoni, 2015; Stadtfeld, 2018). Yet, no attention was placed on the link between local network mechanisms and specific lobal network structures, namely blockmodels. This will be addressed while investigating this research question. Given that there are many blockmodel types and many local network mechanisms, the research will be narrowed down to a set of selected blockmodel types and selected local network mechanisms since it is believed the selection of local network mechanisms and blockmodel types depends strongly on the

social context being studied. For each social context (e.g., friendships in kindergarten or the flow of knowledge within a company) considered, the initial and final blockmodels together with local network mechanisms will be selected based on previous studies.

1.7 Structure of the dissertation

This dissertation is organized in several chapters. The common research methodology is described in Chapter 2, while some methodological specifics are outlined in the following chapters which address the stated research questions.

Chapter 3 addresses first research question on generating networks with different blockmodels by considering only different triad types. The next two chapters discuss the context of the interactions among preschool children (Chapter 4 and Chapter 5). Here, an examination unfolds of the relationship between the most common local network mechanisms and the newly proposed (symmetric and asymmetric) blockmodel type, where the symmetric version is studied in the interactional preschool networks. The relationship between the most common local network mechanisms and the blockmodels is described in Chapter 6.

The emergence of hierarchical blockmodels in the knowledge-flow context is studied in Chapter 7. A global network structure very close to the hierarchical blockmodel was found in knowledge-flow networks collected in a medium-sized knowledge-based company in Slovenia.

The research results that are obtained are synthesized and critically evaluated in the Discussion chapter (Chapter 8).

2 Research methodology

Different methodologies are applied to respond to the stated research questions.

In Chapter 3, networks with a given blockmodel type (without errors or with a certain level of errors) are compared to random networks, according to the number of chosen local structures (e.g. different triad types). To this end, networks with a given blockmodel type are generated by using the mixture model for random graphs , while the random networks are generated using the $R(n, m)$ models. Considering the local network characteristics of the obtained networks, the NEM is defined and used to generate random networks (by considering the number of selected local network statistics) with a given blockmodel structure.

In the latter chapters, different NEMs are used to generate networks. Common to them all is that they consider the defined network statistics, which operationalize the selected local network mechanisms. In Chapter 4, asymmetric networks are described while in Chapter 5 symmetric networks are analysed. Growing asymmetric networks are presented in Chapter 6.

Many different blockmodels and local network mechanisms exist. They depend on the study context. Therefore, the analysed blockmodels and local network statistics are selected by considering the chosen social context. When the context is taken into account, relevant local network mechanisms can be selected and operationalized according to e.g. the literature review or intuition. For example, in Chapter 4 and Chapter 5 the preschool environment is considered, while in Chapter 6 the context of advice-giving in a medium-sized company is examined. Real-world networks are analysed by using ERGM or SAOM in order to gain a deeper insight into possible local network mechanisms. Where the data were already analysed by other researchers, the results are simply summarized and applied to this study.

The described models for generating random networks with a given blockmodel type are used in Monte Carlo simulations. The global network structures of the generated networks are evaluated by using the number of inconsistent blocks and the proposed Mean Improvement Value and the Relative Fit value, as described in more detail in the sections that follow. Here, the main interest is typically given to whether the desired blockmodels can be generated by considering selected local network structures (e.g. triad types) or selected local network mechanisms (e.g. popularity).

In the following sections, different models for generating networks are presented. Then, different approaches to blockmodeling are described, followed by several approaches to evaluating global network structures in the blockmodel context. Some of these approaches are newly proposed and their main characteristics are therefore illustrated by simulation studies.

2.1 Erdős-Rényi models, Network Evolution Models and Nodal Attribute Models

One model which generates networks with a specific density (namely, a network with exactly n nodes where each link is created with equal probability p independently of the other links) is $G(n,p)$ that was proposed by Gilbert (1959). Similarly, Erdős & Rényi (1959) proposed the $G(n,m)$ model which assumes exactly n nodes and m links, where each link is equally likely. In the case of a smaller number of links, nodes (called isolates) without any link can appear. With the number of nodes approaching infinity, the characteristics of networks generated by the $G(n,p)$ model approach the characteristics of networks generated by the $G(n,m)$ model, where $m = \binom{n}{2}p$.

The $G(n,p)$ and $G(n,m)$ models can be used to generate networks with an exact density or with an expected density, but without a pre-specified blockmodel structure. Yet, the models can be modified in such a way that each block of the assumed blockmodel is generated separately using the mentioned models but with different parameters (the density is controlled for each block separately) regarding the type of block (e.g. null or complete), which is in line with mixture models for random graphs (Daudin, Picard, & Robin, 2008). Even though these models can be used to generate networks with a given blockmodel type, they cannot be used to directly study the impact of different mechanisms on the global network structure. For this aim, Network Evolution Models (NEM) or Nodal Attribute Models (NAM) are available.

Toivonen et al. (2009) defined three core characteristics of NEMs: (i) each generated network is produced by an iterative procedure where the initial network is usually a network without any link (an empty network) or a small seed network; (ii) the links in the network are changing throughout the iterative procedure based on exactly specified stochastic rules (mechanisms) which include the selection of a subset of nodes and the links (or non-links) between them. Based on these mechanisms, both the nodes and links can be removed or added. In a certain case, the mechanisms can also include the nodes' characteristics; and (iii) NEMs can be further classified in two

subgroups, based on the stopping criterion of the iterative process, on a growing NEM (the iterative procedure stops when the network reaches a certain size) and a dynamic NEM (the iterative procedure stops when selected network statistics converge).

Within NEMs, the inductive class of graphs (ICG) must also be mentioned (Curry, 1963). It includes the class of initial graphs $\mathcal{B}$ and the class of rules (mechanisms) $\mathcal{R}$ which can be further divided into the left element on which the rule is implied and this produces the right element. Kejžar, Nikoloski and Batagelj (2008) extended the ICG by defining the probability space. They introduced the probability inductive class of graphs (PICG) which includes (besides $\mathcal{B}$ and $\mathcal{R}$) the probability distribution related to the selection of the initial graph $\mathcal{B}$, the probability distribution related to the selection of the rules from $\mathcal{R}$ and the probability distribution for the selection of the left element for each class of rules $\mathcal{R}$.

One of the main differences between NEMs and NAMs is that with NEMs a given link between two nodes depends on the configuration of links in the network (and not also on the nodes' attributes) while, with NAMs, the probability of a link is only determined by the attributes of the nodes. For example, it is often the case that links are established between more similar nodes (this is the so-called homophily effect (McPherson, Smith-Lovin, & Cook, 2001)). For instance, a study relying on observational network data collected among preschool children shows the children spent more time with those who had similar levels of preschool competency (DeLay et al., 2016). The results are significant even when controlling for covariates such as sex, age, language, family financial strain, parent education and receptive vocabulary, which are often considered as covariates related to the homophily mechanism.

2.2 Exponential Random Graph Models (ERGM)

Toivonen et al. (2009) classified ERGM as a separate category in their social network models. ERGMs are "a family of statistical models for social networks that permit influence about prominent patterns in the data, given the presence of other network structures" (Robins, 2011, pp. 484–485). They are used to check to what extent the global network structure can be explained, considering the structure of links and/or characteristics of the nodes. Compared to NEMs, ERGMs do not consider the (evolution) process, although the MCMC algorithms (see section 1.4.3) can be used to model the evolution of social networks (Snijders, 2001).

The above definition of ERGM is mostly summarized by Hunter et al. (2008). Let us assume a given random network Y (the concrete realization of a random network is denoted by y), consisting of N nodes. In such a network, the link between the i-th and j-th nodes is represented by a random variable Y_{ij}. The set of all possible networks is denoted by $\mathcal{Y}$. Therefore, the distribution of Y can be written as

$$P_{\theta,\mathcal{Y}}(Y = y) = \frac{\exp\{\theta^T g(y)\}}{\kappa(\theta, \mathcal{Y})}, \ \ y \in \mathcal{Y} \tag{2.1}$$

where θ is a vector of coefficients and $g(y)$ is a vector of statistics, estimated based on the matrix y. $\kappa(\theta, \mathcal{Y})$ is a normalizing constant which ensures that the sum of probabilities equals 1. The estimation is computationally very intensive, especially when networks with a higher number of nodes are involved. In the context of generating random networks based on the model above, the change statistics (Wasserman & Pattison, 1996) must be mentioned. It is defined as

$$\delta_g(y_{ij}) = g(y_{ij}^+) - g(y_{ij}^-) \tag{2.2}$$

where $g(y_{ij}^+)$ is the vector of statistics obtained on the network with the link between the i-th and j-th node, and $g(y_{ij}^-)$ is the vector of statistics obtained on the network without a link between the i-th and j-th node. It can be shown that probability $P_{\theta,\mathcal{Y}}(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$ (where Y_{ij}^c represents the rest of the network other than the single variable Y_{ij}) depends on y_{ij}^c only through $\delta_g(y)_{ij}$, which holds practical implications when generating random networks since it is often easier to compute $\delta_g(y)_{ij}$ than $g(y_{ij}^+)$ and $g(y_{ij}^-)$ separately.

There are several types of Markov Chain Monte Carlo (MCMC) algorithms for generating random networks. Generally, the start is represented by an empty network. Then, based on the uniform distribution one of the links or non-links is chosen. According to the model, the probability of establishing or dissolving a link is calculated and then, based on this probability, the chosen link or non-link is established or dissolved. The process is iterative. For each iteration, the change in the values of the estimated statistics before and after the change in the link between i and j is estimated – in other words, for each iteration $\delta_g(y)_{ij}$ is calculated. The iterative process stops when the approximate convergence to $P_{\theta_0,\mathcal{Y}}(Y=y_{proposed})$ is reached (Hunter et al., 2008). When the Metropolis algorithm is used, the network with a changed link is accepted with the probability

$$\min\left\{1, \frac{P_{\theta_0,y}(Y=y_{proposed})}{P_{\theta_0,y}(Y=y_{current})}\right\} \tag{2.3}$$

More general is the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) which is currently implemented in the "ergm" package of the R computer language (Hunter, Goodreau, & Handcock, 2013). However, in the case of the Metropolis-Hastings algorithm, the $y_{proposed}$ is chosen based on an auxiliary distribution which depends on $y_{current}$. $P(y_2|y_1)$ represents the probability in the Metropolis-Hastings algorithm that y_2 becomes a new proposed network on the condition that current network y_1 is given. Therefore, the proposed network is accepted with the probability

$$min\left\{1,\ \frac{P_{\theta_0,y}(Y = y_{proposed})}{P_{\theta_0,y}(Y = y_{current})}\frac{P(y_{proposed}\ |y_{current})}{P(y_{current}|y_{proposed})}\right\} \tag{2.4}$$

When using the Metropolis or Metropolis-Hastings algorithms (or Gibbs sampling), the $\kappa(\theta, \mathcal{Y})$ disappears from the ERGM likelihood ratio (Equation 2.1) and the ratio is simplified to

$$\frac{P_{\theta_0,y}(Y=y_{proposed})}{P_{\theta_0,y}(Y=y_{current})} = exp\{[g(y_{proposed}) - g(y_{current})]\} \tag{2.5}$$

Different methods can be used to estimate the parameters. A well-known one is the maximum pseudo likelihood estimator (MPLE) for which the estimates can be biased in some cases (Corander, Dahmström, & Dahmström, 1998), which arises from the assumption of independence between the nodes. The latter can also result in higher values of standard errors (Van Duijn, Gile, & Handcock, 2009). Desmarais & Cranmer (2012) proposed an efficient approach (based on non-parametric bootstrapping) to compute the confidence intervals for MPLE estimates.

The parameters' estimates are less biased and less variable when the Markov Chain Monte Carlo maximum likelihood estimator (MCMC-MLE) is used. The estimates are obtained using Monte Carlo simulations, which can be computationally very intensive, especially when the analysed network consists of a higher number of nodes. However, with the number of nodes the MPLE estimates approach the MCMC-MLE estimates (Hyvarinen, 2007; Strauss & Ikeda, 1990). Besides the limitation on networks with a smaller number of nodes, the main issue with ERMG is that the distribution of sufficient statistics can be multimodal (Snijders, 2002), particularly when statistics related to transitivity are included in the model (Jonasson, 1999). So-called degenerativity emerges

when the empirical data do not fit the estimated model (e.g. due to inadequate selection of the terms or explanatory variables). In such cases, the generated networks based on the model do not fit the empirical network on average (the distribution of estimated parameters is not appropriate). Usually, such generated networks are either completely empty or completely full (Handcock et al., 2003).

Model degenerativity can be avoided with a well-defined research question and thus the appropriate selection of explanatory variables or terms as they are called in the ERGM context. Use of the terms is typically based on the theory and the context (Goodreau, 2007; Morris et al., 2008). Different terms can be added and removed – at the end, the model that best fits the empirical data is chosen. Another technique for choosing the terms is by considering different forms of dependencies between the nodes (Frank & Strauss, 1986), which can be ordered in a partly ordered dependence hierarchy for ERGM (Block et al., 2016). Block, Stadtfeld & Snijders (2016) summarized the following most frequently used terms in practice: (i) when the independence of ties (all links are established with a certain probability, which does not depend on other links) is assumed, only the term "edge" (a term for the density) should be included in the ERGM model; (ii) when dyadic independence (the link $i \rightarrow j$ depends only on $j \rightarrow i$ and is independent of all other links in the network) is assumed, only the term for the density and the term for mutuality must be included in the ERGM model; and (iii) Frank and Strauss (1986) proposed Markov dependence (two links are conditionally independent of all other links in the network unless they have at least one node in common). When a Markov dependence is assumed, terms related to transitive triads and in-stars and out-stars must be considered in ERGM; (iv) in the case of social circuit dependence (the links between i and j, h and l are conditionally independent of other links in the networks if there are links exist between nodes i and j and also between h and l) proposed by Pattison & Robins (2002) the terms of a type 4-cycle and the terms related to different types of a geometrically-weighted edgewise shared partners must be included in the ERGM model.

Nevertheless, when studying empirical networks, the term edge is almost always included in the model since it controls the density (when only the term edge is included in the model the networks are generated by the $G(n,p)$ model). The density is generally treated as a random variable in ERGM models, which is more realistic in the case of modelling social networks compared to an assumed fixed density (although it is also possible to construct models with a fixed density), since

the density is seen as a product of social processes and it therefore cannot be known in advance (Hunter et al., 2008).

2.3 Stochastic Actor-Oriented Models (SAOM)

SAOM (Block et al., 2016; Snijders, 2001) are similar to ERGM yet different in some important details. First, it should be emphasized that ERGM models are tie-oriented, while SAOM ones are actor-oriented[9]. Both model types come from the tradition of generalized linear models, but the linear predictor is defined for the whole network with ERGM while with SAOM the linear predictor is defined based on the nodes

$$f_i(\beta, x) = \sum_k \beta_k s_{ki}(x) \tag{2.6}$$

where $s_{ki}(x)$ are chosen effects with the corresponding values of coefficients β_k for the i-th node. The above equation is usually named an objective function and gives the foundation for calculating the probability of changing the link of node i. The probability of changing a tie is calculated before the chosen node (called ego) has an opportunity to change a tie. An ego can be chosen at random or based on its characteristics.

Using ERGM or SAOM, hypotheses can be tested about the presence of different local network mechanisms (e.g. reciprocity, transitivity, homophily), considering the local network structures and/or characteristics of the nodes[10]. However, SAOM is more appropriate for testing a hypothesis about processes where the nodes have control over the changing of links. As an example, the network of friendships can be given. The friendship process is usually operationalized by a change of the 021C triad type to the 030T triad type, i.e. if node i is a friend of node j and node j is a friend of node k, then there is a higher probability that node i will become a friend of node k. In contrast to the friendship network, the example of a network of flight connections can be

[9] In this dissertation, the terms “node” and “actor” and the terms “tie” and “link” are used as synonyms.

[10] In ERGM, hypotheses about the number of different small network patterns (called configurations) are considered (Robins, 2011) and thus the local network “mechanisms” (defined as a set of rules for creating links) are not directly addressed since that is more the case of SAOM. However, Robins et al. (2009) provide some theoretical interpretations of different parameters in the context of different sociological concepts (e.g. path closure, tendencies for a structural hole to close, closure in the form of non-transitive cycles etc.).

considered. Here, two airports are linked if an air service between them is provided. This means airports do not have a direct impact on the establishing and dissolving of the links (this example is described in greater detail in Block, Statfeld & Snijders (2016) with the explanation that there is an organization which manages the flight connections, but its rule is different to the rule of a node in an, e.g., friendship network). In the case of the first example given, the use of SAOM is appropriate while for the second one ERGM models should be used.

Using SAOM, networks observed at a minimum of two points in time can be analysed ($k \geq 2$), where time is treated continuously. This assumption is usually satisfied with social networks when data are collected by a survey over several waves. Although several generalizations of ERGM have been proposed[11] to enable temporal networks to be analysed, ERGMs remain the most often used for studying empirical networks observed at one time point. When studying networks with different blockmodel types, this means comparing the given network with a blockmodel structure with a random network according to selected types of local network mechanisms (or local network structures).

Another important assumption of SAOM is that only one link can be changed at a time (a link can be established, dissolved or remain unchanged). It is therefore impossible for two nodes to establish a mutual tie at once (in the case of ERGM, a mutual tie can be established in a single step). Instead, one node has must establish a tie to another one, and then another node can establish the tie to the first node, resulting in a mutual relationship. The nodes control the outgoing ties, which means the ties are established based: (i) on the characteristics of those nodes which have an opportunity to change a link; (ii) the characteristics of other nodes and; finally (iii) on how other links in the network are configured. The final (empirical or generated) network is the outcome of a Markov process, implying that the network's structure is a social context which influences how that network's structure changes.

Given an empirical network, several methods for estimating parameters can be used in actor-based models, such as (generalized) Method of Moments (Snijders, 2001), Maximum Likelihood (Snijders, Koskinen, & Schweinberger, 2010) or a Bayesian estimation (Koskinen & Snijders,

[11] Many variations of ERGM for temporal networks have been proposed, e.g. separable temporal ERGM (Krivitsky & Handcock, 2014) and temporal ERGM (Desmarais & Cranmer, 2012; Hanneke, Fu, & Xing, 2010).

2007). These methods produce similar results with bigger data sets (Snijders, 2011). When the selected local network mechanisms are highly correlated, the standard errors of the estimated parameters are high and the estimated parameters' values vary highly from one estimation to another on the same empirical data. This can be avoided by removing or adding some effects (or node attributes) when analyzing empirical network data. However, it is not desirable to remove or add effects when the set of local network mechanisms is pre-defined.

Whereas in SAOM, ERGM and NEM, the estimated values of the parameters are not directly comparable, several approaches are proposed to determine the relative importance of the local network mechanisms (or terms/effects). One simple approach is to consider the odds ratios of choosing between two alternatives regarding the change of an outgoing link by node i (Snijders, Van de Bunt, et al., 2010). This may be accomplished based on the raw SAOM coefficients and may give an initial insight into the strength of an individual local network mechanism. Another possibility is to use the measure of explained variation proposed by Snijders (2004), which primarily considers the effect of adding extra local network mechanisms to the model rather than quantifying the relative importance of all mechanisms already included (where the change in the proportion of the variance explained is interpreted). The author says that "further experience with this measure will have to be collected to obtain better insights into what may be considered low and high values". Moreover, the approach is very computationally demanding and hence not useful in many real-network analyses (Indlekofer & Brandes, 2013).

Indlekofer & Brandes (2013) proposed an approach to calculate how strongly the probabilities (with which node i may change one of the outgoing links in a mini step) depend on each local network mechanism (effect) that is included. This may be used to compare the relative importance of the local network mechanisms within a model, among different models and on different data sets. Since the proposed measures are calculated on an individual level, they are averaged over nodes as implemented in the "Rsiena" package (Ripley et al., 2019) for the R programming language.

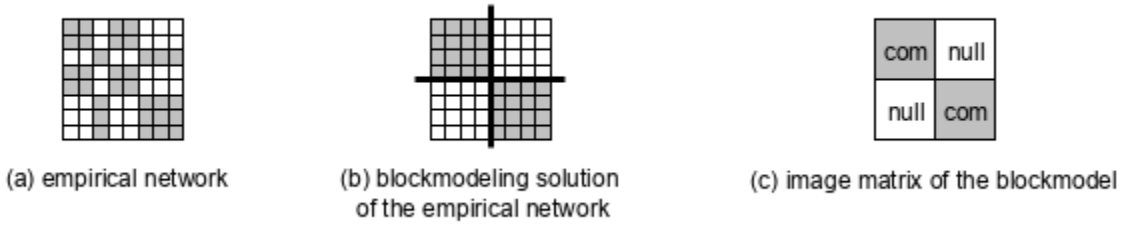
With NEM, the degree of comparability of the strength of different local network mechanisms varies according to their mathematical definition and the level of dependence between them. In this study (see the sections on generating networks in each chapter for more details), a partial level of comparability is achieved by normalizing the network statistics, which corresponds to the local

network mechanisms. They are usually normalized in such a way that they take values on the interval between 0 and 1. But none of these guarantee the absolute comparability of the strengths of the local network mechanisms.

2.4 Blockmodeling

Blockmodeling is a procedure for deriving a blockmodel from a given empirical network. "The goal of blockmodeling is to reduce a large, potentially incoherent network to a smaller comprehensible structure that can be interpreted more readily" (Batagelj et al., 2004, p. 456).

Figure 2.1: Example of an empirical network and its blockmodeling solution



A demonstration of blockmodeling considering structural equivalence is given in Figure 2.1. The original network is presented in matrix form in Figure 2.1a. Here, each row and column represent a unit. Gray-coloured cells in the matrix represents a link from the i-th node (row) to the j-th node (column). Cells on the diagonal represent loops (a given node is linked to itself). The nodes are permuted (see Figure 2.1b) such that those with the same pattern of links to the other nodes are placed together and form a cluster (group). Two clusters are shown in Figure 2.1b.

In the blockmodeling context, the equivalent nodes are 'shrunken' into new nodes. The blockmodel that is obtained is visualized in Figure 2.1c. The obtained blockmodel has two nodes (shrunk clusters). Here, two types of blocks appear: complete and null. Complete blocks are on the diagonal of the matrix because the nodes from both clusters are internally linked to each other. Off-diagonal blocks refer to the relationships between different clusters. Since the nodes from different clusters are not linked to each other, the off-diagonal blocks are null blocks.

The example represents an ideal case, meaning that there are all possible links in complete blocks and there is no link in the null blocks. However, this is unrealistic for the empirical networks. In

such networks, there are usually some non-links in complete blocks and some links in null blocks (see Figure 2.2). Such links are called errors or inconsistencies.

Figure 2.2: An example of blockmodeling solution with two inconsistencies

Several blockmodeling approaches have emerged to establish the best blockmodel structure. Batagelj et al. (1992) classified the blockmodeling approaches into the classes of indirect or direct blockmodeling. Both approaches are implemented in the computer program Pajek (Batagelj et al., 2004; De Nooy, Mrvar, & Batagelj, 2018) and in the package (for the R programming language) called "blockmodeling" (Žiberna, 2018).

In this dissertation, the terms pre-specified blockmodeling and non-specified blockmodeling are also used. In the case of pre-specified blockmodeling, the whole image matrix is specified while with non-specified blockmodeling only the number of clusters and the allowed block types are specified, and not also the relationships between the clusters.

2.4.1 Indirect blockmodeling

The indirect approach is based on two steps (Doreian et al., 2005). In step one, the dissimilarity matrix is calculated by considering the dissimilarity measure, which is consistent with the selected type of equivalence. Then, considering this dissimilarity matrix, the nodes are clustered by using one of the clustering approaches, e.g. Ward's agglomerative method (Ward, 1963). The indirect blockmodeling approach is not computationally very intensive and can be applied to networks with a higher number of nodes. The deficiency of this approach is that it does not allow the blockmodel structure to be pre-specified (therefore, the approach is seen as exploratory).

There are two well-established types of equivalence (Doreian et al., 2005) (Faust, 1988): structural (Lorrain & White, 1971) and regular (White & Reitz, 1983). Structural equivalence was formally defined by Lorrain & White (1971, p. 63) as follows:

> Objects a, b of category C are structurally equivalent if, for any morphism M and any object x of C, aMx, and xMa if and only if xMb. In other words, a is structurally equivalent to b if a relates

> to every object x of C in exactly the same way as b does. From the point of view of the logic of the structure, a and b are absolutely equivalent, they are substitutable.

The lack of this definition means that that, in a network without loops, the connected nodes cannot occupy the same position. Therefore, an alternative formulation of structural equivalence proposed by Everett et al. is more appropriate (1990, p. 164):

> Suppose G is a labelled graph with vertex set V and edge set E. Then two vertices $a, b \in V$ are structurally equivalent if and only if the permutation $(a\ b)$ produces an automorphism of G'.

On the other hand, the original definition of regular equivalence was provided by White & Reitz (1983, p. 200):

> If $G = (V, R)$ and $\equiv$ is an equivalence relation on V then $\equiv$ is a regular equivalence if and only if for all $a, b, c \in V$, $a \equiv b$ implies: (i) aRc implies there exists $d \in V$ such that bRd and $d \equiv c$; (ii.) cRa implies there exists $d \in V$ such that dRb and $d \equiv c$.

To summarize, structurally equivalent describes nodes linked to other parts of the network in the same way while regular equivalent describes nodes which are linked in the same way with the clusters of equivalent nodes. Each structural equivalence is also regular equivalence. Structural equivalence is probably one of the most often used types of equivalence (Žnidaršič, 2012) while regular equivalence has never achieved widespread use in practice (Žiberna, 2013) mainly because it is rarely present in empirical data (Boyd & Jonas, 2001) and is very sensitive to small changes in the network (Žnidaršič, 2012; Žnidaršič et al., 2012). Concerns have also been voiced regarding regular equivalence's applicability to social theory (Boyd, 2002).

In order to apply indirect blockmodeling by considering structural equivalence, the corrected Euclidian distance can be used as a measure of the dissimilarity among the nodes. It is defined as (Batagelj, Ferligoj, et al., 1992)

$$d(r, i, j, p) = \sqrt{p\left(\left(r_{ii} - r_{jj}\right)^2 + \left(r_{ij} - r_{ji}\right)^2\right) + \sum_{\substack{s=1 \\ s \neq i,j}}^{n} \left(\left(r_{is} - r_{js}\right)^2 + \left(r_{si} - r_{sj}\right)^2\right)} \qquad (2.7)$$

where r is the network in the form of an adjacency matrix, while i and j denote nodes for which the dissimilarity is calculated. Parameter p takes the values 1 or 2 (the value $p = 0$ would means that the correction is not considered).

For regular equivalence, the REGE algorithm (White, 1985) may be used to compute the similarity matrix and the REGDI algorithm (White, 1985) to compute the dissimilarity matrix. For categorical data, the CATREGE algorithm was proposed (Borgatti & Everett, 1993). One can also find other versions of these algorithms (Žiberna, 2008).

The selected type of equivalence imposes the possible block types in a network. It was shown (Batagelj, Doreian, & Ferligoj, 1992; Batagelj, Ferligoj, et al., 1992; Doreian et al., 2005) that in the case of structural equivalence only complete and null blocks exist, while with regular equivalence the regular and empty blocks are possible. When indirect blockmodeling is used, a researcher must select the equivalence type and the number of clusters only. The latter usually entails considering a dendrogram obtained by using am agglomerative hierarchical clustering procedure.

2.4.2 Direct blockmodeling

With generalized blockmodeling, the blockmodel is obtained directly from the network data. This means a researcher does not need to calculate the dissimilarity matrix. To obtain a blockmodel, a local optimization procedure is generally used.

Batagelj et al. (Batagelj, Ferligoj, et al., 1992) obtained the solution by optimizing a criterion function with a relocation algorithm. The criterion function (Batagelj, 1997; Batagelj, Ferligoj, & Doreian, 1998; Doreian, Batagelj, & Ferligoj, 1994) reflects the difference between the ideal blockmodel and the empirical (current) solution. The generalized blockmodeling considers different types of equivalence according to different types of blocks.

When implementing direct blockmodeling in “blockmodeling” (Žiberna, 2018) and also in Pajek (De Nooy et al., 2018), the iterative relocation algorithm works in such a way that it relocates one node from one cluster to another cluster or interchanges two nodes from two different clusters.

In general, compared to indirect blockmodeling, direct blockmodeling produces a solution with a lower or equal criterion function value. With respect to direct blockmodeling, the risk of obtaining a local optimum exists and therefore the algorithm must be repeated several times in the hope of obtaining the global optimum. Its computational complexity is high when a larger number of nodes is analysed.

With direct blockmodeling, the allowed types of blocks (and number of clusters) or the whole image matrix must be pre-specified (equivalence is defined by the set of allowed block types). The first scenario (a researcher specifies the selected block types) is usually seen as partly confirmatory approach while the second scenario (a researcher specifies the whole image matrix) is seen as a confirmatory approach (Doreian et al., 2005).

The blockmodeling approach on empirical networks often reveals the approach is problematic when relatively sparse binary networks are being analysed. Specifically, when regular equivalence is used, all nodes are often classified in the same equivalence group. In contrast, structural equivalence finds only very small complete blocks (Žiberna, 2013). Therefore, Žiberna (2013) proposed several approaches to tackle this problem. The first is to use density blocks (Batagelj, 1997) which have zero inconsistencies if the density of the block is equal to or above γ (the parameter of the density block type) and equal to the number of missing ties to achieve this density otherwise. This approach's drawback is that there is no incentive for these density blocks to have densities above the selected threshold. The second possible approach is to employ structural equivalence with different weights assigned to the null and complete blocks. As Žiberna (2013) reports, the advantage of this approach is that the incentive remains for complete blocks to be as dense as possible. The third proposed approach is to use sum of squares (homogeneity) by structural equivalence. The advantage of this approach is that it differentiates complete blocks of different densities when searching for complete blocks with similarly dense rows and columns.

When the weights are assigned to the links (valued network), binarization is usually done. Since the binarization of networks can cause the loss of a considerable amount of information, Žiberna (2007) suggested two approaches to analysing valued networks, whereby the loss of information is minimized. One approach is the generalization of ideal blocks for binary blockmodeling to valued networks. The three conditions are stated to describe the most common block types: (1) a certain link value must be at least m; (2) a certain link value must be 0; and (3) the f over each row (or column) must be at least m, where f is a function with the property $f(a) \geq \max(a)$ and a is a valued vector. Parameter m represents the minimal value that characterizes the link between two nodes (for complete blocks) in such a way that this tie satisfies the condition of the block. The value of m can be determined either by previous knowledge about the nature of the links and/or based on the distribution of the link values.

Several approaches to blockmodeling signed networks have also been proposed (Doreian, 2008; Doreian & Mrvar, 2009; Wang et al., 2016; Brusco & Doreian, 2019), but are not discussed in this dissertation because the research is limited to non-signed networks.

2.4.3 Stochastics blockmodeling

Compared to the approaches to blockmodeling previously described, stochastic blockmodeling (Anderson, Wasserman, & Faust, 1992; Holland, Laskey, & Leinhardt, 1983; Snijders & Nowicki, 1997) relies on estimating the statistical model from which the data were generated. If simple binary networks without loops are considered, the stochastic blockmodel is defined as

$$P(A|p,b) = \prod_{i<j} p_{b_i,b_j}^{A_{ij}} \left(1 - p_{b_i,b_j}\right)^{1-A_{ij}} \tag{2.8}$$

where A is the adjacency matrix, $p = \{p_{rs}\}$ refers to the probabilities of a link existing between any two nodes belonging to clusters r and s, and b is a vector of entries $b_i \in \{1, \dots, B\}$ specifying the cluster membership of node i (Peixoto, in press).

The fact that a statistical model can be estimated based on empirical data brings many very useful features when analysing empirical data. For example, different hypotheses regarding the global network structure (e.g., the number of clusters or the general fit) can be tested (Bickel & Sarkar, 2016; Lei, 2016; Peixoto, 2015). Since the probability of the existing and non-existing links can be calculated, the model obtained from the data can be used to predict missing and spurious links (Clauset, Moore, & Newman, 2008; Guimerà & Sales-Pardo, 2009; Peixoto, in press).

Different algorithms may be used to estimate the stochastic blockmodel parameters and cluster the nodes of the network. One is the variational Bayes EM algorithm (Latouche, Birmelé, & Ambroise, 2012; B. Yang, Liu, Li, & Zhao, 2017). There are versions of the stochastic blockmodeling in which the nodes can belong to several clusters (Airoldi, Blei, Fienberg, & Xing, 2008).

2.5 Evaluating the errors (inconsistencies) in the blockmodel

This section presents: (i) the approach to quantify the empirical network's fit with the obtained blockmodel; (ii) the approach to quantify the difference between two blockmodels; (iii) and introduces the term "level of errors", which is closely related to generating (partially) randomized

networks because it is defined based on the number of links to be randomly relocated in a network with an ideal blockmodel in order to achieve a randomized network.

2.5.1 Number of inconsistent blocks

The number of inconsistent blocks[12] (Žnidaršič et al., 2012) is used to evaluate how much two blockmodels differ in terms of the number of different blocks. The definition of the number of inconsistent blocks is based on the image matrix of the first blockmodel (e.g. obtained from empirical data) and the image matrix of the second blockmodel (e.g. the ideal or desired blockmodel). Both blockmodels must have the same number of clusters. The number of inconsistent blocks is then defined as the number of different block types by cells in the image matrices.

When the relationships between clusters are not specified in the blockmodeling procedure (e.g. non-specified generalized blockmodeling or direct blockmodeling) and the blockmodels are represented in the form of image matrices, one must consider the order of the rows and columns. More precisely, one has to order the rows and columns of an image matrix in such a way that the difference between the number of inconsistent blocks is minimized. With some blockmodel types, ordering by considering the corresponding density can be sufficient, while in other cases one needs to calculate the number of inconsistent blocks for all possible permutations of rows and columns of one of the image matrices. In the latter case, the number of inconsistent blocks is the minimal difference. Reordering is not necessary in some situations, for example, when one of the blockmodels to compare is cohesive.

2.5.2 Relative fit and mean improvement value

Relative fit (RF) and mean improvement value (MIV) are used to evaluate the extent of inconsistencies in a blockmodel. They are proposed since the values of the criterion function[13]

12 The concept of inconsistent blocks was previously used by Žnidaršič et al. (2012) while studying the impact of non-response on the stability of a blockmodel. They calculated the proportion of incorrect blocks to evaluate the similarity of two blockmodels. The term “incorrect” in their study is used in the same way as “inconsistent” in this study.

13 Usually, the one defined as the total number of inconsistencies with a pre-specified blockmodel is used (Batagelj, 1997; Doreian, Batagelj, & Ferligoj, 1994).

obtained in different blockmodel types (including a different number of clusters) are not generally comparable (e.g. increasing the number of clusters lowers the value of the criterion function in the case of structural equivalence). Compared to the number of inconsistent blocks, RF and MIV are more detailed measures of a given blockmodel's fit to the empirical data and their use holds greatest validity when the presence of a given blockmodel type is confirmed by non-specified blockmodeling (i.e. the number of inconsistent blocks is 0).

Simulations confirm that a higher number of iterations is needed to estimate the maximum value of a criterion function (in the case of a random network under the assumed blockmodel) than for the expected value of a criterion function. Therefore, relative fit (RF) is defined as

$$RF = 1 - \frac{P^m}{\frac{1}{k}\sum_{i=1}^{k} P_i^r} \tag{2.9}$$

where k is the number of randomized networks, P^m is the value of the criterion function of the network of interest (e.g. empirical) and P_i^r is the value of the criterion function of the i-th random network. The mean criterion function value in the case of random networks (which is $\frac{1}{k}\sum_{i=1}^{k} P_i^r$) is estimated by simulations since it depends on many factors (e.g. the algorithm that is implemented for generalized blockmodeling, the density …) and therefore cannot be analytically calculated (also see subsection 2.5.5).

Mean improvement value (MIV), on the other hand, is defined as

$$MIV = 1 - \frac{1}{k}\sum_{i=1}^{k}\left(\frac{P_i^m}{P_i^r}\right) \tag{2.10}$$

where the P_i^m is the value of the criterion function obtained on the i-th empirical network and P_i^r is the value of the criterion function obtained on the i-th randomized empirical network (there are k empirical networks and, for each empirical network, one randomized network is generated).

RF and MIV are comparable among different blockmodel types with different numbers of clusters, yet one must be aware that RF and MIV values obtained on blockmodels with a different number of clusters are not independent. Typically, the values of the indices increase along with the number of clusters. When the whole image matrix is specified in generalized blockmodeling, the values

start decreasing after a certain number of clusters is reached. The values are equal to 1 if the empirical network perfectly fits the assumed blockmodel. Negative values indicate a lower fit than would be expected in the case of random networks.

The expected value of RF is the same as the expected MIV value (on the assumption that all networks of interest are generated from the same model). In this sense, RF may be seen as a special case of MIV. The biggest practical difference between RF and MIV is that RF can be calculated for only one network of interest, while to ensure a reasonable interpretation of MIV several observed networks are needed.

When the global network structure of a network generated from the same model (e.g. when networks are generated using the NEM model) can be obtained several times, MIV can then be reported. As mentioned, the equivalent to this would be reporting the mean RF value. In that case, the number of randomized networks could be equal (when $k = 1$) or higher (when $k > 1$) as in the case of MIV. A higher number of randomized networks decreases variability of the measure. This means that when the mean RF measure is used and $k > 1$, the variability of RF is lower than the variability of MIV.

The proposed relative values of the criterion function could also be referred to as a relative criterion function. However, since criterion functions are generally defined so that lower values indicate a better fit (while lower RF and MIV values indicate a worse fit), different names are used to avoid any misunderstanding.

2.5.3 The level of errors

The level of errors (LE) represents the share of links that have to be randomly relocated from complete to null blocks (in a blockmodel without errors) to obtain the same (expected) density in all blocks. More precisely, LE can be defined by generating totally randomized networks based on an ideal network: k links in complete blocks are randomly chosen and replaced by non-links. At the same time, k non-links in null blocks are randomly chosen and replaced with links. The number of links is relocated in such a way that the overall density in complete blocks and overall density in null blocks are equal. The number of relocated links k is calculated as

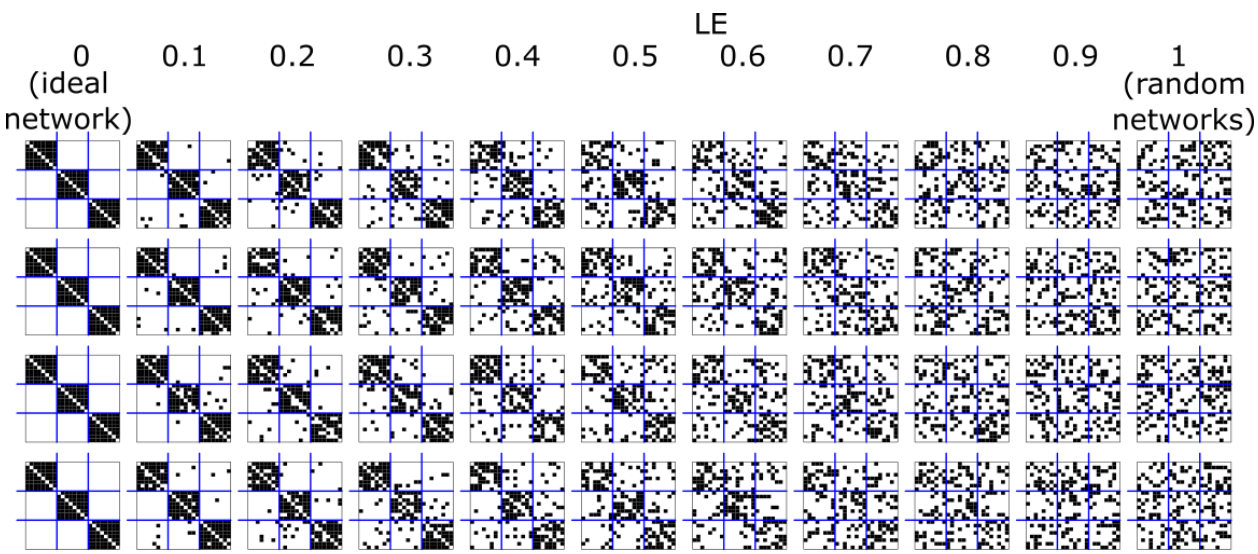
$$k = m - \frac{m^2}{n^2 - n} \tag{2.11}$$

where m is the number of links and n is the number of nodes in a selected blockmodel type. Instead of totally randomised networks, blockmodels with a certain level of errors can be analysed. In such case, when a network with a given blockmodel structure must be generated with a certain level of errors, the number of relocated links is calculated as

$$k = m - \left(\frac{m^2}{n^2 - n}\right) * LE \tag{2.12}$$

where LE can take a value on the interval $[0, 1]$ (0 stands for an ideal network and 1 for a random network). Figure 2.3 visualizes the cohesive blockmodel with different levels of errors. The level of errors increases linearly as the links are moved from complete blocks to null blocks, until the densities of both block types are the same (the level of errors then equals 1). It is impossible to distinguish between blocks in such a network.

Figure 2.3: Cohesive blockmodel with different levels of errors (blue lines separate groups defined based on the ideal network)



2.5.4 Simulation study

The characteristics of the proposed indices are studied in this section by using Monte Carlo simulations. The main questions are: (i) whether one can differentiate between different types of blockmodels based on RF (if the true number of clusters is known or not); (ii) whether one can determine the true number of clusters (if the true blockmodel is known or not); and (iii) whether one can select the true blockmodel and the true number of clusters based on the value of RF.

To this end, networks with different blockmodel types (cohesive, hierarchical, hierarchical-cohesive, transitivity and transitive-cohesive), with different levels of errors, $LE = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, and with different numbers of clusters, $k = \{3, 4, 6, 8\}$ are generated. For each combination of blockmodel type, level of errors and number of clusters, 30 networks (with each consisting of 24 nodes) are generated. In total, $5 * 11 * 4 * 30 = 6600$ networks are generated.

To study the number of inconsistent blocks, direct blockmodeling with a non-specified model is applied to each network generated. Complete and null blocks are assumed. The number of clusters in the blockmodeling procedure is set to the same value as used to generate the networks to be blockmodeled. The number of inconsistent blocks is obtained and analysed for each generated network.

To analyse RF, direct blockmodeling with a pre-specified model is assumed. More specifically, for each network, all considered blockmodel types with all numbers of clusters between 3 and 9 are pre-specified. This means that $4 * 9 = 36$ RFs are obtained for each network. Each RF is calculated by considering 30 random networks.

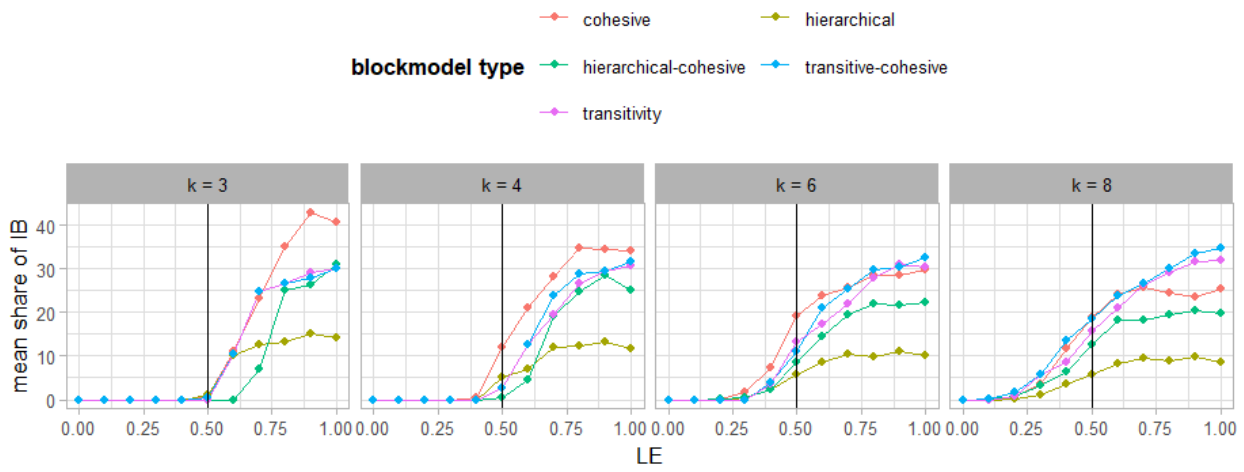
For what LE is it possible to discuss the presence of a given blockmodel type?

Nevertheless, a very clear definition of LE is not trivial when it comes to interpreting high and low levels of errors. An important indicator of which LE is high and which is low is the ability to recognize the presence of the true underlying global network structure – the pre-specified blockmodel. This can be measured by the number of inconsistent blocks. Here, it is assumed that the number of inconsistent blocks must be 0 in order to be able to declare the presence of a given blockmodel type. This is a very strict assumption.

The density of real networks is generally a consequence of the underlying local network mechanisms. Since it can affect a given blockmodel's fit with a given empirical network, the relationship between the network density and an ideal blockmodel must be considered for the purposes of this simulation study. For example, in the case of all of the blockmodels considered, except for a transitive-cohesive one, the density is reduced as the number of clusters increases. In the case of a transitive-cohesive blockmodel, the density increases as the number of clusters rises. This affects the mean number of inconsistent blocks, which increases faster with an increase in LE

in the case of denser networks. Further, the mean number of inconsistent blocks is increasing faster with the number of clusters in the case of denser networks. Therefore, instead of the number of inconsistent blocks, the share of inconsistent blocks is used to increase the comparability of blockmodels with different numbers of clusters. The relationship between the mean share of inconsistent blocks and the LE, blockmodel type, and number of clusters is visualized in Figure 2.4.

Figure 2.4: Impact of the level of errors on the mean share of inconsistent blocks, controlling for the number of clusters and blockmodel type (the vertical line at 0.5 LE indicates the LE at which the true blockmodel type can still be determined with a very high probability)



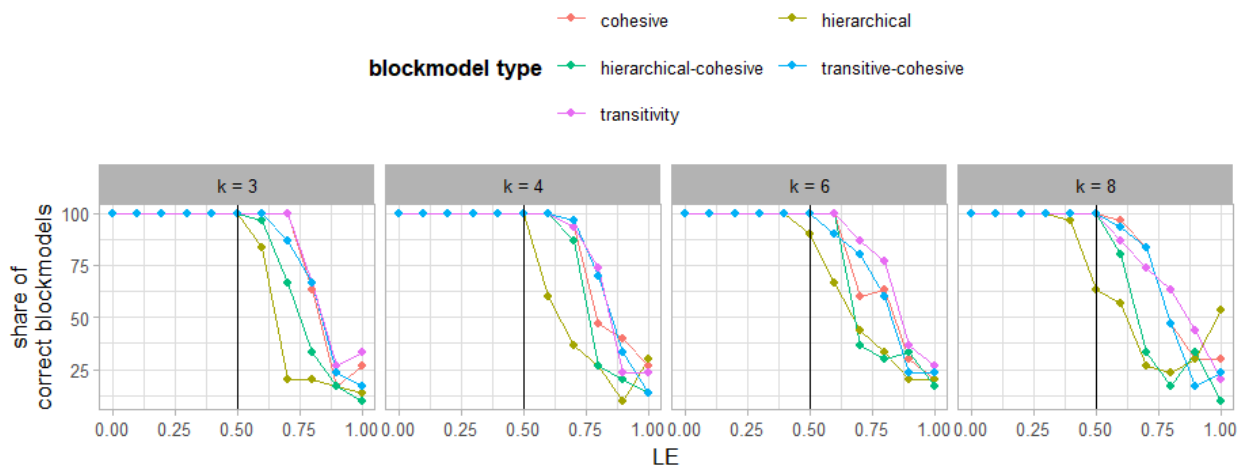
In general, the correct blockmodel (with zero inconsistent blocks) has been identified when there are three clusters (of equal size, each with 8 nodes) and the LE is below 60% or when there are four clusters (with 6 nodes in each) and the LE is below 50%. Networks with a higher number of smaller clusters are rare and unlikely to found in reality although, for such cases, one could determine (by using the algorithm for generalized blockmodeling described earlier) the correct global network structure at a lower LE. Differences may be found among the blockmodel types: the mean share of inconsistent blocks grows the slowest with the hierarchical blockmodel.

Can RF be used to distinguish different blockmodel types (when the true number of clusters is known)?

When the number of clusters in a blockmodel is known, it is possible to use the RF value to select the true blockmodel type. Here, one needs to be aware that density is strongly related to blockmodel type and might therefore be one of the key predictors of the true blockmodel type.

Moreover, a given empirical network may be made up of many different blockmodel types (whose exact number depends on the number of clusters, types of links etc.). Therefore, the set of possible blockmodel types must be based on prior knowledge about the mechanisms underlying tie formation in the network being studied (where it assists to know the underlying social mechanisms).

Figure 2.5: Impact of the level of errors on the share of blockmodels correctly detected, controlling for the number of clusters and blockmodel types (the vertical line at 0.5 LE indicates the LE at which the true blockmodel type can still be determined with a very high probability)



As visualised in Figure 2.5, the share of correctly classified blockmodel types falls as the LE increases while the true number of clusters does not affect the share of correctly determined blockmodel types. Differences occur between different blockmodel types when the level of errors exceeds 50%. This indicates that the true blockmodel type is hard to be recognized at this LE, which can be due to no existing initial blockmodels at such LE or due to the inability to identify the initial blockmodel type (which is assumed to the true blockmodel type) by the selected blockmodeling approach. To sum up, the use of the RF is generally not recommended to determine a blockmodel type when the level of errors is above 50%.

Can RF be used to determine the true number of clusters (when the true blockmodel type is known)?

The number of clusters should be chosen with regard to previous knowledge of the studied networks, including the hypothesized blockmodel type. However, different statistical approaches are proposed to select the most appropriate number of clusters for different blockmodeling

approaches. Many different methods have been proposed within stochastic blockmodeling. They can generally be separated into three classes (Chen & Lei, 2018): (i) the likelihood-related methods; (ii) the Bayesian-inference-related methods; and (iii.) methods related to the information theory.

There is no commonly used method to select the number of clusters for the case of indirect blockmodeling, yet the number of clusters can be chosen by applying many approaches that are regularly used in (hierarchical) cluster analysis. One of these entails determining the number of clusters based on the dendrogram by considering the appropriate type of distance measure (e.g. corrected Euclidean distance), as was done in Cugmas et al. (2016). A deficiency of this approach is that, to some extent, it is subjective.

The approximate number of clusters for the case of direct blockmodeling can be estimated by obtaining the dendrogram based on indirect blockmodeling (as described above) and using this number of clusters within the direct blockmodeling procedure. A more suitable approach is to select the number of clusters by observing the CF values for a different number of clusters, as in Žiberna (2013). This approach may be used with either pre-specified or non-specified blockmodeling. When the whole image matrix is pre-specified, the CF values usually first decrease and later increase along with the number of clusters (a researcher take the number of clusters at which the corresponding CF value is the lowest). In the case of non-specified blockmodeling, the CF values are monotonically decreasing along with the number of clusters and, therefore, a researcher selects the number of clusters at which the CF values decrease more slowly.

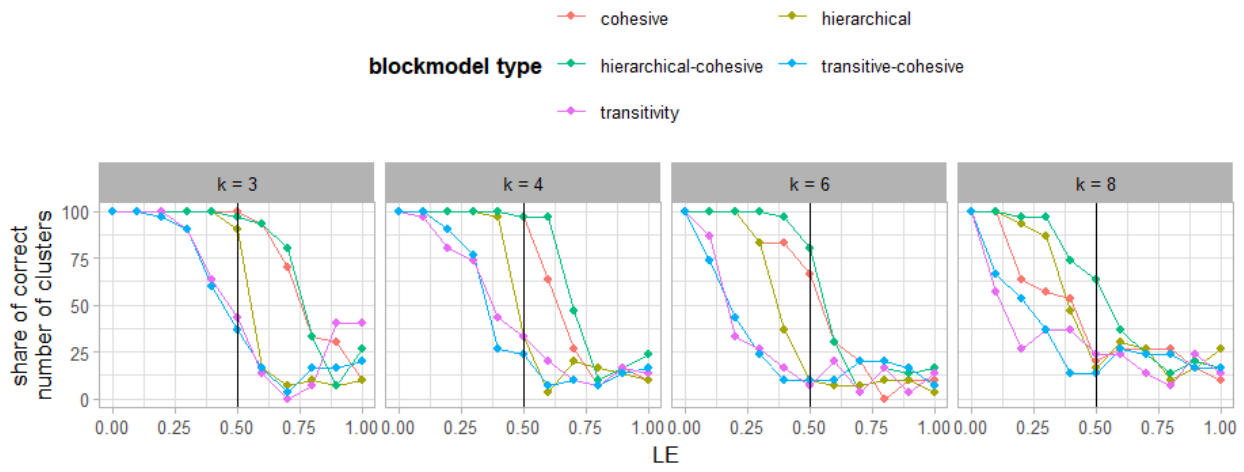
A similar approach can be taken with the RF values where it is assumed that the most appropriate number of clusters is the one at which the RF value is at its highest (because RF is defined such that higher values indicate a better fit). Therefore, this section looks at whether RF values can be used to select the most appropriate number of clusters when using blockmodeling with structural equivalence assumed.

The following factors are considered: the true number of clusters, LE and blockmodel type. As shown in Figure 2.6, the probability the estimated number of clusters is wrong is increasing with the level of errors, as expected. Further, when the true number of clusters is growing, the probability the estimated number of clusters is wrong is also growing. However, when the number

of true clusters and the level of errors are low, the estimated number of clusters is probably correct. As mentioned, the cluster sizes are very small when 6 or 8 clusters are assumed. When examples of networks from real life are analysed, the clusters are probably bigger and the probability of a correctly identified number of clusters is higher.

Knowing the true number of clusters is the sole item of importance since establishing the difference between the true and the estimated numbers of clusters may also prove valuable. Therefore, the variable "difference" is defined as the difference between the estimated number of clusters (based on the RF value) and the true (known) number of clusters (from which random networks with a given level of errors are generated).

Figure 2.6: Impact of the level of errors on the share of the correctly determined number of clusters, controlling for the number of clusters and blockmodel type (the vertical line at 0.5 LE indicates the LE at which the true blockmodel type can still be determined with a very high probability)



The deviation of the estimated number of clusters from the true number (Figure 2.7) is relatively low when the level of errors is equal to or lower than 50%. This is particularly the case with the hierarchical-cohesive blockmodel as well as the cohesive blockmodel. The estimated number of clusters is somewhat less accurate when it comes to transitive blockmodels. In all cases, the number of clusters tends to be overestimated (when the level of errors equals or is below 50%). At a higher level of errors, the deviation from the true number of clusters first increases, then decreases along with the number of clusters. Yet, these results might be overlooked since it is assumed that the initial blockmodel structure disappears with smaller networks containing a

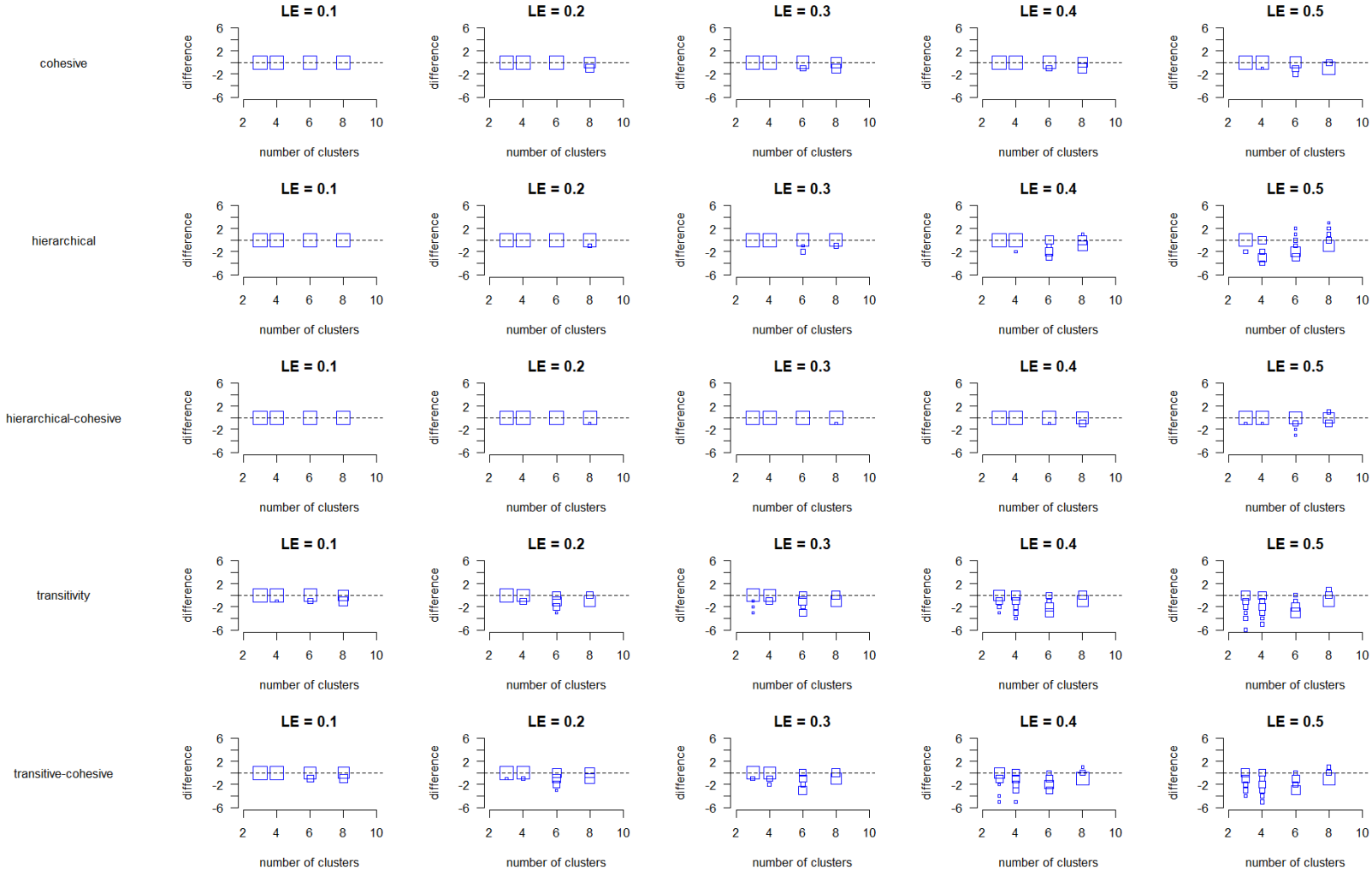
greater number of clusters and relatively high level or errors (see above). Moreover, the smaller difference that emerges when the true number of clusters is higher might be due to the fact that blockmodeling with up to 9 clusters is being applied. This means the maximum difference, e.g. is 1, when the true number of clusters is 8.

RF may be used to estimate the true number of clusters in a blockmodel. When the level of errors is higher, the number of clusters could be overestimated. A more systematic and extended (e.g. by considering clusters of different sizes) study should be conducted on this subject in order to obtain deeper insights into possible applications of RF.

It happens sometimes that a researcher wishes to estimate the number of clusters without knowing (or without paying attention to) the underlying true blockmodel. The relationship between the pre-specified number of clusters in the blockmodeling procedure and RF (by considering the true number of clusters) is visualized in Figure 2.8. The mean value of RF rises considerably along with the pre-specified number of clusters until the true number of clusters is reached[14]. Thereafter, the increase in the mean RF is much smaller. This indicates well that the mean RF value can be used to detect the true number of clusters. The mean RF decreases along with the LE, while differences in the mean RF for a different pre-specified number of clusters are lower when LE has higher values. When the LE exceeds 0.5, the mean RF value cannot be used to estimate the true number of clusters.

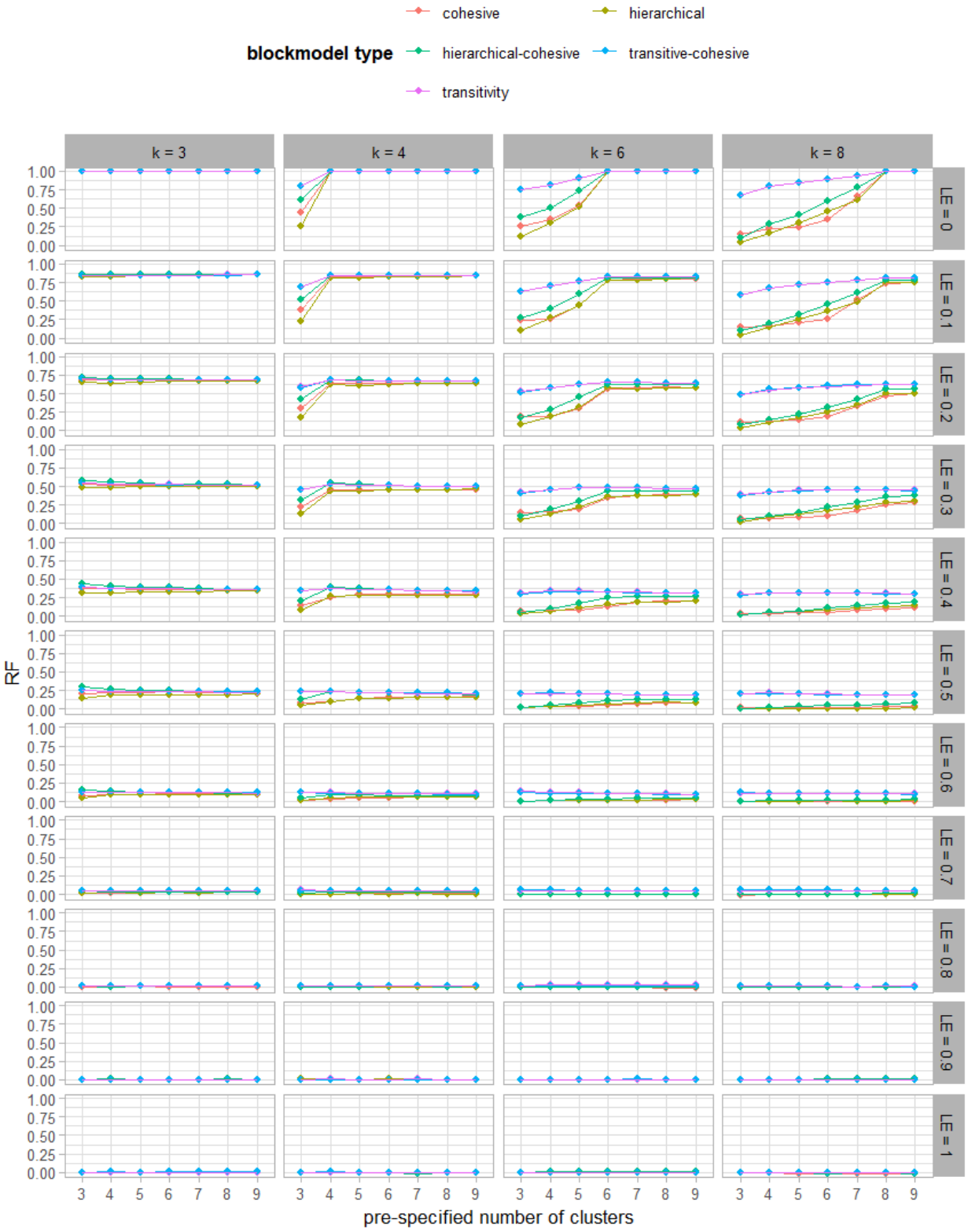
[14] When $LE = 0$, RF reaches 1 at the true number of clusters. The value remains the same ($RF = 1$) when a higher number of clusters is considered. In some cases, the value of RF starts to decrease after the true number of clusters is reached (e.g. the case of a transitive-cohesive blockmodel, $LE = 0.2$, with the true number of clusters equal to 4).

Figure 2.7: Association between the true cluster number and the difference between the true and estimated cluster numbers for different levels of errors and blockmodel type (only the plots for the $LE \leq 0.5$ are shown)



Note: In the scatterplots, the rectangles' sizes are in proportion to the number of networks with the corresponding value of the difference between the true and estimated number of clusters.

Figure 2.8: Impact of the pre-specified cluster number on the mean relative fit value, controlling for the true number of clusters, level of errors, and blockmodel type



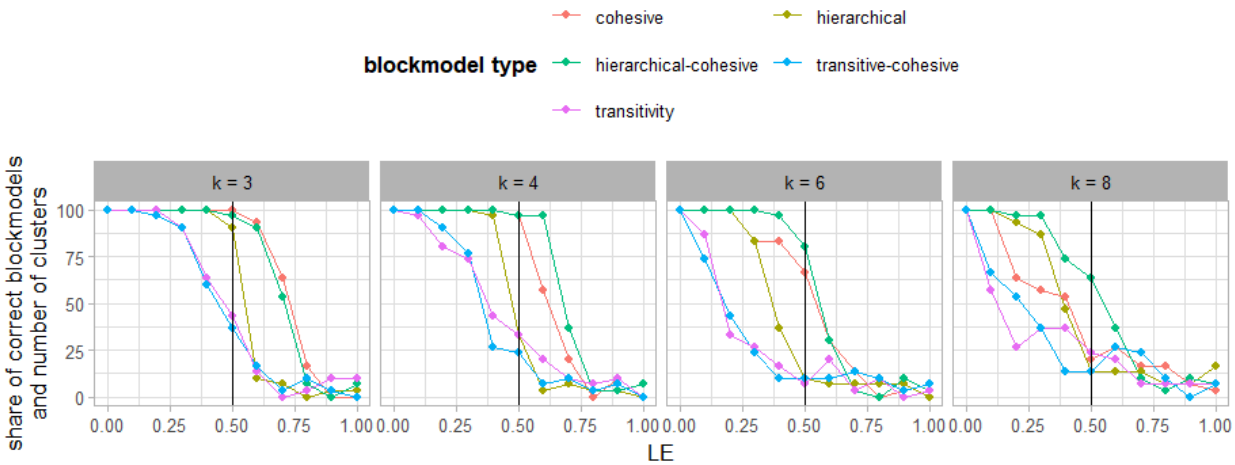
Can RF be used to select the true blockmodel type and true number of clusters?

The previous examples assumed the researcher knows either the true blockmodel type or the true number of clusters. Yet a more realistic scenario is when neither the true blockmodel type nor the true number of clusters is known. In that eventuality, the researcher can use RF to select them, but (as stressed in the previous sections) considering only a limited set of possible blockmodel types or number of clusters should be prioritized since it brings a considerable computational burden.

The correct blockmodel type and the correct number of clusters are assigned to 48% of all networks generated in the current study (or 78% when considering just those with a level of errors below or equal to 60%). As seen in Figure 2.9, all of the factors considered influence the share of correct blockmodel and the number of clusters. The highest odds for a correct guess are seen with the cohesive and hierarchical blockmodels while the lowest odds are found for the transitive blockmodels. The probability the wrong blockmodel type and incorrect number of clusters is selected increases along with the number of clusters and the level of errors.

RF may be used to detect the true blockmodel type and the true number of clusters if the level of errors is relatively low and the cohesive and hierarchical blockmodels are considered.

Figure 2.9: Impact of the level of errors on the share of correctly determined blockmodel type and the number of clusters, controlling for the number of clusters and blockmodel type (the vertical line at 0.5 LE indicates the LE at which the true blockmodel type can still be determined with a very high probability)



2.5.5 A note on randomizing networks

The MIV and RF functions compare the value of the CF obtained for an empirical network with that obtained for randomized networks. Networks can be randomized (i.e. generated under null hypotheses) in several ways (Artzy-Randrup et al., 2004).

Determining the correct way for randomization can prove difficult because mechanisms which generate networks with all possible constraints (until no link in generated random networks is impossible) must be considered. One example of such a constraint is maximum degree. Constraints like this may be due to natural limitations (Dunbar, 1992) or emerge from the chosen data-collecting technique (e.g. the limitation on listing one's three best friends). Not considering all relevant constraints during randomization can lead to overestimated or underestimated MIV and/or RF values.

Two types of constraints (which do not take the characteristics of the nodes into account) and randomizing techniques are discussed below.

Fixed network density

The most common constraint is fixed network density (or fixed expected density of the generated networks). Here, it is assumed that all nodes have equal probabilities of creating links to all the others. When only this constraint is considered, a researcher can generate random networks by using either the $G(n, p)$ model proposed by Gilbert (1959) or the $G(n, m)$ model put forward by Erdős & Rényi (1959) where parameters p and m can be estimated based on an empirical network. Valued (symmetric or asymmetric) networks may be randomized in two ways:

1. by randomly relocating values among cells of the adjacency matrix (values to preserve randomization); or
2. iteratively increasing the values on randomly selected links until the sum of all the values in randomized networks reach the sum of all of the initial network's values (density to preserve randomization).

Fixed degree

When the assumption of fixed (or expected) density is not sufficient or inappropriate, one can state another constraint that relates to the fixed degree of each unit. In this case, the weights (or simply links in binary networks) must be generated such so that the row sums and column sums are equal

in the empirical and randomized networks (degree to preserve randomization). This approach can be applied when the nodes have limited opportunities for establishing links[15], when the degree is fixed by the data-collection technique etc.

Several approaches are proposed to randomize binary networks by keeping the marginal frequencies fixed (Miklós & Podani, 2004; Sanderson, Moulton, & Selfridge, 1998). Rao and Bandyopadhyay (1996) put forward a Markov Chain Monte Carlo method for generating nearly random binary networks with fixed densities while many of the other proposed approaches are based on a swap of checkboard nodes where a checkboard unit is a $2 * 2$ sub-matrix defined as $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. In an iterative process, the two columns and two rows are randomly selected. If the submatrix on the interception of the selected rows and columns is a checkboard unit, this checkboard configuration is then replaced with another one. Since this approach tends to generate networks which are not completely random, a new *trial-swap* algorithm was proposed by Miklós & Podani (2004). An approach for generating randomized networks based on random walks was also proposed (Zaman & Simberloff, 2002), but the algorithm is very computational intensive.

The alternative to the approaches mentioned above is the Curveball algorithm (Strona, Nappo, Boccacci, Fattorini, & San-Miguel-Ayanz, 2014), which is computationally less demanding and produces unbiased networks. The algorithm identifies in a single step all possible swaps between a pair of matrix rows (or columns) and performs them all with equal probability.

Another approach to randomization that also often used is the relabeling algorithm. Here, the names of the nodes are randomly mixed (or the order of the rows and columns is randomized). This approach does not affect the global network structure and hence cannot be used by way of randomization for the MIV and RF functions. Yet, the approach can be applied to study the relationship between the nodes' attributes and their position in the network.

[15] For example, in the given empirically observed interactional network collected in a preschool environment the weights represent the number of interactions between two children. If the network is based on several observations in time, it may be that not all children are present in class for all of the time (e.g. due to illness) and therefore they are considered as having limited opportunities to create links to others.

An empirical example

Let us consider an interactional network collected among preschool children. The network is symmetric and valued (the network is analysed in Chapter 5). Three types of randomization are applied to this network (Figure 2.10). The row and column sums are the same as in the empirical network subjected to degree-preserving randomization while, when values-preserving randomization is performed, the values remain fixed, but not the row and column sums. In the case of the density preserving randomization, the sum of all the values remains the same, but the row and column sums and the individual link values change.

Figure 2.10: Different types of randomization

	node 1	node 2	node 3	node 4	node 5	node 6	node 7	node 8	node 9	node 10	node 11	sum
node 1		9	3	3		2	1	1	1		4	24
node 2	9					2		4	1	1	1	18
node 3	3			7	2	1	1	7	3	4	1	29
node 4	3		7		1	2	1	5	1	3		23
node 5			2	1		2	1		1	1	3	11
node 6	2	2	1	2	2		2		1	3	2	17
node 7	1		1	1	1	2		3	2	1		12
node 8	1	4	7	5			3			4		24
node 9	1	1	3	1	1	1	2			5	7	22
node 10		1	4	3	1	3	1	4	5		1	23
node 11	4	1	1		3	2			7	1		19
sum	24	18	29	23	11	17	12	24	22	23	19	222

(a) original network

	node 1	node 2	node 3	node 4	node 5	node 6	node 7	node 8	node 9	node 10	node 11	sum
node 1		3	2	3	2	2	2	2	1	6	1	24
node 2	3		1	4	3	2		3		2		18
node 3	2	1		3	3	3	1	5	3	4	4	29
node 4	3	4	3		1	2	3	1	1	4	1	23
node 5	2	3	3	1							2	11
node 6	2	2	3	2				3	3	1	1	17
node 7	2		1	3				1	3		2	12
node 8	2	3	5	1		3	1		4	4	1	24
node 9	1		3	1		3	3	4		1	6	22
node 10	6	2	4	4		1		4	1		1	23
node 11	1		4	1	2	1	2	1	6	1		19
sum	24	18	29	23	11	17	12	24	22	23	19	222

(b) degree preserving randomization

	node 1	node 2	node 3	node 4	node 5	node 6	node 7	node 8	node 9	node 10	node 11	sum
node 1			1	5	4	1		1	2	9	1	24
node 2			3	3		2	2	7	1	1	4	23
node 3	1	3		2	4	1		2	1	1	1	16
node 4	5	3	2		7	3	1	1			3	25
node 5	4		4	7				2	1	2		20
node 6	1	2	1	3			3	7	5	1	1	24
node 7		2		1		3		3	1		3	13
node 8	1	7	2	1	2	7	3		2		4	29
node 9	2	1	1		1	5	1	2		1	1	15
node 10	9	1	1		2	1			1			15
node 11	1	4	1	3		1	3	4	1			18
sum	24	23	16	25	20	24	13	29	15	15	18	222

(c) values preserving randomization

	node 1	node 2	node 3	node 4	node 5	node 6	node 7	node 8	node 9	node 10	node 11	sum
node 1			5	3	1	2		1	2	2	2	18
node 2			2	6	2	1	3		1	5	1	21
node 3	5	2		1	3	1	1		4	2	1	20
node 4	3	6	1		2	1	2	2		4	5	26
node 5	1	2	3	2		4	1	2	5	2	2	24
node 6	2	1	1	1	4		1	2	3		3	18
node 7		3	1	2	1	1		1	2	1	3	15
node 8	1			2	2	2	1				1	9
node 9	2	1	4		5	3	2			2	4	23
node 10	2	5	2	4	2		1		2		4	22
node 11	2	1	1	5	2	3	3	1	4	4		26
sum	18	21	20	26	24	18	15	9	23	22	26	222

(d) density preserving randomization

3 Generating networks with different blockmodels by considering triads

This chapter addresses the first research question of this dissertation which is concerned with generating different global network structures by considering the selected local network structures. Local network structures are defined as subgraphs, which can be of different sizes. In this study, the triadic census (the collection of all possible networks of size three, as visualised in Figure 1.2) (Davis & Leinhardt, 1967; Holland & Leinhardt, 1970) is considered. Even though the triadic census is well studied within different network types (Faust, 2006, 2007, 2010), no attention was paid to the dependencies between the triadic census and different global network structures, operationalized by the types of blockmodels. This is especially important when thinking about the factors that drive a network towards a certain global structure in social mechanisms terms (Hedström & Swedberg, 1996).

Therefore, the main goal of this chapter is to study whether it is possible to generate networks with a given blockmodel structure, taking only different types of triads into account. The primary objective is further elaborated: is it possible to generate networks with a given blockmodel type while considering only allowed or only forbidden triad types? The classification of allowed and forbidden types of triads is determined for each blockmodel type separately. Allowed types of triads are those with frequencies higher than zero in an ideal blockmodel structure. On the other hand, forbidden triad types are those with frequencies equal to zero in an ideal blockmodel. The sets of allowed/forbidden triad types are then reduced following comparisons of the different blockmodel types with respect to different levels of errors in the network according to the ideal blockmodel being considered.

The sets of all triad types, the sets of allowed/forbidden triad types and the sets of reduced (called "selected") allowed and forbidden triad types are then used to generate networks with a given blockmodel structure. For a blockmodel type which cannot be generated successfully based only on the triad type, some other local network structures are considered.

Beside the different types of triads, other subgraph types of a size higher than three can be used to generate networks with a given blockmodel. Here, different triad types are considered chiefly because they are the smallest sociological unit from which the dynamic of a multi-person relationship can be observed (Davis & Leinhardt, 1967).

Various types of algorithms can be applied to generate networks with a given blockmodel when considering only different triad types. In this study, the proposed Relocating Links algorithm (RL algorithm) and the MCMC algorithm are used. If the structures generated by the selected set of triads obtained by both algorithms are very similar and close to the assumed ideal structure, one may conclude it is possible to generate networks with the assumed blockmodel structure by only considering the selected triad types. On the other hand, if the generated networks are not similar and consistent with the assumed blockmodel structure, one must consider whether this is a consequence of the algorithms' characteristics (see Section 3.3) or that the set of selected local structures is insufficient to generate this specific blockmodel.

In this study, it is assumed that the assignment of a node to a cluster is unknown. Considering information on the cluster assignment would require a different methodological approach. It is also assumed that the nodes' characteristics are not known. Kogut (2000) reported that a certain structure's emergence in a network is often the consequence of rules that generate self-organization dynamics. These rules do not need to be technological in origin but can also reflect institutional or cultural norms, and are also deeply embedded in the social identity of individuals, meaning they are often invisible or unknown when an empirical social network is being studied.

Moreover, the study does not address the question of how the specific selected social mechanisms affect the emergence of given blockmodel types. Instead, it examines the possibility that selected global network structures (blockmodels) are a consequence of local (social) mechanisms. For example, when a given global network structure is strongly characterized by a very high number of transitive triads (or when a given blockmodel type can emerge due to nodes' tending towards the creation of transitive links), one can discuss several social mechanisms which are related to transitive triads. In this regard, one must note that different social mechanisms can lead to a specific social output and a specific social mechanism can lead to different social outputs (Hedström & Ylikoski, 2010).

This chapter is organized in the following way: first, the blockmodel types being considered are listed (Section 3.1) and then a classification of the different triad types is proposed. The classification of allowed/forbidden triad types is given, followed by a further selection based on the proposed A-measure (Section 3.2). The algorithms for generating the networks are then

presented in Section 3.3, followed by the research results (Section 3.5) and conclusions (Section 3.6).

3.1 Global network structures

Several of the well-known and studied blockmodel types are presented in more detail in Section 1.3 and other sections. Here, the most common blockmodel types are considered: the cohesive blockmodel type, the symmetric and asymmetric core-periphery blockmodel type, the transitivity blockmodel, transitive-cohesive blockmodel, hierarchical blockmodel and hierarchical-cohesive blockmodel. The three clusters are set in the case of all blockmodels, expect in the core-periphery blockmodel which by definition consists of two clusters.

3.2 Choosing triads for different types of ideal blockmodels

There are 16 different triad types in the case of directed networks (see Figure 1.2). When generating networks with a specific type of a blockmodel (according to different triad types), all triad types or only a subset of all of them can be examined. Considering only a subset of all possible triad types is particularly important when generating networks with the RL algorithm. This is because the distribution of triads must be known in advance for each type of ideal blockmodel separately. It should be pointed out that the distributions of triads can vary among the same type of blockmodel with a different number of clusters or different number of nodes per cluster.

Since the number of different triads is also affected by the network density (Faust, 2006), the value of the A-measure can be used to select a smaller number of different triad types (see subsection 3.2.2) needed to generate networks with a selected blockmodel type. The A-measure is defined as the ratio between the number of triads of a certain type in an ideal blockmodel and the mean number of such triads in a totally randomized network of the same density (see subsection 2.5.3 for more information on generating totally randomized networks and networks with a given level of errors).

The classifications of allowed/forbidden triad types for different blockmodel types are presented in the next section, followed by classifications of selected allowed/selected forbidden triad types based on values of the A-measure.

3.2.1 Allowed and forbidden triad types

Triad types can be classified in the set of allowed or the set of forbidden triad types for each blockmodel type based on counts of the triad types in an ideal blockmodel. Triad types with a count equal to zero are said to be forbidden in a given blockmodel and are thus classified in the set of forbidden triad types (for a given blockmodel). All the other triad types are classified in the set of allowed triad types. This classification is essential for the MCMC algorithm because it determines the values of the appropriate parameters in the ERGM model (see Section 3.3.2).

Table 3.1: A-measure values and the classification of allowed/forbidden triad types for different blockmodel types

	Cohesive	Asymmetric core-periphery	Symmetric core-periphery	Hierarchical	Hierarchical-cohesive	Transitivity	Transitive-cohesive
003	2.3	7.1	7.2	1.5	0.0	1.1	0.0
300	96.3	7.5	2.7	0.0	3.7	0.0	1.2
120D	0.0	8.2	0.0	0.0	4.1	0.0	5.1
120U	0.0	0.0	0.0	0.0	4.1	0.0	5.1
102	10.2	0.0	0.0	0.0	5.8	0.0	0.0
021C	0.0	0.0	0.0	2.2	3.1	0.0	0.0
021U	0.0	8.2	0.0	4.0	0.0	5.1	0.0
021D	0.0	0.0	0.0	4.0	0.0	5.1	0.0
030T	0.0	0.0	0.0	0.0	0.0	3.5	3.5
201	0.0	0.0	6.6	0.0	0.0	0.0	0.0
120C	0.0	0.0	0.0	0.0	0.0	0.0	0.0
111D	0.0	0.0	0.0	0.0	0.0	0.0	0.0
030C	0.0	0.0	0.0	0.0	0.0	0.0	0.0
210	0.0	0.0	0.0	0.0	0.0	0.0	0.0
012	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note: Values greater than zero denote allowed types of triads, while values equalling zero denote forbidden types of triads; grey colours denote selected triad types.

Reducing the number of triad types used to generate networks with a given blockmodel type can bring several benefits. For example, it can help to identify the main (e.g. social) mechanisms that cause a given blockmodel structure to be formed.

In addition, there are practical reasons that vary according to the algorithm being used. For the RL algorithm (see subsection 3.3.1), the reduction to only forbidden triad types (or a subset of forbidden triad types) is especially appealing since it does not require knowledge of the exact distribution of triad types in the ideal network structure (as this algorithm otherwise requires) because the frequency of all forbidden triad types is 0.

The frequencies of different forbidden triad types are also not affected by the sizes and number of clusters as these frequencies are always 0. This means that, when generating networks by considering only the forbidden triad types, this information is not taken into account, which may be either desired or not. On the other hand, the frequencies of all allowed triad types contain all the information that is included in all (allowed and forbidden) triad types.

For the MCMC algorithm, these issues are not relevant because the exact distribution of triad types is never taken into account when setting the parameter values. However, the MCMC algorithm is affected by multicollinearity, which can be reduced by selecting only a subset of all triad types. Given the characteristic of this algorithm, it is best to select only a small number of relatively different triad types.

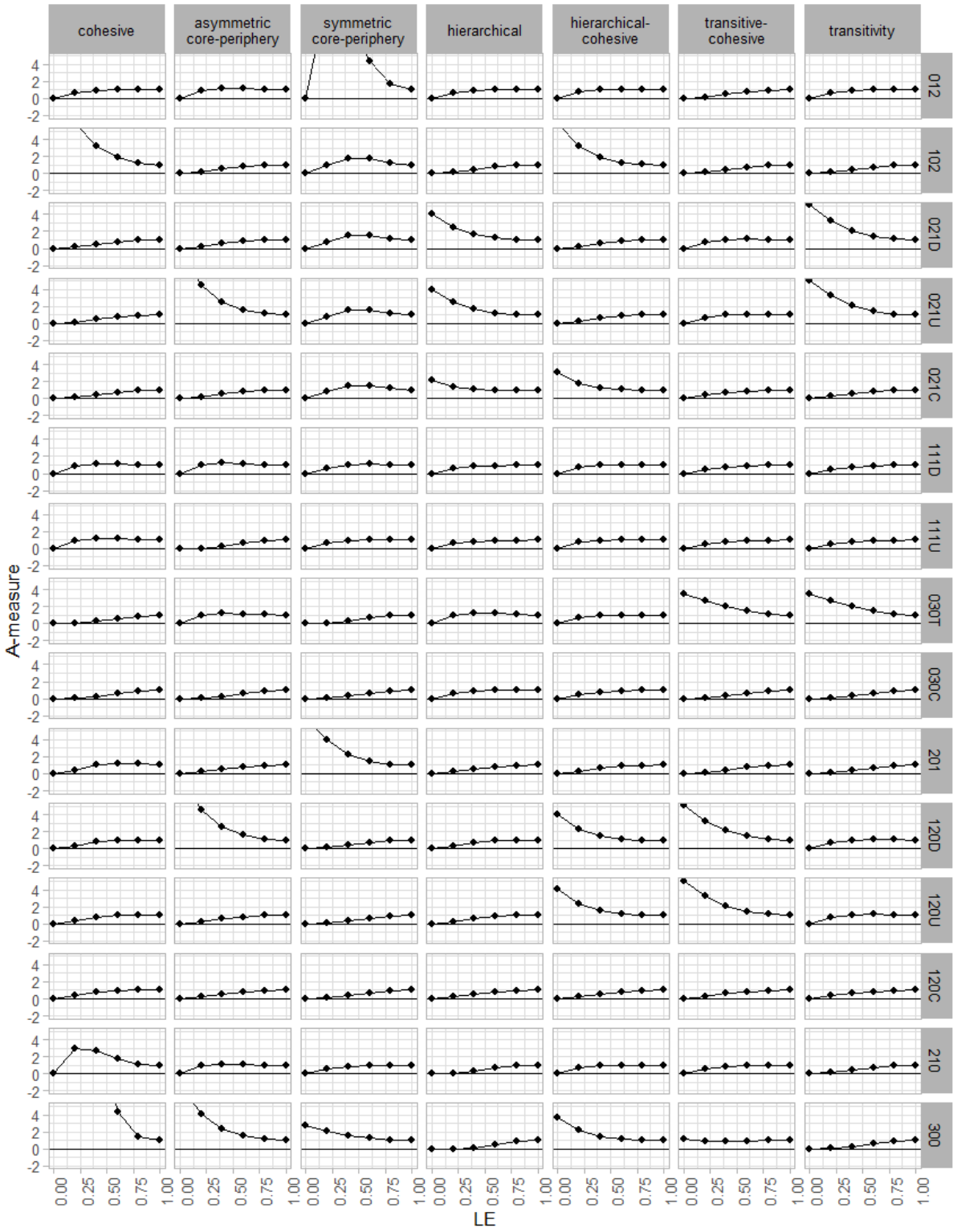
3.2.2 Selecting subsets of allowed and forbidden triad types

The sets of allowed/forbidden triad types can be reduced to selected allowed/selected forbidden triad types. There are several ways for selecting the subset of triad types. In this study, the choice of triad types is based on their sensitivity to different levels of errors (from 0.2 to 1 with step 0.2; see subsection 2.5.3 for the definition of LE), where the sensitivity is evaluated through the value of the A-measure (10,000 random networks were generated for each LE).

The A-measure values are presented in Table 3.1. Values greater than 1 indicate triad types that are more likely to occur in an ideal network structure than would be expected in randomized networks. Such are complete subgraphs of size three (a triad of type 300) in a cohesive blockmodel.

When the A-measure value is close to 1, the number of triads in the case of an ideal network structure is close to the number of triads in totally randomized networks. This may indicate that their occurrence is mainly due to the density rather than the type of blockmodel. In the present case, all values corresponding to the allowed triad types are higher than 1. The A-measure values in the cells without any number in Table 3.1 equal zero and therefore denote forbidden triads.

Figure 3.1: A-measure values for different levels of errors and different blockmodel types



Note: Values higher than 4 are truncated to save space; values above 4 are considered as high.

The most common (and uncommon) triads for each blockmodel type can also be recognised by their sensitivity to different levels of errors. The idea is as follows: the most important triads are those with the highest absolute A-measure values for all levels of errors and with as close to a linear trend as possible through all levels of errors, indicating that a certain triad is not greatly affected by the LE (see Figure 3.1 for the visualised relationship between the LE and A-measure for different blockmodel types and triad types).

For some triads, the A-measure values are nearly constant for all levels of errors. Such a triad is triad type 300 in the case of a transitive-cohesive blockmodel. The value of the A-measure for these triad types is not associated with the LE and this triad type is therefore not relevant for this blockmodel type. On the other hand, for many types of triads a sharp change in the A-measure value at a certain LE is common. For example, in the case of a hierarchical blockmodel, the value of the A-measure for triads of types 012, 111D, 111U, 030T, 030C and 210 is zero in the case of an ideal network while it approaches 1 at very low levels of errors (i.e. between 0.2 and 0.4) and then remains constant. Values for some types of triads first increase very fast at low levels of errors and then decrease at higher levels of errors. One example is the number of complete subgraphs of size three with one missing link in a cohesive blockmodel.

The values of the A-measure for some types of triads are increasing or decreasing nearly linearly with the LE. These types of triads can be seen as triads that should be considered when generating networks with a given blockmodel structure.

Still, these types of triads can be further differentiated. For example, there are many types of triads with similar A-measure values for different levels of errors within some types of blockmodels. This could indicate that certain types of triads are defined similarly and are therefore not needed when generating networks with a given blockmodel structure.

Some types of triads which are strongly influenced by the level of errors at low levels and less influenced by the level of errors at a high level of errors (and vice versa) could also be chosen. In this case, it may happen that one should choose different types of triads for networks with higher levels of errors and those with lower levels.

The study does not focus on how to select the smallest sufficient subset of triad types to generate networks with a given blockmodel, and therefore we do not imply our procedure is the best

possible one. Instead, the aim is to test one can generate networks with a given blockmodel by considering a smaller set of triad types.

The selected triad types are shown in grey in Table 3.1. It is shown that only a few triad types are allowed for each blockmodel type. Almost all of these allowed types of triads are selected (in the case of each blockmodel type). The only allowed triad types that were not selected are triad type 021C in the case of a hierarchical blockmodel and triad type 300 in the case of a transitive-cohesive. On the other hand, for some blockmodel types only a small number of all forbidden triad types is selected, e.g. in the case of an asymmetric core-periphery only one, and in the case of a cohesive blockmodel only two.

3.3 Algorithms for generating networks

As described in Section 1.5, different statistical models have been developed (Toivonen et al., 2009) to explain the impact of local network mechanisms on global network structures or to characterize the global network structures in terms of local network structures. Two similar algorithms are used in this study: the RL algorithm and the MCMC algorithm implemented in the "ergm" package (Hunter et al., 2008) for the R programming language. They both assume that the nodes tend to create such a constellation of links that results in a desirable distribution of subgraphs of size three or other network characteristics. Both approaches are described and compared in more depth in the following sections.

3.3.1 Generating networks with the Relocating Links algorithm (RL algorithm)

The RL algorithm (see Algorithm 3.1), which is based on the approach of relocating links, requires that all considered local network statistics for an ideal network be represented by the vector $\mathfrak{I}$. The number of elements g of this vector equals the number of local network statistics considered. The numbers of different types of triads are considered here, but other local network statistics could also be chosen. The distribution of all or only a subset of all triad types can be given (for forbidden triad types, corresponding values of $\mathfrak{I}$ equal zero). Beside $\mathfrak{I}$, the initial random network Y_r has to be given. The density of the network does not change over the iterations and therefore the density

of the initial network must be in line with the desired global network structure[16]. Before the iterative procedure starts, Y_r is saved as a new network Y_{new}.

The iterative procedure is repeated many times. Upon each iteration, a pair of linked nodes i and j and a pair of unlinked nodes k and l are randomly chosen. Then, the link between i and j is dissolved and a link between k and l is established. The modified network is saved as the proposed network Y_p. From Y_p, the number of each triad type considered $\mathfrak{I}^p$ is calculated. The proposed network is saved as Y_{new} (the new network) if the CR ratio is greater than 1. The CR ratio is defined as

$$CR(\mathfrak{I}, \mathfrak{I}_i^p, \mathfrak{I}_i^{new}) = \frac{\sum_{i=1}^{g}\left(\mathfrak{I}_i^p - \mathfrak{I}_i\right)^2}{\sum_{i=1}^{g}(\mathfrak{I}_i^{new} - \mathfrak{I}_i)^2} \tag{3.1}$$

Then, the new iteration is performed and, after many iterations, the last Y_{new} is the final solution. Besides the Y_{new}, the values of CR can be saved and further analysed.

Algorithm 3.1: The Relocating Links algorithm

Require: $\mathfrak{T}$ ▷ $\mathfrak{T}$ denotes the distribution of local network statistics in an ideal network
Require: Y_r ▷ Y_r denotes a random network
Require: M ▷ M denotes the number of iterations

```
Y_new ← Y_r
Y_p ← Y_r
for m in 1 : M do
    randomly select a tie y_{i,j} in Y_new
    randomly select a non-tie y_{k,l} in Y_new
    transform a tie y_{i,j} to a non-tie in Y_p
    transform a non-tie y_{k,l} to a tie in Y_p
    if CR > 1 then
        Y_new ← Y_p
    else
        Y_p ← Y_new
    end if
end for
return Y_new
```

[16] The blockmodel type and number of clusters are closely related to the network density.

Compared to the MCMC algorithm introduced in the next subsection, the RL algorithm is deterministic since a link is only allocated if the distribution of the triads of the proposed network is closer to the distribution of the triads in the case of an ideal blockmodel. This may result in lower variability of the global network structure of networks generated when the RL is used since, compared to the MCMC algorithm, RL strive to generate networks with the exact number of the selected types of triads. However, the risk of a local optimum exists, which could be avoided by further improving the algorithm. Moreover, RL is computationally very intensive: a higher number of iterations is required, especially in the case of denser networks.

3.3.2 Generating networks with the MCMC algorithm

To generate networks, the Metropolis-Hastings algorithm was used, as implemented in the "ergm" package (see Section 2.2. for more details on MCMC algorithms in ERGM). The benefit of this approach is that by selecting a suitable proposal distribution one can place appropriate restrictions on the network, e.g. fixed density.

The definition of the probability of accepting the proposed network for the Metropolis-Hastings algorithm is similar to the definition of the CR: both compare the proposed network with the current one through the values of the proposed statistics. The elements of $\mathfrak{I}$ are the exact values from the network with the ideal global network structure (where the number of nodes plays a significant role) while the values of θ are regression coefficients and are, therefore, less directly related to the global network structure. In the case of the RL algorithm, a link is relocated (i.e., one link is dissolved and one is established) always when CR is greater than 1 and never when it is below 1. The RL algorithm could be defined in line with the logic behind the Metropolis-Hastings algorithm. In this case, the link is relocated if the value of CR is higher than 1. If the value of CR is below 1, the link is relocated with a given probability. This approach would incorporate an extra level of randomness in the generating process.

As described in Section 2.2, the method most often used to estimate parameters θ is MCMC-MLE, which can be computationally hard to estimate. In this study, the parameters can be estimated based on networks with a given blockmodel without or with only very low levels of errors. When using this approach, in many cases the estimation algorithm does not converge, probably due to

the high level of multicollinearity of the triads. In addition, from a researcher's point of view, estimating the values of all parameters for each blockmodel type would be very difficult.

Instead, the values of the ERGM parameters θ are arbitrarily set to 2 (allowed) or -2 (forbidden). It has been shown (see subsection 3.2) that some triad types are much more likely to appear in an ideal network (compared to a random network). By setting all the parameters' values to 2 or -2, we essentially assume that all types of allowed triads have the same importance (and similar for all forbidden triad types). Such a setting is critical when all types of triads are included in the model and result in a relatively unstable model, particularly when the density is not fixed.

All types of triads are used by considering the two approaches of considering the number of links: (i) the number of links is fixed (to the same value as in ideal networks) and (ii) the number of links is free (with the density being the variable). With the latter, the value of parameter edge is set to such a value that the mean density of 30 generated networks lies within the ideal-density interval ± 0.05.

3.4 Simulation design

To address the objective of this study, the degree to which networks generated using the previously described algorithms (the number of iterations is set to 6,000 in the RL algorithm and to 10,000 in the MCMC algorithm) match the assumed blockmodel type is assessed. With each algorithm, 50 networks (each with 24 nodes) with a given blockmodel structure are generated for each selected set of triads. Each generated network is randomized.

Pre-specified direct blockmodeling is applied to model networks and randomized networks where the number of clusters is set as in the ideal networks (to two or three clusters, see Figure 1.3). The partition is determined by 100 restarts of the blockmodeling algorithm[17]. For each generated network, the minimal value of the blockmodeling criterion function is preserved.

Here, it should be highlighted that there may be bias in the values of the criterion function, where the networks are generated by the RL algorithm and all allowed triad types are considered. This is because the information on the number and sizes of the clusters is embedded in the frequencies of

[17] The "blockmodeling" (Žiberna, 2018) package implemented in R is used for the blockmodeling.

different allowed triad types when using the RL algorithm. Yet, this is not the case when the MCMC algorithm is used and/or other subsets of triads are considered.

As the criterion function is not generally comparable for different blockmodels, the MIV is used (see Section 2.5.2) for each type of blockmodel and each combination of triad types examined. The corresponding values of the MIV are visualised in Figure 3.2, Figure 3.3 and Figure 3.5.

3.5 Results

This section is organized in several parts. First, to evaluate whether one can generate networks with a given blockmodel by considering different triad types, the global network structures of the networks generated with the RL algorithm and the MCMC algorithm (fixed and non-fixed density) are evaluated. For each algorithm, the networks generated by considering different sets of triad types are compared. Then, considerations of certain additional local network structures are presented in the event triads do not generate the networks with the expected blockmodel. Finally, a general statement is made concerning the generating of networks using triads.

3.5.1 Networks generated with the RL algorithm

When the RL algorithm is used to generate networks and all triad types are considered, the overall MIV is around 72%, which is more than for all other sets of triads studied (Figure 3.2)[18]. On the other hand, the MIV corresponding to the networks generated based only on all forbidden triads or all allowed triads is slightly lower or the same. What is outstanding is the symmetric core-periphery with the lowest MIV varying between 11% and 32% among the different models (all, all allowed, or all forbidden triad types). As has been emphasized, when the network is very dense the RL algorithm is less effective at finding the right link to relocate. This is expressed in the very small peripheral part in the case of a symmetric core-periphery blockmodel.

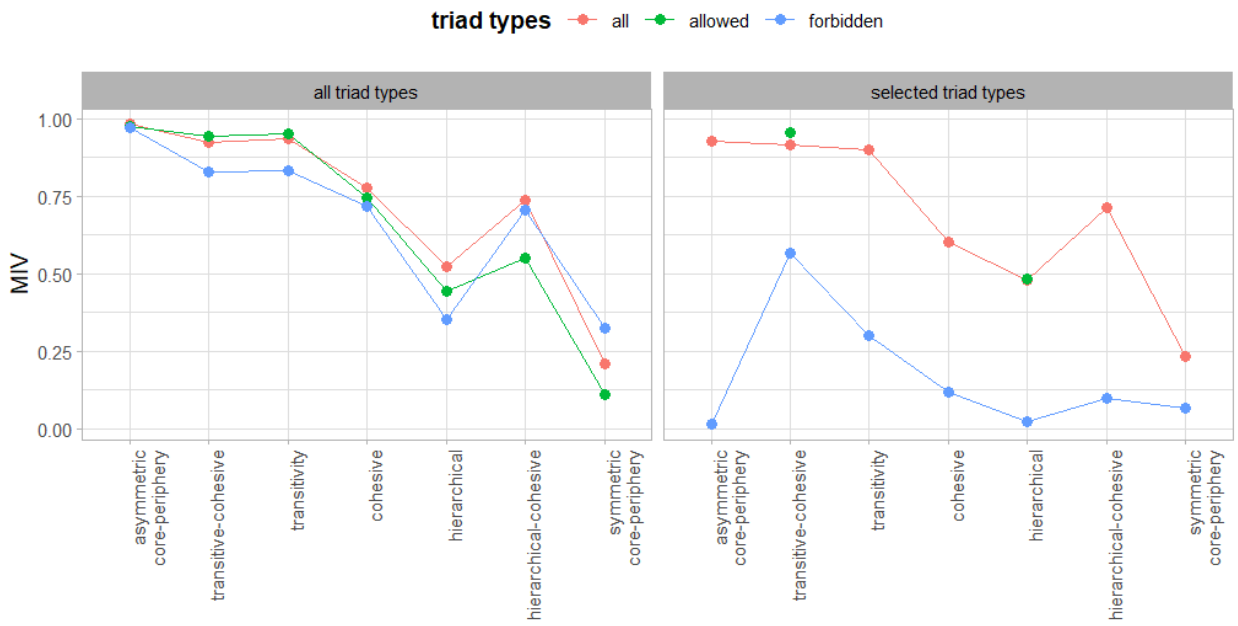
The MIVs are typically lower when all forbidden triad types are considered. The MIVs corresponding to the cohesive blockmodel are very similar, yet the structure of the blockmodels

[18] The distributions of improvement values (the ratio between the value of a criterion function obtained on a generated network and the value of a criterion function obtained on a randomized network), which are used to determine the MIV, are given in Appendix B.

so generated is different when only the set of forbidden triad types is considered (the cluster sizes are more variable).

When comparing the different blockmodel types, the highest MIV is observed with an asymmetric core-periphery blockmodel (98% when all allowed or all forbidden types of triads are considered), a transitivity blockmodel (95% when all allowed types of triads are considered; 94% when all types of triads are considered) and a transitive-cohesive blockmodel (94% when all allowed triad types are considered; 92 % when all triad types are considered). In the latter case, quite considerable variability is seen among the cluster sizes when all forbidden types of triads are considered. More precisely, the tendency to form one cluster with a relatively high number of nodes and two clusters with a smaller number of nodes is present. This happens because the vector with the frequency of forbidden triad types (which is a vector of 0s) provides less information about the target global network structure than the vector with the frequency of allowed triad types.

Figure 3.2: The mean improvement value for each blockmodel type (generated by the RL algorithm) and selected set of triad types



When generating networks with a hierarchical blockmodel, a blockmodel structure, which is not assumed, emerges. Instead, links exist in the blocks below the diagonal of the matrix and in the blocks above the diagonal. This means there are links from the top to the lowest clusters and the

other way around. On the level of nodes, only asymmetric links are possible. However, the density is still higher in complete than in null blocks (Figure 3.7), which may be due to the optimization algorithm for pre-specified blockmodeling.

When all allowed triads are included in the process of generating networks, one would expect a similar MIV as when all triads are included in the model because all the information for generating the networks embedded in all triad types is also embedded in only allowed triad types (since all the rest have a count of 0). Yet the results might differ due to the different ways of computing errors.

The set of all allowed triad types and the set of triads with selected allowed types of triads vary only in the case of a hierarchical-cohesive blockmodel and a transitive-cohesive blockmodel. The selection of triad types slightly improves the MIV with both blockmodel types. In the former case, the blockmodel structure can be visually recognized in most, but not all, networks that are generated. On the other hand, there are very low levels of errors in all generated networks with a transitive-cohesive blockmodel.

Comparing the networks generated with all forbidden triad types and the networks generated with only the selected forbidden triad type, the MIV is generally lower in the latter case for all types of blockmodels. By visually observing some generated networks, it is hard to recognize the assumed blockmodel structure, except for some transitive-cohesive blockmodels.

3.5.2 Networks generated with the MCMC algorithm: fixed density

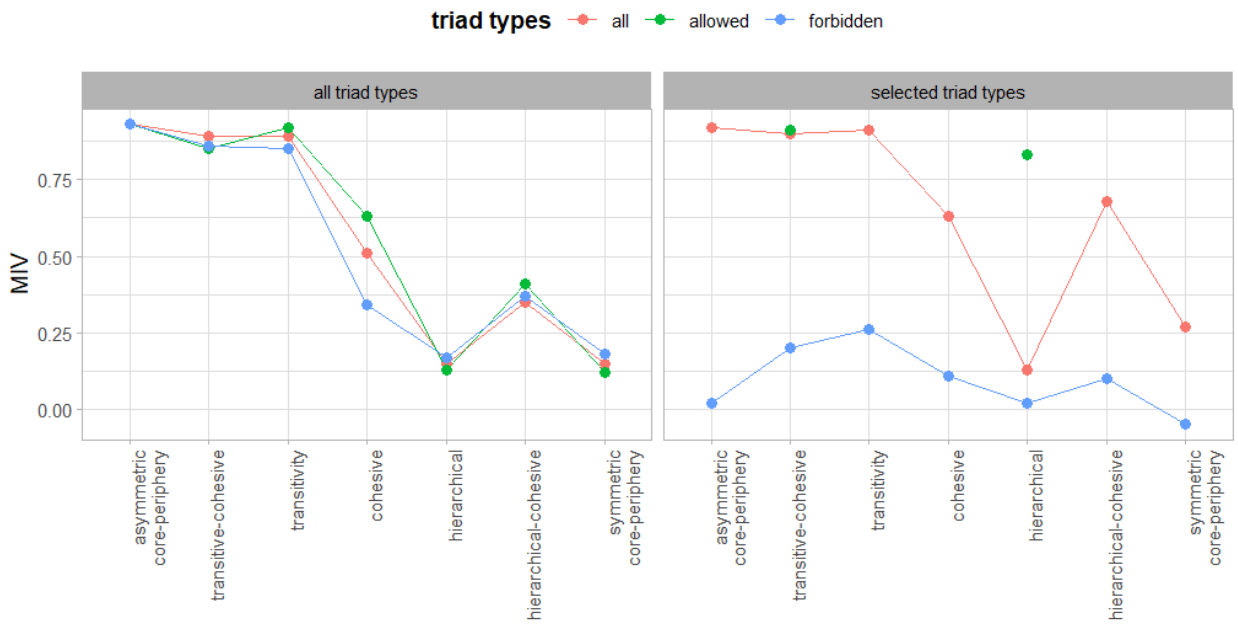
Since the RL algorithm is more deterministic, it generally performs better than the MCMC algorithm. But when networks are denser, the MCMC algorithm might perform better such as when e.g. considering the set of all allowed types of triads while generating a symmetric core-periphery blockmodel. This is another reason for considering different algorithms while studying microstructures in the context of various global network structures using simulations.

When all possible triad types are considered, the overall MIV among all blockmodel types is higher when the networks are generated using the RL algorithm and lower when the networks arise from the MCMC algorithm with a fixed density (Figure 3.3). Yet, generated networks have an assumed blockmodel structure (Figure 3.4) with a relatively low level of errors, except the hierarchical one where the global network structure obtained is similar to that produced with the RL algorithm

(considering selected allowed triad types) (see Figure 3.7). Further, the hierarchical-cohesive blockmodel and the cohesive blockmodel are not as clear as the others.

Considering only all allowed or only all forbidden triad types does not produce networks with a significantly higher level of errors. The MIVs are lower when selected forbidden triad types are considered compared to the case when all forbidden triad types are considered. In this instance, the generated networks do not have the expected blockmodel.

Figure 3.3: The mean improvement value for each blockmodel type generated by the MCMC algorithm with fixed density and selected set of triad types



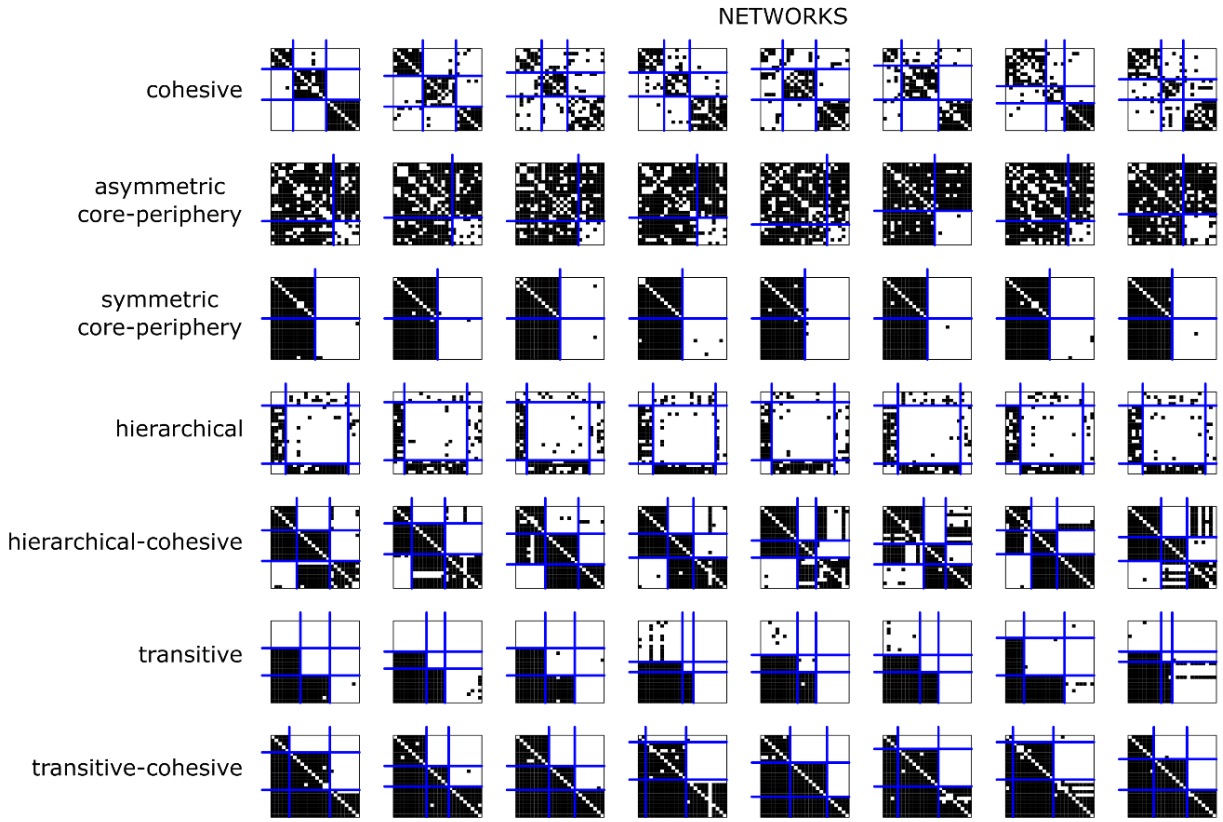
Further selection of the different types of triads that are allowed does not improve a hierarchical-cohesive blockmodel, even though some MIVs indicate the opposite. The MIVs are higher in the case of the selected triad types because the global network structure is (in some generated networks) closer to the cohesive blockmodel with two clusters. Because pre-specified blockmodeling is applied, one of the obtained clusters has only two or three nodes, which leads to the overestimated MIV.

The number of errors in complete blocks of the generated blockmodels is lower than when considering all allowed triad types, but the generated blockmodels are not hierarchical-cohesive.

The further selection of all possible triad types (allowed and forbidden) improves all the MIV values, especially those corresponding to the hierarchical-cohesive blockmodel and the cohesive blockmodel.

The generated networks with the expected hierarchical blockmodel structure are not in line with the expected global network structure. This is true for any set of triad types considered. Possible treatments are considered in subsection 3.5.4.

Figure 3.4: Some randomly selected empirical networks generated using the RL algorithm by considering all triad types



3.5.3 Networks generated with the MCMC algorithm: non-fixed density

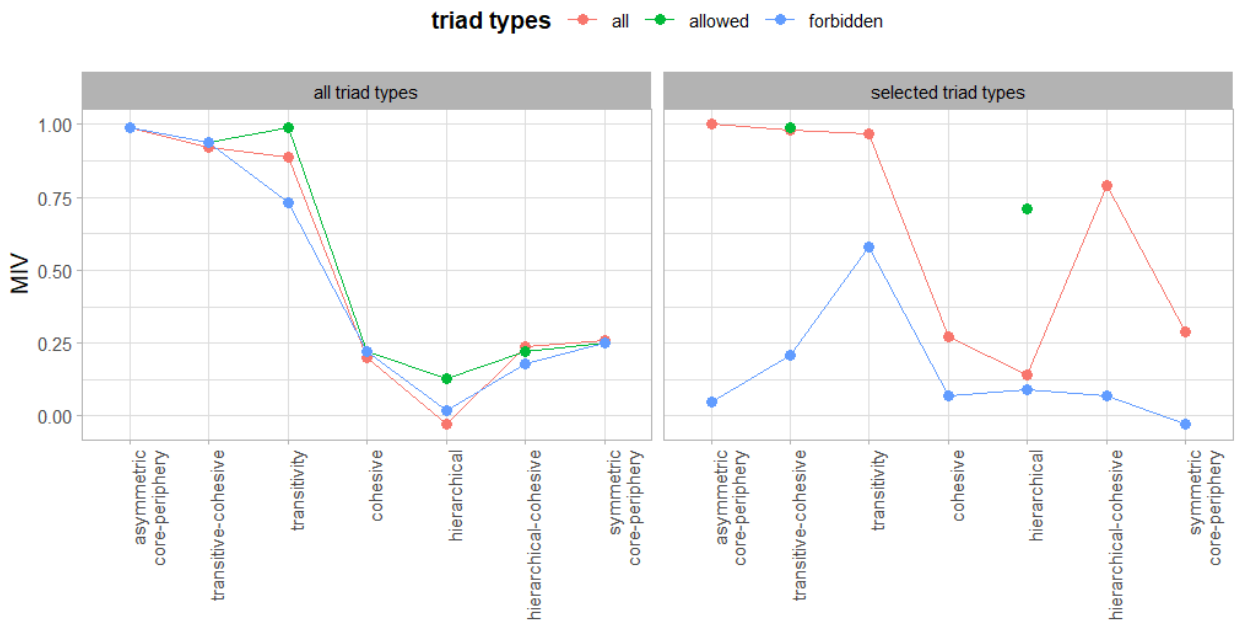
In the event the initial networks are totally randomized ideal networks, networks generated using the MCMC algorithm with a non-fixed density are close to the networks with a fixed density (Figure 3.3 and Figure 3.5).

The further selection of different triad types does not considerably increase the MIV for most blockmodel types when using the MCMC algorithm with a non-fixed density. An increase in the

MIV is observed when all selected triad types are considered (compared to the case of all possible triad types) when generating a hierarchical blockmodel (from around zero to 14 %). The global network structure of the generated networks is similar in both cases (Figure 3.6), but in the latter case there are fewer links between those in the highest to those in the lowest hierarchical position. A bigger increase in the MIV is noted in the case of considering the set of selected allowed triad types compared to considering all allowed triad types (from 13% to 71%). However, the MIV are overestimated in this case because the true underlying global network structures consist of two, not three, clusters.

A very significant increase in the MIV (from 14% to 71%) is noted in the case of generating a hierarchical blockmodel by considering the set of selected triad types compared to the case when the set of all triad types is considered. The global network structure is consistent with the hierarchical-cohesive blockmodel.

Figure 3.5: The mean improvement value for each blockmodel type and selected set of triad types generated by the MCMC algorithm with variable density



Here, it is noted that the way the initial networks are chosen has a great impact on the networks that are generated. In the case of the MCMC algorithm with a non-fixed density, considering the random networks (as initial networks) with the expected (the actual number becomes a random

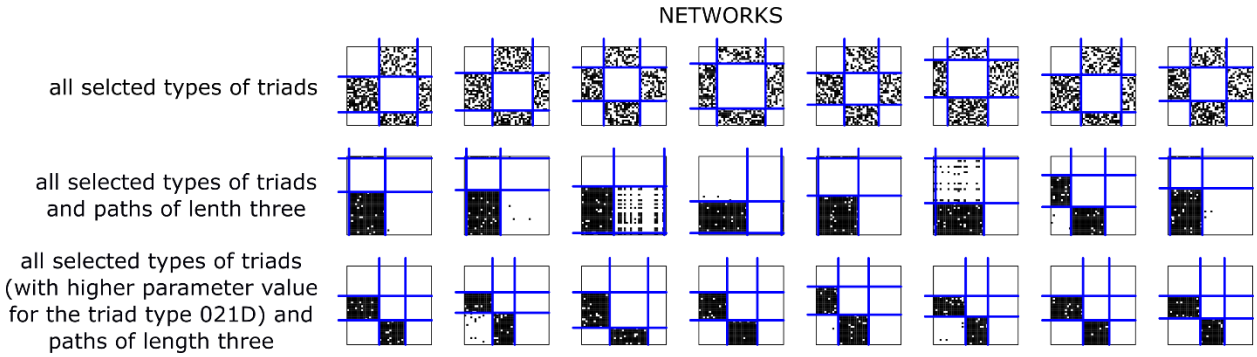
variable) number of links being equal to the number of nodes usually produces a very high number of totally empty or full generated networks. This is especially when all triad types are included in the model. In this study, randomized ideal networks are used as initial networks, meaning the density of the initial networks is not variable and is the same as in ideal networks.

3.5.4 Improvement of the hierarchical blockmodel

The proposed models for generating networks with a hierarchical blockmodel structure perform poorly. This is seen by the MIVs and the empirical examples (see Figure 3.7 and Figure 3.4).

The obtained blockmodel structure is often hierarchical but has additional links from the upper to the lower clusters and with all asymmetric links. This is especially typical of networks generated using the MCMC algorithm. Therefore, the focus is put on networks generated using the MCMC algorithm with a non-fixed density. The resulting global structure probably emerges since all considered triad types appear in all parts of the network. Their combination produces a network that is highly determined by paths of length three (e.g., $1 \rightarrow 2 \rightarrow 3 \rightarrow 2$, where digits denote clusters).

Figure 3.6: Some randomly selected generated networks with a hierarchical blockmodel generated by the MCMC algorithm with a non-fixed density using different local network statistics



Therefore, by considering paths of length three, the links from the upper to the lower positions are omitted. Here, it should be pointed out that the number of triads is unit-based while the number of paths of length three is an edge-based count. However, an additional parameter paths of length three (in the case of networks with a different number of positions, paths of different lengths should be considered) is added to the model with the value of -2 (as forbidden). Networks generated using this model have the expected hierarchical structure but with only two clusters (Figure 3.6).

From time to time, networks with a transitivity blockmodel are also produced ($MIV = 0.74$). To obtain three positions (instead of two), the parameter's value of triad type 021C must be increased, e.g. to the value to 4. Such a model produces networks with a very clear hierarchical structure (Figure 3.6). There are no errors in all generated networks in null blocks while some appear in complete blocks ($MIV = 0.93$).

All of the described networks were generated using the MCMC algorithm. When the RL algorithm is used, all allowed types of triads and paths of length three can be considered. In that case, some errors appear in both null and complete blocks, which is a consequence of the fixed density. However, with a higher number of iterations, the number of errors could also be lower.

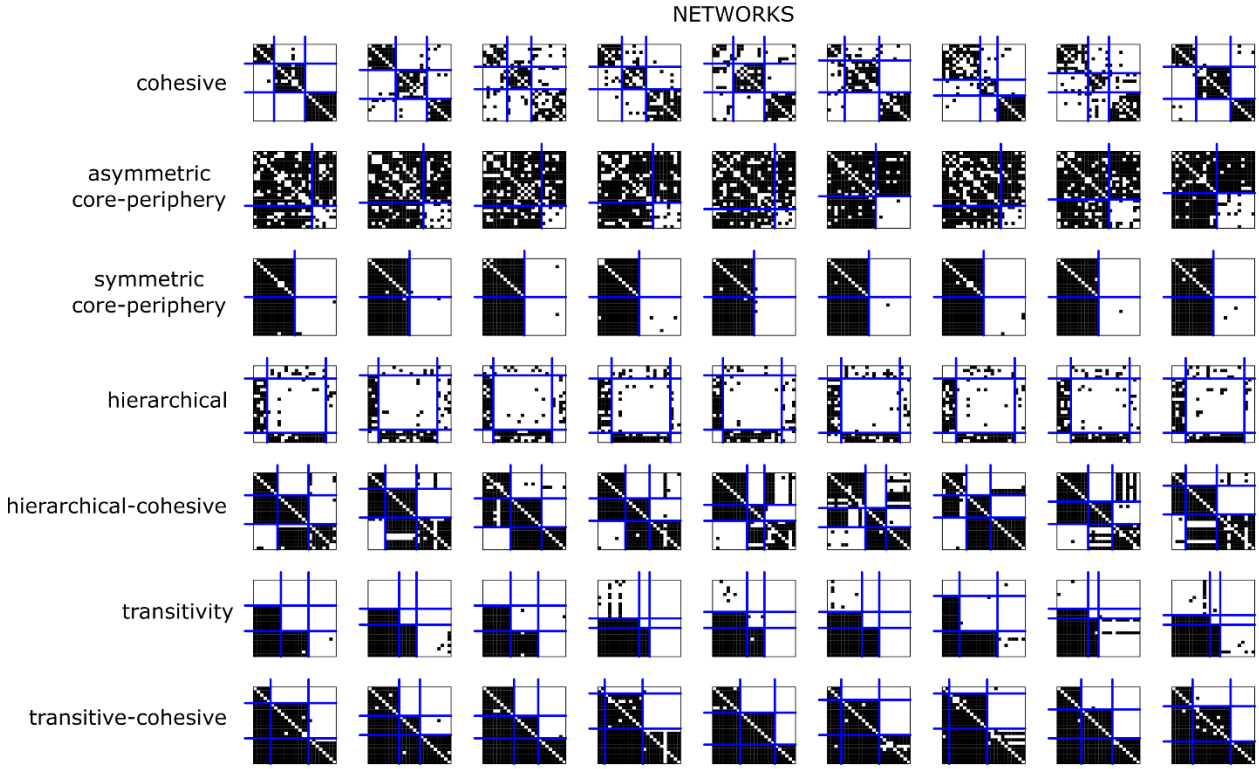
3.6 Conclusion

The aim of this chapter is to test whether networks with the chosen blockmodels can be generated by considering only different triad types. To this end, two different algorithms were used: the proposed deterministic Relocating Links (RL) algorithm, and the Monte Carlo Markov Chain (MCMC) algorithm (specifically the Metropolis-Hastings algorithm) (Hunter et al., 2008). The RL algorithm randomly selects a link and exchanges it with a randomly selected non-link. The change is accepted if the new network's local structure count is closer to the target count than in the previous network. With the MCMC algorithm, the same local structures are used as parameters in the ERGM model.

To determine the target count for the RL algorithm and the parameter values for the MCMC algorithm, the count of different triad types in ideal networks (namely, those that perfectly comply with a certain blockmodel) has to be determined. This is achieved by considering the specific blockmodel type and corresponding cluster sizes. All types of triads are classified in the set of forbidden or in the set of allowed triad types (for each blockmodel). Allowed triad types are those that are present in ideal networks and forbidden triad types are those that are not present in ideal networks. The RL algorithm uses counts of a selected local structure in an ideal blockmodel while for the MCMC algorithm the parameter values are determined based on the classification into allowed and forbidden triad types.

Both algorithms perform well in the case of asymmetric core-periphery blockmodels. However, in the case of the symmetric core-periphery blockmodel, the MIVs are usually small, which is reflected by the insufficiently small periphery in the generated networks. It is also hard to generate a hierarchical blockmodel when considering only different triad types, regardless of the algorithm that is used to generate the networks. By adding paths of length three, the empirical networks produced have the expected blockmodel type with a very low level of errors (see Figure 3.6).

Figure 3.7: Some examples of networks generated using the RL algorithm by the selected types of triads, for each type of blockmodel



The main finding is that the chosen blockmodel types can be generated by considering different triad types. This study also confirms that the number of different types of triads reflects the assumed global network structure (all generated networks have the expected blockmodel, see Figure 3.4 and Figure 3.7) where it is often enough to consider only some of all possible types of triads. Considering further local network statistics can considerably improve the global network structure of the networks that are generated.

This chapter also explored whether one can reduce the required local structure information by using only allowed or only forbidden triad types. Using only forbidden types of triads is especially

desirable for the RL algorithm as the count for this triad type is zero. In addition, the reduction of all these sets (all, allowed, forbidden) of triad types was studied based on their sensitivity to errors according to the blockmodel structure. Most of these reductions of sets of different triad types overall resulted in only a slightly worse fit and in some cases even in an improved performance. The only exception is when using only selected forbidden triad types, which often did not generate the assumed blockmodel structure.

Some considered blockmodel types are defined for only two clusters (symmetric and asymmetric core-periphery blockmodels). The other blockmodel types can consist of more than three clusters. The initial tests suggest that for the cohesive blockmodel, the transitivity blockmodel and the transitive-cohesive blockmodel, the results presented in this chapter can be generalized to blockmodels with higher numbers of clusters, while for other blockmodels one cannot make such speculations.

4 Emergence of the asymmetric core-cohesive blockmodel

The previous chapter showed that networks with the most common blockmodels can be generated from random networks by considering only different types of triads. This is important because it shows that very clear global network structures can emerge solely due to these local network characteristics (triad types) and suggests that one can define some local network mechanisms that produce a given global network structure.

The focus of this chapter is on studying the relationship between local network mechanisms and blockmodels. The difference between this and the previous chapter is that the mechanisms are defined as rules for creating links (by way of the nodes) in a dynamic process while the earlier chapter considered local network configurations (i.e. the number of different triad types.

Since there are many possible blockmodel types and local network mechanisms, in this chapter the context of liking and friendship networks in kindergarten is considered (it is arbitrarily chosen), to find an appropriate blockmodel type and corresponding local network mechanisms.

The blockmodel type is selected based on the observation that children start to form groups when they enter kindergarten. Those within a child's group spend more time with each other than with those from other groups. At the same time, a group (a popular group) of children is formed with which all the other children want to spend considerable amount of time. Therefore, the selected blockmodel is the asymmetric core-cohesive blockmodel type, which is assumed to be present in empirical preschool children networks. This blockmodel type is a mixture of two very common blockmodel types, namely: the asymmetric core-periphery and cohesive types.

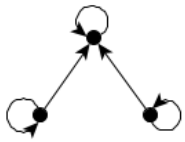
The local network mechanisms are selected based on existing studies on the evolution of popularity and friendship networks among the preschool children (see Section 4.2), assuming that such a structure can be found in these networks. Another reason explaining the choice of mechanisms is the assumption that the popularity mechanism leads to the asymmetric core-periphery blockmodel type and transitivity-related mechanisms lead to the cohesive blockmodel type. All mentioned mechanisms and blockmodel types are described in more detail in the sections below.

4.1 Asymmetric core-cohesive blockmodel

The asymmetric core-cohesive blockmodel type has at least three clusters of nodes. The nodes of each cluster are internally well linked. All nodes are linked to the core cluster of nodes. Such a blockmodel type with three clusters is shown graphically in Figure 4.1. Here, the nodes represent the clusters of nodes. A cluster, as visualized at the top of the graphic presentation in Figure 4.1a, consists of internally highly linked nodes and is called a core cluster (or core group). The nodes in cohesive clusters (or cohesive groups) are internally highly linked and are linked to the core cluster. The blockmodel may be extended in such a way that the nodes from one cohesive cluster are not highly internally linked.

The core nodes could be named popular nodes since the term "popularity" is often associated with a high in-degree, which is a characteristic of the core cluster of nodes. The term core is more general than the term popular.

Figure 4.1: An asymmetric core-cohesive blockmodel with three clusters



com	null	null
com	com	null
com	null	com

4.2 Mechanisms which might lead to the core-cohesive blockmodel

Since the asymmetric core-cohesive global network structure has not yet been formally presented in the form of a blockmodel, it is not mentioned in empirical studies concerned with the evolution of global network structures. However, on the assumption that the proposed blockmodel type is present in the liking friendship networks among the preschool children, several mechanisms that might lead to the proposed blockmodel type may be identified based on the previous literature.

It has to be stressed here that the characteristics of the nodes are not considered in this study, even though they can play an important role in how links are formed in real networks[19]. The ability to

[19] Kerns (2000) studied the friendships among children aged between 42 and 84 months and was able to confirm the existence of several types of friendships (e.g. a cluster of harmonious, responsive and interactive friendships and a cluster of harmonious but independent friendships). Some types are expected to last a longer time while others are

generate networks with the desired blockmodel types when considering only the selected local network mechanisms may suggest personal characteristics are not needed for the emergence of the proposed global network structure.

Moreover, when evaluating studies conducted on networks among preschool children according to the global network structure, one must differentiate between popularity and friendships[20]. Popularity refers to the view that the group holds with respect to an individual, in terms of different levels of liking and disliking, while friendship is conceptualized as a bilateral construct (although empirical measurements of friendship networks are usually asymmetric). Considering liking as the basic link between popularity and friendship, Bukowski et al. (1996) showed that the positive association between popularity and friendships decreases with age.

Since conducting longitudinal sociometric interviews with a high level of reliability and validity among preschool children might be too demanding for both the children and the researcher, the data which are analysed are often observational. In such studies, a link is typically operationalized as an interaction and therefore the links that are observed are non-directed. If such interactions are considered as an indicator of friendship, popularity or liking, the same mechanisms must be

expected to not survive for as long. However, with respect to the latter it can also happen that they last relatively long due to some personal characteristics, e.g. in some cases friendships can be the outcome of graphical proximity between two children or the result of one or both children in the current friendship lacking social skills. Similarly, Proulx & Poulin (2013) showed that personal characteristics such as aggressiveness, social pro-activeness and shyness affect the number and stability of friendships among children in kindergarten (also see Engle, McElwain and Lasky (2011)). Depending on the children's age, gender is another important personal characteristic when forming cohesive groups of friends (Johnson et al., 1997). Adams & Torr (1998) highlighted that friendships are much more influenced by the cultural context than any other institutionalized relationship.

[20] This was also considered when operationalizing popularity through in-degree in empirical networks. For example, when thinking about friendships in a preschool environment the number of peer contacts relates more to the perception of popularity among peers than the quality of contacts. Those with a higher number of peer contacts and those in the centre of the network are hence seen as more popular while the isolates are seen as more unpopular by the others (La Fontana & Cillessen, 2002). Popularity can be achieved by positive or negative behaviour (Cillessen & Rose, 2005). While groups of aggressive and popular students were found in many empirical studies among older children, a study by Estell (2007) did not confirm this was the case for children in a kindergarten.

accounted for when testing for the emergence of the symmetric core-cohesive blockmodel type. The following mechanisms are often discussed in the literature:

- **Mutuality** or reciprocity is defined as the reciprocation of ties and is one of the most fundamental local network mechanisms and a basic feature of social life (Daniel, Santos, Peceguina, & Vaughn, 2013). Analysing 49- to 62-month-old preschool children, Snyder et al. (1996) not only found that children spend considerable time with selected friends and less with others, but also observed the strong mutual affiliation of friendships. The mutual links observed in the empirical global network structures can also be due to the fact that children prefer to interact with peers who are similar to themselves. This tendency often fosters the emergence of mutual peer relationships during childhood (Block, 2015; Kandel, 1978; McPherson et al., 2001; Schaefer et al., 2010).
- **Popularity** is defined through an in-degree in social network analysis and is usually an operationalization of likeability or social status (Daniel et al., 2013). As a local network mechanism, popularity expresses the tendency to create links with others with a relatively high (in)degree (popularity level). The fact that some nodes become more popular than others may relate to their personal attributes (e.g. wealth, being good at something etc.) or positive or negative behaviour (Cillessen & Rose, 2005).
- **Transitivity** measures the tendency for triadic closure in networks – "the friends of my friends are also my friends". Transitivity in peer groups may arise from the increased propinquity of individuals who share mutual friends, or from a psychological need for balance – a convergence of third parties' evaluations (Schaefer et al., 2010).

Many other studies conducted among older individuals in a school environment have focused on the local network mechanisms underlying different kinds of networks. Some of these are mentioned below to show that: (i) similar local network mechanisms might also be at work among older individuals; and (ii) that the global network structure of liking or friendship networks observed in preschool or among older individual might be similar. Yet these studies also show that the strengths of the local network mechanisms might change with age.

Schaefer et al. (2010) studied the three most common network-formation mechanisms (reciprocity, popularity, and triadic closure) among preschool children throughout a school year in four waves using SIENA (Block et al., 2016; Handcock et al., 2003; Snijders, Van de Bunt, et al., 2010). They

found the reciprocity effect is constant over time while the popularity effect is most important midway through the school year. The importance of the triadic closure effect increases in time, which is expected since very early friendships are typically play-oriented dyads that primarily socialize children into group life (Hartup & Stevens, 1997). When children gain more social contacts and greater confidence, they move into larger groups (Hartup, 1993).

Daniel et al. (2013) used ERGM (Robins, Pattison, Kalish, & Lusher, 2007) to study the mutuality, reciprocity, popularity and transitivity mechanisms on the forming of affiliative ties in 19 Portuguese preschool peer groups. They found that all of these mechanisms are important for forming affiliative ties.

Later, Daniel et al. (2019) collected and analysed interactional network data collected in several waves among children aged 3 to 5 years. The data were collected in several classes and analysed using SIENA. Compared to the data collected by Schaefer et al. (2010), the data collected by Daniel et al. (2019) also contain information on who initialized the interaction. Based on the analysed data, the researchers were unable to confirm that the importance of local network mechanisms changes over time. As they explain, this might be because they did not start collecting the data immediately after the school year had commenced.

Dijkstra, Cillessen & Borch (2013) found that higher-status adolescents strive to maintain their status by keeping lower-status adolescents at a distance. Using SIENA, they analyzed longitudinal data obtained from students from grades 6 to 8 of middle school. Adolescents strongly prefer similar or more popular others since this can lead to a higher status for themselves. Here, popularity increases the receipt of best-friend nominations, but decreases the giving of them. The idea that lower-status individuals are the initiators of friendships was also raised by Hallinan (1978).

Crockett, Losoff & Petersen (1984) conducted semi-annual interviews among 335 boys and girls who were followed longitudinally from grade 6 to grade 8. They confirmed the perceived importance of being part of a clique increases with the students' age. On the other hand, a study by Brown, Eicher & Petrie (1986) shows the importance of being part of a clique decreases among adolescents from grade 7 to grade 12. Shrum & Cheek (1987) observed that the share of students who are classified as members of the group first increases until grade 6 but starts decreasing in later grades. This is not due to the higher number of isolates but to the increasing number of

liaisons[21]. Liaisons are "less likely to have friends who interact predominantly with one another" (Shrum & Cheek, 1987, p. 222).

Lubbers (2003) studied the structure of the within-class social networks of students (on average aged 13 years) with special attention to the differences with respect to the network structures across classes. To this end, she used a multilevel application of ERGM. She observed very strong evidence of the tendency toward mutuality, transitivity and a very strong tendency against cyclicity and 2-mixed-stars. The combination of the latter is interpreted as showing that the relationships among students are hierarchically structured, especially when the link between two students is defined through co-operation and when boys are analysed. Indicators of a hierarchy were found by Gonzales et al. (2007) who studied school friendship networks (grade 7 to grade 12) drawn from the Add Health study database using the threshold analysis. They extracted networks based on 3-clique communities and networks based on 4-clique communities. Based on the distribution of degrees, the authors suggest the network structure that is obtained may be due to the "rich get richer" effect while the clustering coefficient analysis revealed a hierarchical structure is present in the friendship network.

4.3 Research question

This chapter addresses the second general research question of this dissertation. The research question this chapter examines is as follows: Can the selected mechanisms lead the network to the core-cohesive blockmodel structure?

The research question is then broken up into several parts entailing several sub-questions as to whether a core-cohesive blockmodel can emerge from: (i) an empty network; (ii) a cohesive blockmodel; or (iii) an asymmetric core-periphery blockmodel. All of these cases consider popularity, assortativity, two transitivity-related mechanisms and mutuality mechanisms as local network mechanisms.

[21] A liaison is defined as (Shrum & Cheek, 1987, p. 220): "an individual who is linked as a (1) tree node, that is, a link connecting into branching structures with isolates (type 2) at one end and group members or other liaisons at the other; (2) direct liaison, most (>50.01 percent) of whose interaction is with group members (but not any one group); (3) indirect liaison, most (>50.01 percent) of whose links are with other liaisons".

This chapter is divided into several sections. In the next section, the network evolution model used to generate the networks is described, followed by formal definitions of the listed mechanisms. Methods for evaluating the global network structures are also presented in this section while the results are given in Section 4.5. The latter is further organized in several subsections regarding the global network structure of the initial networks.

4.4 Methods

Probability models have been already proposed for friendship networks consisting of a given number of cliques (Jansson, 1997) and probability models for popularity structure in social networks where popularity is defined through in-degree (Jansson, 2000).

In this analysis, an algorithm from the NEM family is used to generate networks by considering several local network mechanisms with different strengths (weights). The obtained global network structures are then evaluated by the number of inconsistent blocks and the RF (see subsection 2.5.2). The local network mechanisms are then discussed.

The proposed NEM imposes two main assumptions common to the Markov process: two links cannot change simultaneously, and the probability of a link changing can be expressed as a function of the entire network at a certain point in time (Zeggelink, 1994). In the proposed NEM, the next step in the global network structure depends on the global network structure in the current step.

The actors' actions are goal-oriented (Snijders, 1996), meaning that "each actor takes actions in order to fulfil his own goals; these actions are in the domain of his own behaviour or of the directed relationships from him to others". The actions are operationalized through mechanisms.

The alternative approach to studying the global network structure's emergence and evolution is to use SAOM (the same assumptions mentioned above also apply to SAOM). This approach can be applied in several ways.

One way is to use SAOM to generate networks with parameter values that are randomly selected. The main issue in this regard is that the space of all possible parameters' values is very huge, making it very hard to find the most appropriate parameter values. It is also often the case in SAOM

and ERGM that small changes in the parameter values lead to considerable changes in the global network structures. Therefore, even a higher number of different weights' values would have to be considered to use the proposed methodology for the NEM.

The number of all possible parameter values may be reduced by using the initial values estimated on empirical networks, although this is not always possible, especially when the global network structure of interest has not yet been empirically observed and the data are not in longitudinal form.

Another approach is to generate several artificial networks with the global network structure of interest and to estimate the SAOM parameter values based on such networks. These networks could be created in several ways. For example, some random errors could be added to the networks with the chosen blockmodel(s) without errors. In this case, the errors must be generated by considering the dependency in the global network structures between consecutive networks. Otherwise, it could happen that the parameters are left not estimated because the randomly generated errors would not reflect the presence of the local network mechanisms being studied. In other words, the way in which the errors are added to the ideal networks affects the SAOM estimates. This is another rationale for choosing the proposed NEM over the SAOM when studying local network mechanisms.

4.4.1 Network Evolution Model

Before applying the iterative algorithm for generating random networks (see Algorithm 4.1), one must specify the initial network X_0 in the form of a binary $n * n$ adjacency matrix X, the vector of weights of the local network mechanisms θ, the probability of establishing a link at each iteration q, and the number of iterations m.

The algorithm is as follows: at each iteration, node i is randomly selected with probability $1/n$. Then, the network statistics S are obtained by the operationalized selected mechanisms presented in the next subsection. These network statistics are weighted by the vector of weights of local network mechanisms θ producing vector $\varphi = S\theta^T$. The nodes, for which it holds $\varphi \geq Q_3(\varphi)$ (Q_3 stands for the 3rd quartile), are classified in set C and the nodes, for which it holds $\varphi \leq Q_1(\varphi)$ (Q_1 stands for the 1st quartile), are classified in set F. With probability q, the link from i to randomly selected j from set C is set and with the probability $1 - q$ a non-link from i to randomly selected j from set F is established.

Algorithm 4.1: The NEM algorithm for generating interactional networks

```
set initial network X_0
set vector of weights of the mechanisms θ
set probability of establishing a link q
set number of iterations m
repeat m-times
|randomly select node i
|calculate network statistics according to the selected mechanisms for the node i and all
|other nodes; save it in S
|
|calculate φ = Sθ^T
|
|classify node j into the set C if φ ≥ Q_3(φ), where Q_3 is the 3rd quartile
|classify node j into the set F if φ ≤ Q_1(φ), where Q_1 is the 1st quartile
|
|with probability q set i → j where j is randomly selected from the set C
|with probability 1 − q set i ↛ j where j is randomly selected from the set F
return generated network
```

The fact that a node can establish a link to those to which it is already linked can result in no visible change of a link upon a given iteration. Similar is true when node i establishes a non-tie to node j if there was previously no link. The links in the resulting network are represented by 1s while non-links are shown by 0s.

4.4.2 Local network mechanisms

Different local network mechanisms can be considered in the proposed NEM. The term mechanism describes a process that drives concrete actions according to nodes in the networks (e.g. creating a link to a highly popular unit). These mechanisms can be operationalized by different statistics (e.g. the alter in-degree) that reflect the mechanisms.

As described in the previous section, the proposed network statistics S are weighted by a given vector θ. A higher positive weight is associated with the greater importance of a given mechanism while a lower positive weight is associated with a given mechanism being less important. Higher negative weights are also associated with mechanisms that are more important, but the direction of the mechanisms' effect is the opposite. For example, a negative weight for the popularity mechanism is interpreted as a tendency to avoid creating links with highly popular others.

In the current study, the following network statistics are defined and used (see Figure 4.2): mutuality (M), assortativity (A), popularity (P), transitivity (T) and outgoing shared partners (OSP). They are defined on a network with n nodes, represented by the binary $n * n$ adjacency matrix X, and normalized in such a way that the minimum corresponding values are 0 and the

maximum values are 1. The different network statistics considered are schematically shown in Figure 4.2 where dashed lines illustrate the links being evaluated in terms of appearing, confirming or disappearing. The image shown in Figure 4.2a corresponds to the tendency of having a link. Since this is not a focal mechanism, it is implemented in the NEM algorithm as parameter q. The following network statistics are considered in this section:

1. **Mutuality** (Figure 4.2b) reflects the tendency to reciprocate links. It is operationalized by an asymmetric in-ties statistic as

$$asymIn(i,j) = \begin{cases} 1 & if \quad x_{ji} + (-1) * x_{ij} = 1 \\ 0 & if \quad otherwise \end{cases} \tag{4.1}$$

   The first condition is satisfied only when there is a non-reciprocal link from node j to node i. We are using the normalized statistic (M) defined as

$$M(i,j) = \frac{asymIn(i,j)}{\sum_{j=1}^{n} asymIn(i,j)} \tag{4.2}$$

2. **Alter popularity** (Figure 4.2c) (referred to as "popularity" below) reflects the tendency of creating links with the most popular ones. The popularity statistic (P) is calculated for i-th node as the ratio between the in-degree of the i-th node and the total number of nodes

$$P(i,j) = \frac{\sum_{j=1}^{n} x_{ij}}{(n-1)} \tag{4.3}$$

3. **Assortativity** (Figure 4.2d) reflects the tendency to create links to those nodes with the same level of popularity (in-degree). The assortativity statistic (A) is defined as

$$A(i,j) = \max\left(1 - \frac{\left|\sum_{i=1}^{n} x_{ij} - \sum_{j=1}^{n} x_{ij}\right|^2}{4}, 0\right) \tag{4.4}$$

   and can be calculated for each pair of nodes: if i and j have the same in-degree, they have the same assortativity with all other nodes. The $A(i,j)$ has the value 1 when nodes i and j do not differ in the number of incoming ties. In our model, the node that has a chance to change a tie is chosen and therefore for only this node and all the others is $A(i,j)$ applied.

4. **Transitivity-related mechanisms** (Figure 4.3) reflect the tendency to create links to those nodes with whom a node shares a higher number of reciprocated or non-reciprocated links to other nodes in the network. The concept of the selected mechanisms generalizes the ideas of triangles and 2-stars (Snijders, Pattison, & Handcock, 2006). It is very common, especially when analysing networks by ERGM or SAOM. There are four types of shared partners (Robins, Pattison, & Wang, 2009), yet just two of them are considered in this chapter:

a) **Transitivity (T)**, also referred to as an outgoing two-path (OTP) (Figure 4.2e): a transitivity relation is defined as $i \rightarrow j$ where also $i \rightarrow k \rightarrow j$. This means that transitivity is a tendency for a node to directly connect to those nodes to which it is indirectly connected with (one or more) paths of length two (with more paths increasing the tendency).

b) **Outgoing shared partners (OSP)** (Figure 4.2f): node k is a shared partner of the ordered pair (i, j) if $i \rightarrow k$ and $j \rightarrow k$; OSP represents a "structural homophily" effect which is traditionally based on similarity according to the nodes' attributes. In the case of OSP, the effect is defined by similar choices of partners (Robins et al., 2009).

To compute the statistics associated with these mechanisms on the selected pair of nodes i and j, one must identify the other nodes (not i and not j) which are linked with i and j (shared partners) in a given way. When the transitivity statistic is considered, the number of alternative intermediaries on two paths from i to j is computed as

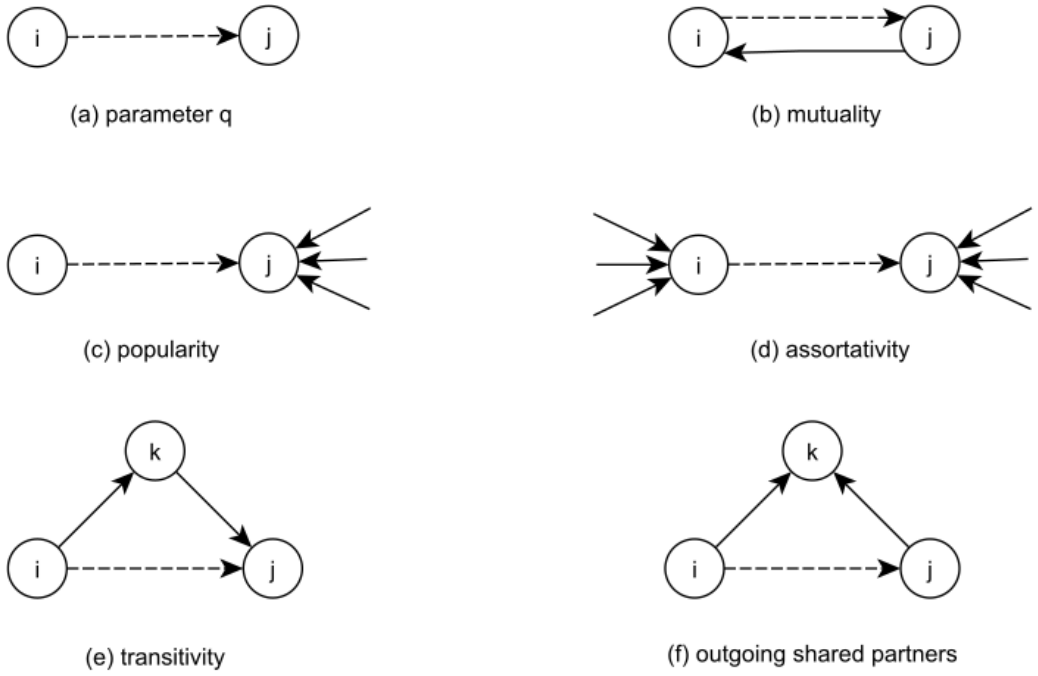
$$T(i,j) = \sum_{k \neq i,j} x_{ik} * x_{kj} \tag{4.5}$$

Similarly, the number of outgoing shared partners is computed as

$$\boldsymbol{OSP(i,j)} = \sum_{\boldsymbol{k \neq i,j}} \boldsymbol{x_{jk}} * \boldsymbol{x_{ik}} \tag{4.6}$$

Both $T(i,j)$ and $OSP(i,j)$ give the number of shared partners of a given type between i and j. By fixing node i, one can obtain vector V with n elements where each value stands for the number of common friends between the node and all the other nodes. The j-th value of vector V can be normalized as $V_j / \sum_{j=1}^{n} V_j$. Such normalized statistics are used to operationalize the transitivity and OSP mechanisms.

Figure 4.2: Illustrations of different local network mechanisms



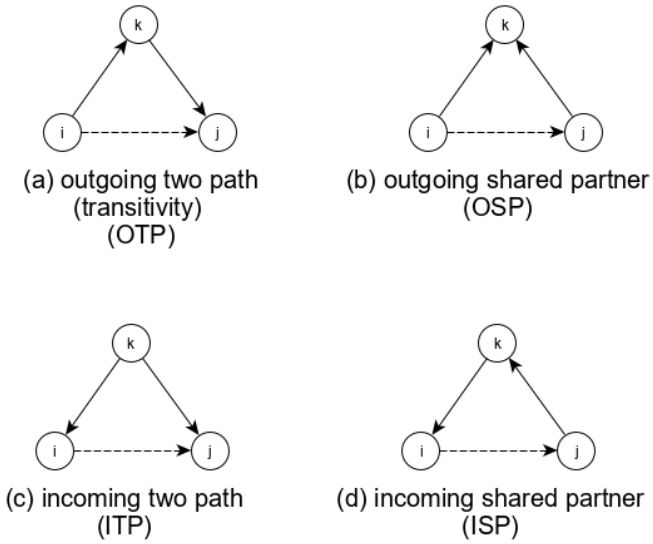
It is assumed the mechanisms' weights are the same for all nodes regardless of their position in the network. However, the effect of a mechanism can vary from node to node due to the weighted network statistics (φ), which may vary among nodes, depending on their network position[22].

The weights of the mechanisms can also vary from node to node and may depend on e.g. age or some other nodes' attributes. Since nodes' attributes are not considered in this study, this issue does not apply, but it is worth mentioning that some authors argue (based on analysis of empirical studies on friendship networks) that the mutuality mechanism might be weaker for those nodes which are involved in transitive triplets, compared to the others (Block, 2015). The author explains this by stating that "embedding in a transitive group is more important for a one-sided tie than a reciprocal one" (Block, 2015, p. 170). After explicitly addressing this issue in their study of

[22] There are some other considerations which often arise for most empirical studies that use SAOM or ERGM. For example, the weights of the mechanisms are fixed for all nodes regardless of their attributes. In addition, a global network structure can influence the selected mechanisms' importance. For example, some empirical studies on friendship networks show the mutuality mechanism might be weaker for those nodes involved in transitive triplets compared to the others (Block, 2015). However, differentiating the strength of the impact of a global network structure from the set of mechanisms' weights of a given node and vice versa is hard and quite context-specific.

interactional networks among children, Daniel et al. (2019) were unable to confirm this effect was present among preschool children.

Figure 4.3: Illustration of different types of shared partners



4.4.3 Selecting the weights of the mechanisms

One way to select the set of mechanisms' weights is to generate networks using the proposed model with different sets of mechanism weights (in this study, 30 networks were generated with each set of mechanism' weights). To access whether the generated networks have the asymmetric core-cohesive blockmodel structure, non-specified direct blockmodeling (Batagelj, Ferligoj, et al., 1992; Doreian et al., 2005) is applied.

To perform blockmodeling, the "blockmodeling" package (Žiberna, 2018) for the R programming language is used. The blockmodels obtained are compared to the ideal asymmetric core-cohesive blockmodel by the number of inconsistent blocks (see subsection 2.5.1). For a given network, one or more inconsistent blocks means a different blockmodel. In the case of many generated networks (generated in the same way, including the same selected local network mechanisms and their weights), a mean number of inconsistent blocks higher than 0.5 indicates a high probability that the selected mechanisms and their corresponding weights generate networks of another blockmodel type than desired (or being compared to). The overall number of errors of the blockmodeling structure obtained vis-à-vis the ideal one is analysed by using the RF (see subsection 2.5.2).

In order to reduce the computational burden, only 300 randomly selected θs are considered. The mechanisms' weights are selected so that the condition $\sum \theta_i^2 = 1$ is satisfied. The random values are generated by first sampling five values from the standard normal distribution ϕ and then multiplying them by the scalar $\theta = \phi / \sqrt{\sum \phi_i}$ (Marsaglia, 1972; Muller, 1959).

4.4.4 Evaluating the global network structure

The number of inconsistent blocks (see subsection 2.5.1) is used to evaluate how close to the proposed blockmodel type is the blockmodel obtained from the empirical network[23]. Although the proposed approach is sufficient for evaluating the presence of the selected blockmodel type in the empirical data, it does not reveal the overall level of errors across blocks. Therefore, the RF is used (see subsection 2.5.2)[24]. The random networks (see subsection 2.5.5) are generated such that the density is fixed while the distribution of the degrees is allowed to vary, as proposed by Boyd et al. (2006) for the core-periphery structure.

Observing the number of inconsistent blocks and the values of the RF after each of several iterations can help evaluate the convergence of the global network structure. The global network structure can stop changing significantly because the selected local network mechanisms reach the most optimal global network structure or because the local optimum has been reached. One or the other can be the case when the RF values stop changing significantly. Because of this, the convergence of the global network structure is also evaluated by looking at some visual representations of the generated networks at different numbers of iterations.

[23] The three clusters are pre-specified in the non-specified direct blockmodeling used in this chapter. This is because the focus is on generating blockmodels with three groups. Yet, in some cases, the generated networks have a core-periphery blockmodel defined on two clusters only. This does not affect visual examination of the networks so generated (e.g. one can identify the presence of an asymmetric core-periphery blockmodel in the early networks shown in Figure 4.7) or other interpreted results. In chapters where the aim is to generate symmetric or asymmetric core-periphery blockmodels (e.g. Chapter 6), the number of clusters being considered is two.

[24] For each generated network, the mean RF values are calculated by considering each selected blockmodel type, regardless of the initial blockmodel. Comparing the trajectories of the mean RF for different blockmodel types can help in understanding and explaining how the global network structure of the generated networks evolves.

4.5 Results

The section is organized into three subsections. Different initial global network structures are assumed in each subsection (empty networks, networks with a cohesive blockmodel structure, and networks with an asymmetric core-periphery blockmodel structure). The remaining methodology is the same for all three parts: 300 vectors containing the randomly selected mechanisms' weights are generated; 30 networks are generated for each set of mechanism weights; each network is generated with 32,269 iterations.

The networks generated along the way are also analyzed. The intermediate number of iterations m, at which the global network structure is analysed is determined as $m_i = m_{i-1} * 1.9$, where $m_1 = 100$. This approach is used since the preliminary analyses showed the global network structure changes more rapidly at a lower number of iterations. Therefore, more frequent insights into the global network structures are needed at a lower number of iterations.

The intermediate networks and "final" networks are analysed by non-specified direct blockmodeling with three clusters assumed and null and complete blocks allowed (structural equivalence is considered). Based on this, the average number of inconsistent blocks is calculated. Ten θs, generating networks with the lowest number of inconsistent blocks, are selected for each type of initial global network structure. The restriction to only 10 θs is made due to the high computational costs of generating networks and obtaining RF values.

The networks are then generated again, based on the selected ten θs with an increased number of iterations up to 108,694 iterations. The intermediate number of iterations for the case of $m_{11}, \dots, m_{15}$ is calculated as $m_i = m_{i-1} + m_{10}$. The values $m_{11}, \dots, m_{15}$ are calculated differently in order to avoid a high number of iterations implied by the constant 1.9. Pre-specified blockmodeling (an asymmetric core-cohesive and an asymmetric core-periphery with three clusters, and a cohesive blockmodel with two clusters are assumed) is applied and the RF values are calculated in order to more precisely evaluate the global network structure.

4.5.1 From an empty network to the core-cohesive blockmodel

The mean number of inconsistent blocks along with the corresponding θs in instances where the initial networks are empty networks are presented in Table 4.1. Note that only the results for the best 10 θs (according to the mean number of inconsistent blocks after 32,269 iterations) are given.

The mechanisms' weights shown in Table 4.1 are not directly comparable, partly because the mechanisms are interdependent and partly because they are not defined for absolute comparability[25]. The highest weights are usually for the transitivity, mutuality and assortativity mechanisms. The values of the popularity mechanism are generally lower, especially when the value of the transitivity mechanism is high.

Table 4.1: Mean number of inconsistent blocks for the selected θs (initial is an empty network and target is an asymmetric core-cohesive blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS										Mean number of IB at 108,694 iterations	Mean RF at 108,694 iterations
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269		
224	.45	.10	.77	.44	-.08	4.3	3.7	3.4	2.8	1.2	0.9	0.4	0.4	0.2	0.1	0.0	0.78
226	.65	-.04	.11	.71	.23	4.2	3.5	3.0	2.2	2.2	1.2	0.4	0.2	0.1	0.1	0.1	0.89
228	.59	.12	.24	.76	-.08	4.4	4.0	3.2	2.5	1.6	1.0	0.6	0.3	0.2	0.2	0.1	0.84
136	.41	-.18	.37	.74	-.34	4.6	4.3	3.6	2.8	1.6	0.8	0.6	0.3	0.3	0.2	0.1	0.84
238	.37	.61	.35	.00	.61	4.4	4.1	3.4	2.4	1.7	1.3	0.5	0.4	0.3	0.2	0.1	0.93
153	.84	-.14	.36	.32	-.21	4.2	4.5	3.5	3.0	1.9	1.0	0.5	0.4	0.3	0.3	0.1	0.75
286	.52	.26	.33	.72	.16	4.0	4.1	3.3	2.4	1.6	1.4	0.6	0.4	0.4	0.3	0.2	0.79
259	.51	.57	.60	.25	.02	4.2	3.9	3.5	2.6	1.5	1.5	0.6	0.4	0.3	0.3	0.1	0.81
40	-.13	-.06	.49	.85	.10	4.2	4.5	3.2	2.5	1.1	0.5	0.5	0.5	0.4	0.4	0.3	0.67
57	-.37	-.33	.59	.63	.09	4.8	4.4	3.5	2.8	1.5	0.4	0.4	0.5	0.4	0.4	0.6	0.61

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

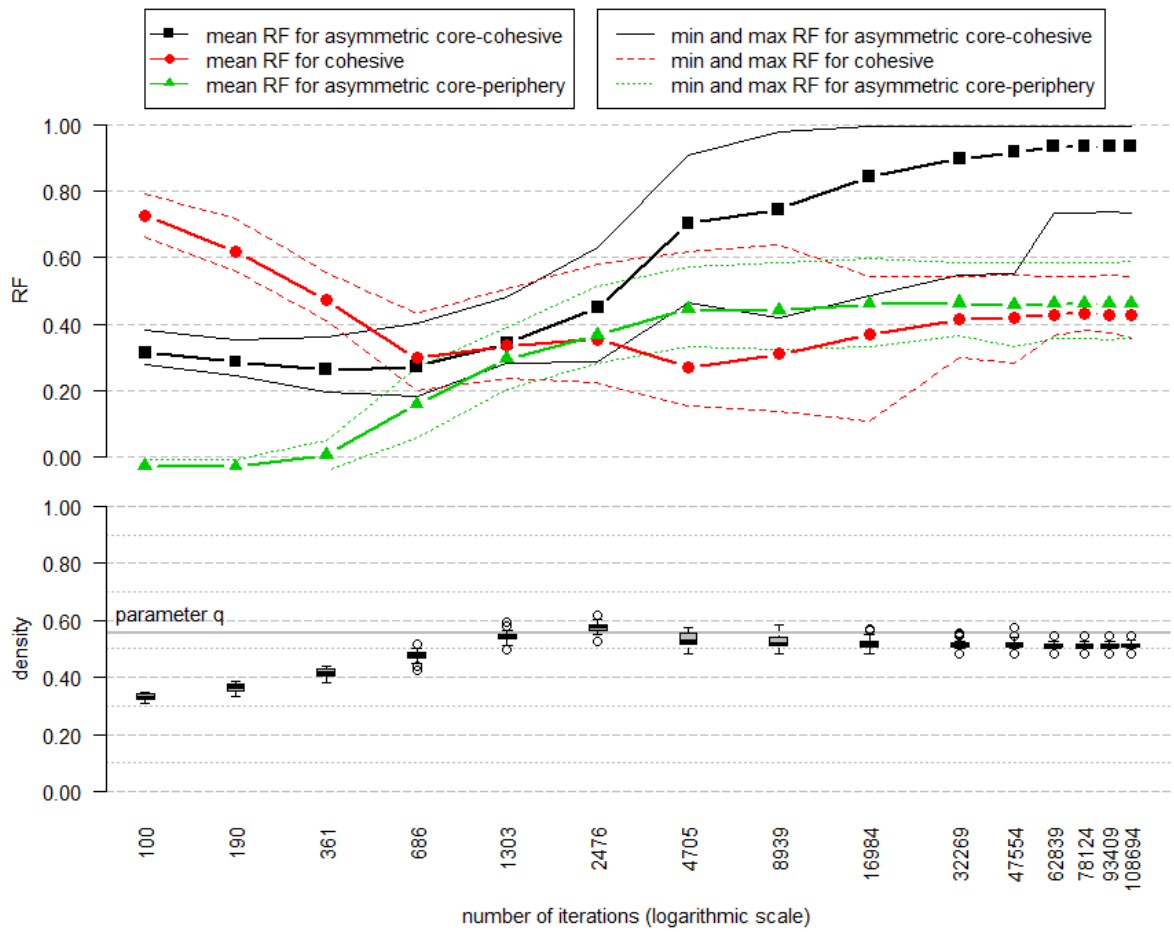
The mean number of inconsistent blocks is relatively low for the generated networks according to the mechanisms' weights given in Table 4.1. Yet the mean number is especially low for those networks generated according to the first five best θs. The mean RF values for the 108,694th iteration are similar for the networks generated by the first ten θs (see Table 4.1 and Appendix A).

The trajectories show that the mean RFs for the different blockmodel types are similar for most of the selected θs. The exceptions are the θs with ID 40, ID 57 (the mean RF values converged at approximately 4,705 iterations) and ID 153 (the mean RF values did not converge). The results for

[25] All statistics, except assortativity, are normalized in such a way that the sum over all nodes equals 1. Assortativity is normalized such that the value corresponding to each node is between 0 and 1 and the sum over all the nodes usually does not equal 1. Higher weights are typically needed in the case of the first normalization that is described to achieve the same impact.

all the selected θs are provided in Appendix A while a more detailed explanation for the θ with ID 238 is given in Figure 4.4 (the mean RF values are similar for the networks generated by the mechanism weights stated in Table 4.1).

Figure 4.4: Mean value of the relative fit for each blockmodel type and the distribution of the density of the generated networks (by considering the θ with ID 238)

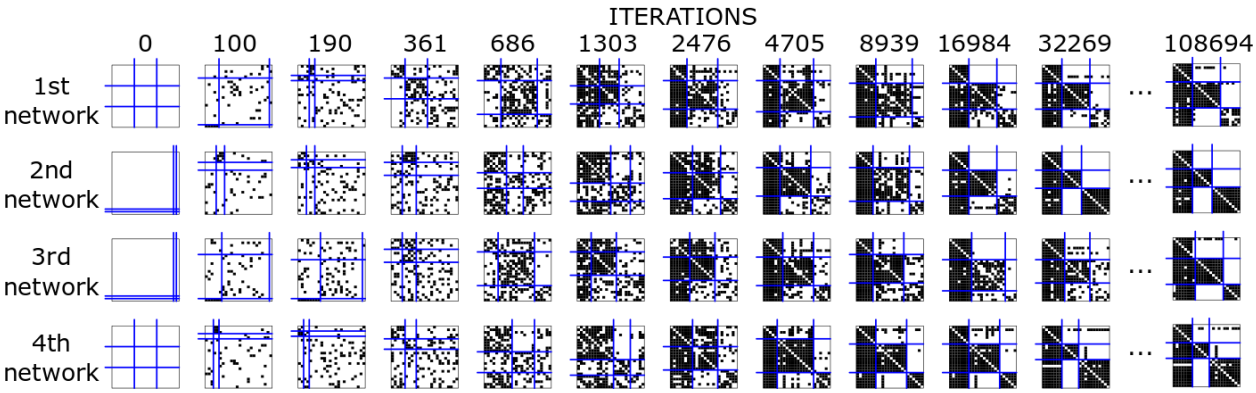


Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = 0.37, P = 0.61, A = 0.35, T = 0.00, OSP = 0.61\}$, $q = 5/9$, $d_0 = 0$. Initial networks are empty networks.

Figure 4.4 shows that the mean RF values corresponding to the asymmetric core-periphery blockmodel type are around zero until iteration 361. The mean RF values corresponding to the core-cohesive and cohesive blockmodel types are increasing with the same speed (the mean values corresponding to the cohesive blockmodel type are usually slightly higher). This suggests that cohesive clusters are formed at very early stages, before the core cluster emerges, which is not in

line with the observations about the mechanisms' strength made by Schaefer et al. (2010) that one would expect the core cluster to first emerge. However, the networks are very sparse at this stage and the mean RF values are very low (below 0.3). At between 1,303 and 2,476 iterations, all the mean RF values are relatively close. The common global network structure cannot be easily recognized by visually observing some generated networks since they look very heterogeneous. Common to all the networks at this number of iterations is that they become much denser. In some networks, this is expressed in the global network structures, which are close to the core-periphery blockmodel type with the core cluster compromising half of the nodes. In some other cases, the global network structure close to the core-cohesive blockmodel type appears. When this occurs, there are usually many links from the core cluster to the largest cohesive cluster. Later, the mean RF values for the asymmetric core-periphery blockmodel and cohesive blockmodel mainly remain constant, while the mean RF values corresponding to the core-cohesive blockmodel type are increasing. This is primarily seen in the decrease in the number of errors in null blocks.

Figure 4.5: Some networks generated (by considering the θ with ID 238) with an asymmetric core-cohesive blockmodel



Note: The networks are generated by considering θ with the highest mean value of the RF; $\theta = \{M = 0.37, P = 0.61, A = 0.35, T = 0, OSP = 0.61\}$, $q = 5/9$, $d_0 = 0$. The networks are drawn in line with the blockmodels obtained (non-specified model). Initial networks are empty networks.

4.5.2 From asymmetric core-periphery to the asymmetric core-cohesive blockmodel

It is assumed in this subsection that the initial network has an asymmetric core-periphery blockmodel (the core cluster consists of eight nodes) with 11 or 12 randomly relocated links (0.1 level of errors, see subsection 2.5.3).

The number of inconsistent blocks is low after the first 100 iterations for all selected θs. Since the blockmodel of the initial network is asymmetric core-periphery, a small number of inconsistent blocks is expected (the asymmetric core-periphery blockmodel is part of the asymmetric core-cohesive blockmodel). The number of inconsistent blocks decreases very quickly. Specifically, four θs generate networks where the mean number of inconsistent blocks is very close to zero after the first 4,705 iterations.

Table 4.2: Mean number of inconsistent blocks for the selected $\boldsymbol{\theta}$s (initial is an asymmetric core-periphery blockmodel with two clusters and target is an asymmetric core-cohesive blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS										Mean number of IB at 108,694 iterations	Mean RF at 108,694 iterations
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269		
9	.10	.29	.72	.58	.23	1.0	1.0	1.0	0.9	0.6	0.4	0.3	0.0	0.0	0.0	0.0	0.88
40	-.13	-.06	.49	.85	.10	1.4	1.0	1.0	0.6	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.86
193	-.03	.21	.44	.28	.83	1.2	1.0	1.0	0.7	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.81
226	.65	-.04	.11	.71	.23	1.1	1.1	1.0	0.8	0.5	0.3	0.3	0.1	0.0	0.0	0.0	0.90
8	.11	-.03	.42	.66	.61	1.0	1.1	1.0	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.87
57	-.37	-.33	.59	.63	.09	1.6	1.1	1.0	0.8	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.83
48	.44	.15	.26	.79	-.32	1.0	1.0	1.0	0.8	0.4	0.5	0.5	0.4	0.2	0.1	0.0	0.86
73	.01	.44	.43	.36	.70	1.0	1.0	1.0	0.9	0.7	0.5	0.2	0.1	0.1	0.1	0.1	0.86
134	-.18	.20	.60	.21	.72	1.3	1.0	1.0	0.7	0.6	0.3	0.1	0.1	0.1	0.1	0.1	0.74
202	-.28	.48	.47	.15	.67	1.8	1.1	1.1	0.8	0.8	0.2	0.3	0.2	0.1	0.1	0.1	0.80

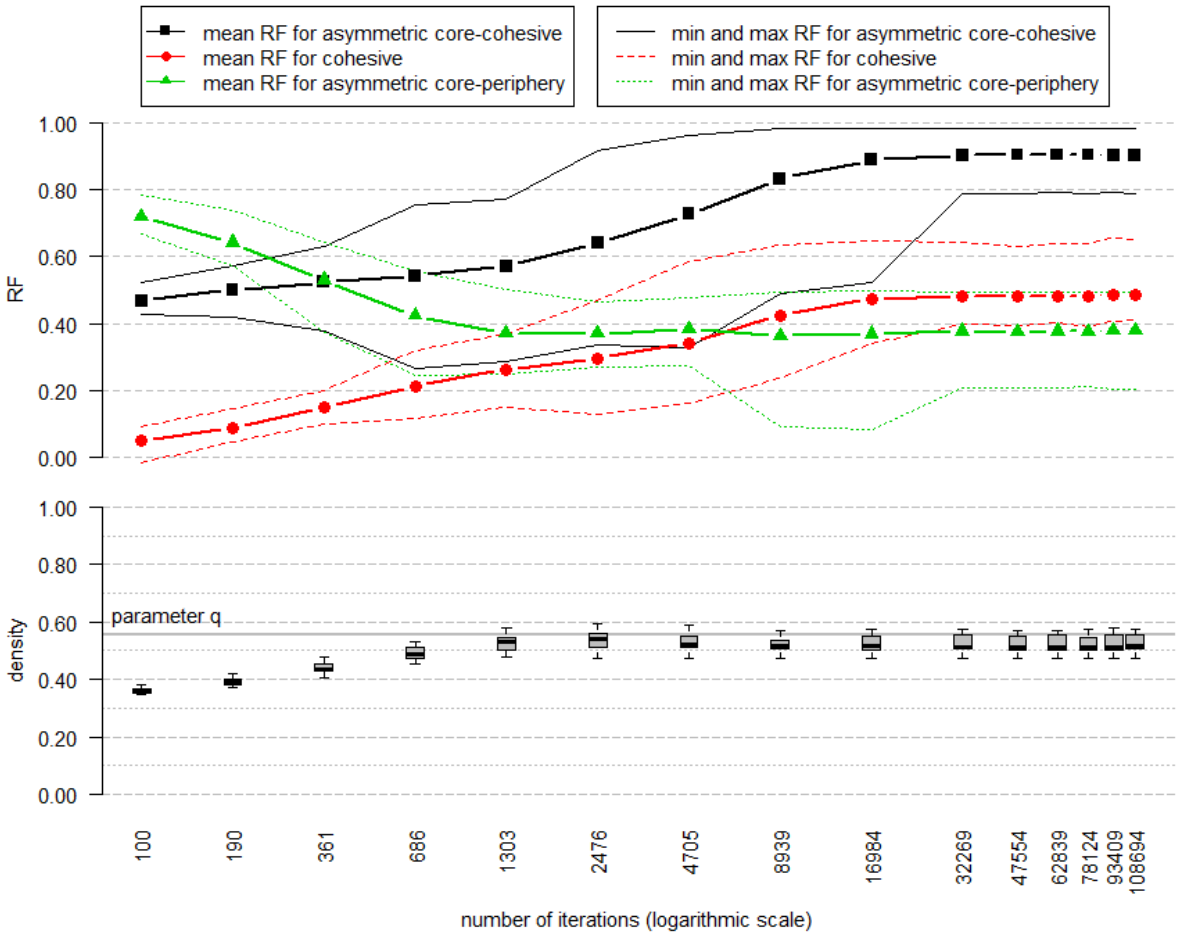
Note: The results for the best ten $\boldsymbol{\theta}$s, according to the mean number of inconsistent blocks, are shown.

The mean RF values are very similar for the networks generated by considering the best 10 θs (see Table 4.2 and Appendix A). By looking at the RF values (Figure 4.6) corresponding to the networks generated by considering the θ with ID 226 in Table 4.2, one can notice an expected decrease in the RF values with the number of iterations, corresponding to the asymmetric core-periphery blockmodel type. On the other hand, the RF values corresponding to the asymmetric core-cohesive and cohesive blockmodel types are increasing with the number of iterations. The trajectories are parallel. This is the logical implication of the emergence of the cohesive clusters in the initial asymmetric core-periphery blockmodel type.

This can be confirmed by visually observing some randomly chosen generated networks (Figure 4.7). The initial global network structure remains relatively unchanged after the first 100 iterations.

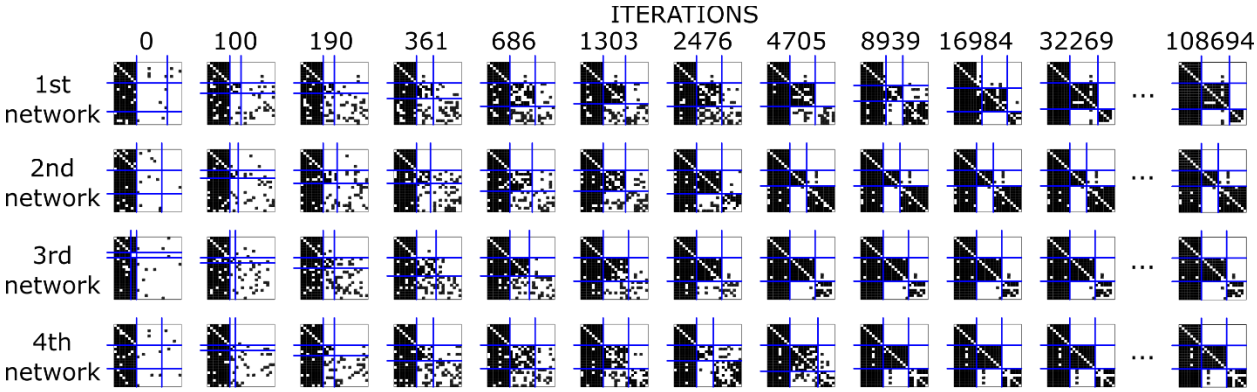
Soon, one cohesive cluster emerges, while another one (which is usually smaller) becomes more prominent at a higher number of iterations.

Figure 4.6: Mean value of the relative fit for each blockmodel type and the distribution of the density of the generated networks (by considering the θ with ID 226)



Note: The networks are generated by considering θ with the highest mean value of the RF; $\theta = \{M = 0.65, P = -0.04, A = 0.11, T = 0.71, OSP = 0.23\}, q = 5/9, \quad d_0 = 3/9$. The initial networks are networks with an asymmetric core-periphery blockmodel (with some number of errors).

Figure 4.7: Some randomly selected generated networks (by considering the θ with ID 226) with an asymmetric core-cohesive blockmodel



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = 0.65, P = -0.04, A = 0.11, T = 0.71, OSP = 0.23\}$, $q = 5/9$, $d_0 = 3/9$. The networks are drawn in line with the blockmodels obtained (non-specified model). Initial is a network with an asymmetric core-periphery blockmodel (with some number of errors).

4.5.3 From a cohesive blockmodel to an asymmetric core-cohesive blockmodel

Here, the initial networks have a cohesive blockmodel structure with 3 clusters and approximately 11 randomly relocated links. Each cluster consists of 8 nodes. The mean number of inconsistent blocks given in Table 4.3 is 2 until the 361st iteration for all best ten θs that generate networks with the lowest mean value of the number of inconsistent blocks at the 32,269th iteration. This is expected, as explained in the previous section (the asymmetric core-periphery blockmodel and the asymmetric core-cohesive blockmodel differ in two blocks). The fact that the mean number of inconsistent blocks is decreasing throughout all iterations (for many θs), instead of stabilizing at higher iterations, indicates the global network structures generally did not converge. The number of iterations is therefore increased, when calculating the RF values, to 108,694.

The trajectories for the mean RF values corresponding to different blockmodel types are given in Figure 4.8 for the θ with ID 238 in Table 4.3. It can be seen that, after the first 100 iterations, the mean RF for the cohesive blockmodel type is the highest, as expected. Following 686 iterations, the values decrease while at a higher number of iterations they remain relatively stable. On the other hand, the mean RF values for the asymmetric core-periphery blockmodel type are increasing with the number of iterations – mostly between iteration 361 and 4,705. The mean RF for the asymmetric core-cohesive blockmodel type usually converges at around 62,839 iterations. This pattern is similar for the other 10 selected mechanisms' weights (see Appendix A).

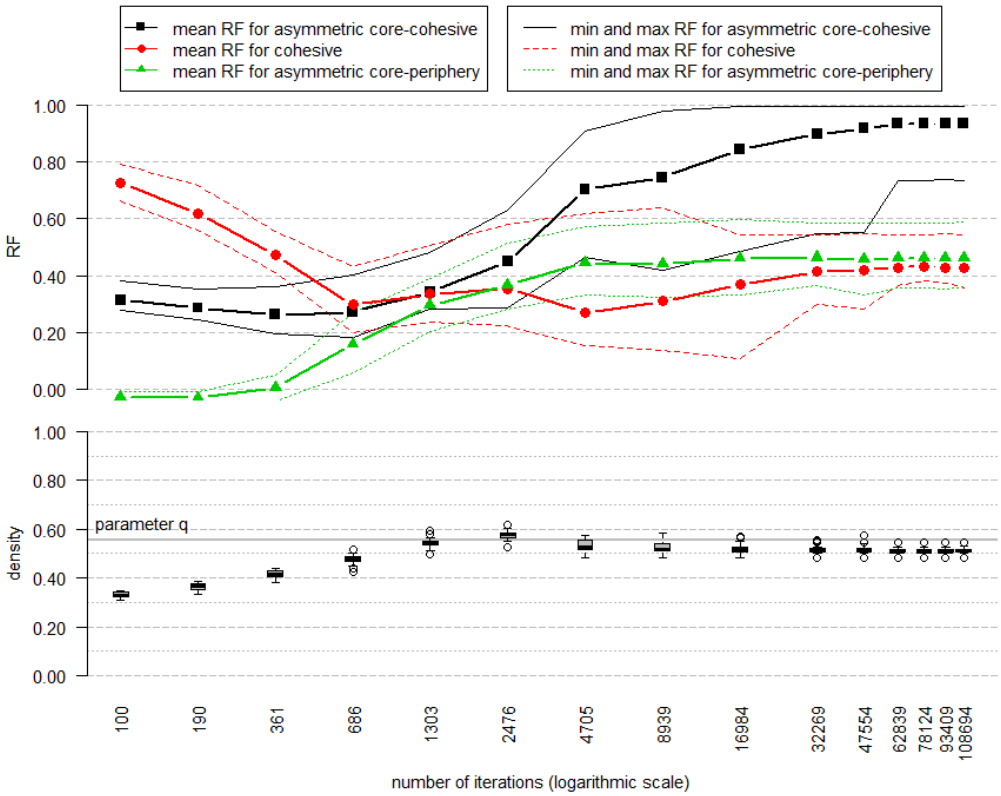
When looking at some generated networks (Figure 4.9), it becomes clear that the asymmetric core-cohesive blockmodel type appears after 8,939 iterations in most networks. Between 361 and 2,476 iterations, there is no recognizable global network structure for three clusters. This transition phase is common for many of the other selected θs.

Table 4.3: Mean number of inconsistent blocks for the selected θs (initial is a cohesive blockmodel with three clusters and target is an asymmetric core-cohesive blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS										Mean number of IB at 108,694 iterations	Mean RF at 108,694 iterations
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269		
238	.37	.61	.35	.00	.61	2.0	2.0	2.0	1.8	1.4	1.0	0.5	0.3	0.1	0.1	0.0	0.93
136	.41	-.18	.37	.74	-.34	2.0	2.0	2.0	1.6	0.9	0.7	0.6	0.2	0.2	0.2	0.1	0.86
228	.59	.12	.24	.76	-.08	2.0	2.0	2.0	1.9	1.5	0.9	0.5	0.3	0.2	0.2	0.0	0.84
224	.45	.10	.77	.44	-.08	2.0	2.0	2.0	1.9	1.0	0.9	0.4	0.3	0.2	0.2	0.0	0.88
153	.84	-.14	.36	.32	-.21	2.0	2.0	2.0	1.7	0.8	0.5	0.3	0.6	0.6	0.2	0.2	0.72
286	.52	.26	.33	.72	.16	2.0	2.0	2.0	2.0	1.5	1.3	0.6	0.3	0.4	0.3	0.3	0.77
226	.65	-.04	.11	.71	.23	2.0	2.0	2.0	2.1	2.8	1.3	0.7	0.3	0.4	0.3	0.2	0.89
259	.51	.57	.60	.25	.02	2.0	2.0	2.0	1.6	1.1	1.3	0.5	0.7	0.5	0.3	0.2	0.73
163	.44	.15	.49	.41	.61	2.0	2.0	2.0	2.2	1.6	1.3	0.7	0.5	0.4	0.4	0.3	0.87
172	.28	.80	.17	-.01	.51	2.0	2.0	2.0	1.9	1.6	1.5	1.0	0.8	0.7	0.5	0.4	0.79

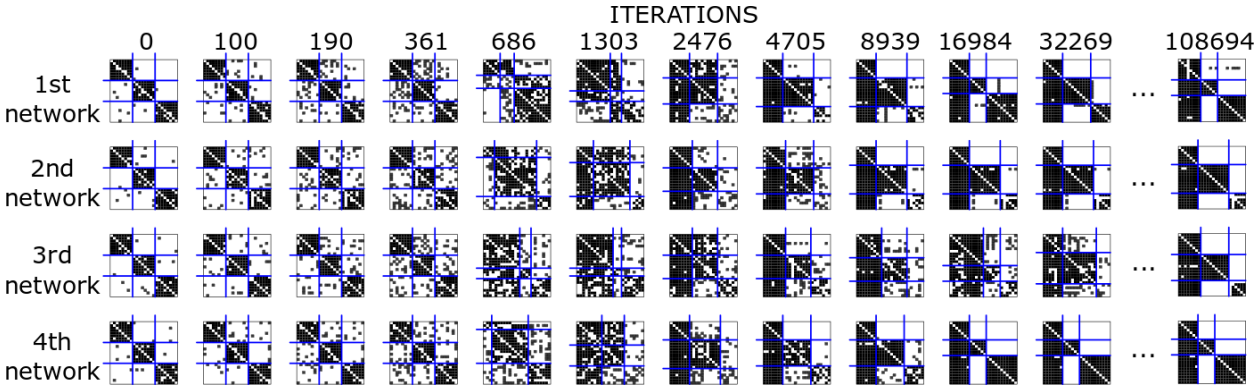
Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

Figure 4.8: Mean value of the relative fit for each blockmodel type and the distribution of the density of the generated networks (by considering the θ with ID 238)



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = 0.37, P = 0.61, A = 0.35, T = 0.00, OSP = 0.61\}, q = 5/9, \quad d_0 = 3/9$. The initial networks are networks with the cohesive blockmodel type (with some number of errors).

Figure 4.9: Some randomly selected generated networks (by considering the θ with ID 238) with an asymmetric core-cohesive blockmodel



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = 0.37, P = 0.61, A = 0.35, T = 0.00, OSP = 0.61\}$, $q = 5/9$, $d_0 = 3/9$. The networks are drawn in line with the blockmodel obtained (non-specified blockmodeling). Initial is a cohesive blockmodel (with some number of errors).

4.6 Discussion on the mechanisms

In order to determine the sets of mechanism weights that generate networks with the global structure closest to the asymmetric core-cohesive blockmodel, 300 different sets of mechanisms weights were randomly selected. The same values were used to generate the networks for different initial global structures. Based on the number of inconsistent blocks, the ten best sets of mechanism weights were chosen for each initial structure. Among the ten best chosen sets of mechanism weights for each initial structure, only 19 different sets of mechanisms' weights were obtained. More specifically, the sets of mechanism weights are very similar when the initial network is an empty network or a network with a cohesive blockmodel yet are very different when the initial network is an asymmetric core-periphery blockmodel. This is shown by a Venn diagram (Figure 4.10) which visualizes the number of common selected θs (the selected θs are shown in Table 4.1, Table 4.2 and Table 4.3).

When the initial network is a cohesive or empty network, one can generate networks with a global network structure sufficiently close to the asymmetric core-cohesive blockmodel type by considering eight different sets of mechanism weights (Figure 4.10). The asymmetric core-periphery initial network structure shares only one common set of mechanism weights with the cohesive initial global network structure and two sets of mechanism weights with the empty

network as the initial network. The seven sets of mechanism weights, which were selected only where the global network structure of the initial networks was asymmetric core-periphery, are strongly characterized by high values for the transitivity and/or OSP mechanisms, which is expected since they promote the forming of cohesive clusters while the most popular one already exists. In brief, the same set of mechanism weights can lead to a network with the same global network structure, especially when the initial global network structure is a cohesive or empty network.

Figure 4.10: Number of the best (based on the mean RF value) common selected θs according to the initial networks with three different global structures

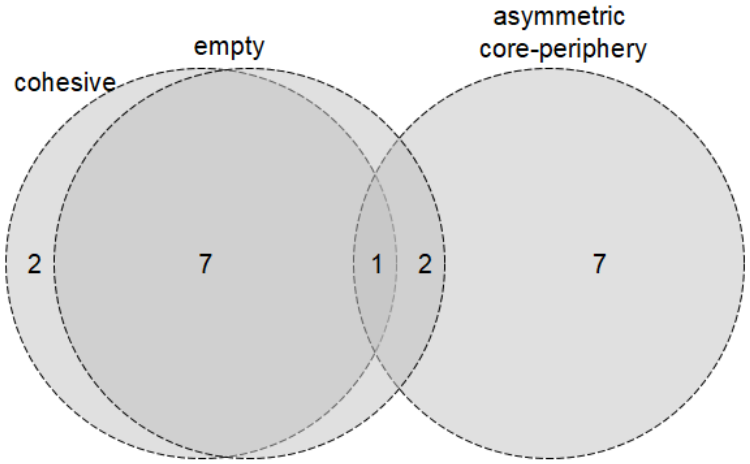


Figure 4.11 illustrates the associations between different weights for the five mechanisms considered. The same 19 chosen sets of mechanism weights are presented as described in the previous paragraph. The sizes of the points are proportional to the mean RF values. Since most of the mean RF values are very similar (and high), the sizes of the points are similar. Different colours indicate the different global structures of the initial networks.

A relatively strong linear negative correlation is observed between the popularity and transitivity mechanisms. This indicates that when one mechanism has a high weight, the other becomes less important for the asymmetric core-cohesive blockmodel to appear. In the case of undirected networks, the transitivity mechanism is generally known for being responsible for the formation of cohesive clusters, yet with the directed networks considered in this study it also promotes links from the nodes of cohesive clusters to the nodes of the core cluster. Let us assume that node j (in

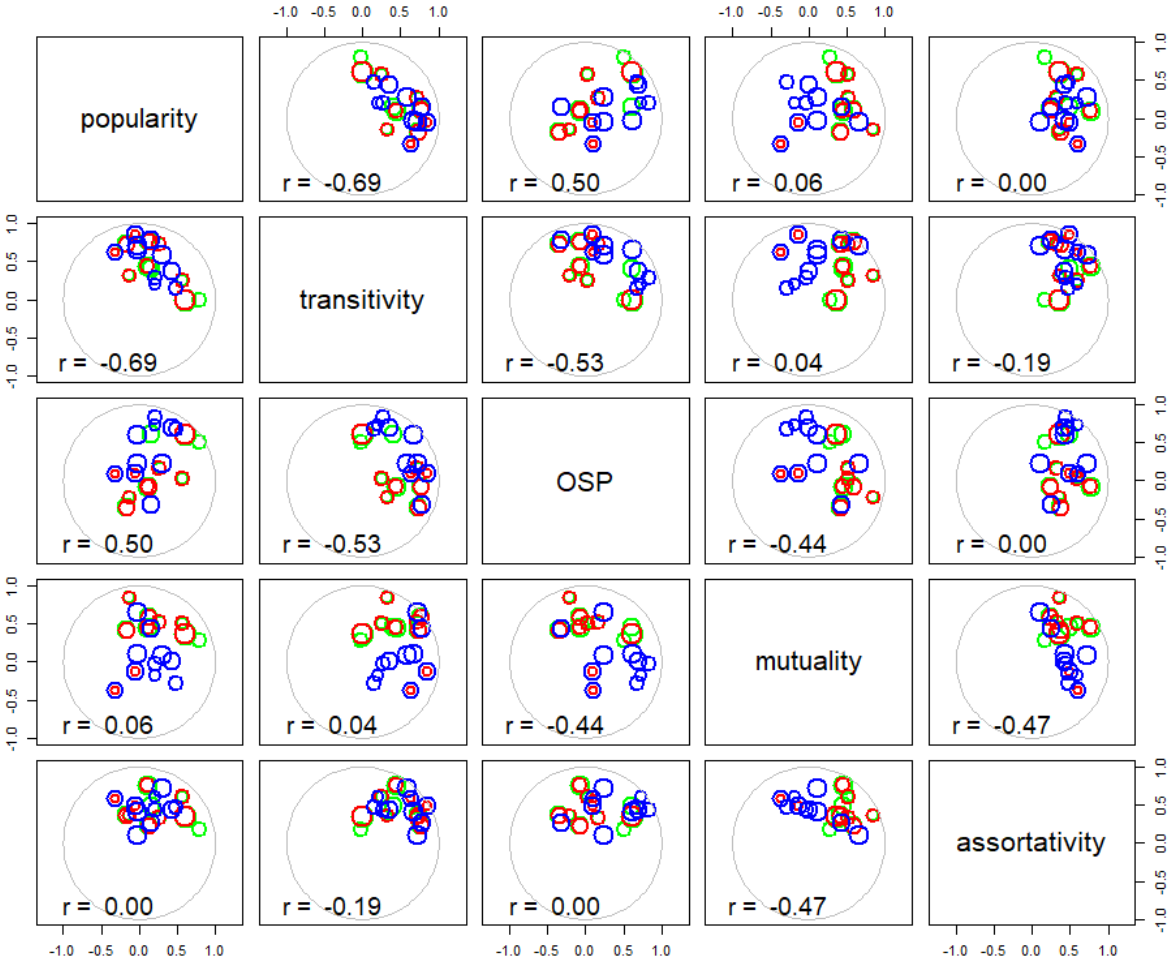
Figure 4.2, transitivity) is a node from the core cluster and nodes i and k are from the same cohesive cluster (a reciprocated or non-reciprocated link exists between them). In such a case, when the transitivity mechanism is applied a link from i to j would be promoted (if $i \rightarrow j$ exists), that is, a link between a node from the cohesive cluster to a node from the core cluster. The OSP mechanism, on the other hand, is also related to the forming of cohesive clusters, but does not promote the establishing of links to the nodes of the core cluster (since that is a symmetrical characteristic). Therefore, when the OSP mechanism is strong, the popularity mechanism is usually also strong. Based on the described principles, the correlation between the popularity mechanism and the OSP mechanism is expected to be positive.

The mutuality mechanism is considerably (negatively) correlated only with the OSP mechanism and assortativity. Since OSP is a very symmetric mechanism, node j can be selected in the proposed algorithm in a specific iteration while, in another iteration, node j can be randomly selected. In both cases, a link from one node to another node would be promoted if they have a sufficient number of links to the same others. This eventually results in a symmetric link $i \leftrightarrow j$ and the additional mutuality mechanism therefore becomes less important for the asymmetric core-cohesive blockmodel type to appear.

The given interpretation is only valid while generating networks with the asymmetric core-cohesive blockmodel type. The results would be different when generating networks with different global network structures.

Depending on the initial network's global structure, one mechanism or the other might be more or less efficient in generating the asymmetric core-cohesive blockmodel type. For example, when the initial network is cohesive, the popularity mechanism may be more important since the links within cohesive clusters are already formed. On the contrary, when the global network structure of the initial network is asymmetric core-periphery, the shared partners' related mechanisms, i.e. the transitivity and OSP mechanisms, are important for forming cohesive clusters.

Figure 4.11: The associations between the mechanisms' weights



Note: Only the 19 selected θs are shown. The points are coloured according to the global structures of the initial networks (red = empty network; green = network with a cohesive blockmodel; blue = network with an asymmetric core-periphery blockmodel). The sizes of the points are proportional to the mean RF values.

4.7 Conclusion

The relationship between the most common local network mechanisms (which were also found to be present in friendship and liking networks among preschool children), and the asymmetric core-cohesive blockmodel type is studied in this chapter. Understanding such a relationship is necessary for studying networks that are empirically obtained.

The local network mechanisms are selected based on previous studies on the network evolution of the liking, popularity and friendship networks among children in kindergarten. The selected mechanisms are: popularity (the tendency to create links to more popular ones), transitivity (the tendency to create links with those with whom a higher number of friends is shared), outgoing shared partners (the tendency to create links to those who "like the same others") and assortativity (the tendency to create links to others with the same level of popularity, i.e. in-degree).

The proposed global network structure (the asymmetric core-cohesive blockmodel) is assumed to emerge in such networks. It is a combination of two well-known blockmodel types, namely cohesive and asymmetric core-periphery.

The research question is addressed by employing the proposed algorithm, which simulates the evolution of a network based on the weights of the selected local network mechanisms. Different global network structures are used for the initial networks (empty network, network with a cohesive blockmodel and network with an asymmetric core-periphery blockmodel) and 300 random sets of mechanisms' weights are generated. For each set of weights and each starting global structure, 30 networks are simulated. To evaluate the obtained global network structure, the number of inconsistent blocks, where the blockmodel structure that is obtained is compared with the ideal asymmetric core-cohesive blockmodel type, is considered. The relative fit function is also applied for each generated network structure. Based on these two criteria, the sets of mechanisms' weights that lead to the desired global network structure are evaluated for each initial network.

The main finding is that the selected five mechanisms with different mechanisms' weights can lead the networks to the asymmetric core-cohesive blockmodel. This is true for each initial global network structure that is considered. Yet, the sets of mechanism weights are very similar if the initial network is empty or has a cohesive network structure. The weights are different when the initial network has an asymmetric core-periphery structure. The findings are especially relevant because in this dissertation it is assumed that the only information that nodes have about the other nodes is their position in the network.

Over the course of an evolutionary process, it takes time for a blockmodel to become apparent. When the asymmetric core-cohesive blockmodel is the final one, the time (i.e. number of iterations) needed to reach this blockmodel type is longer when the initial network is a cohesive

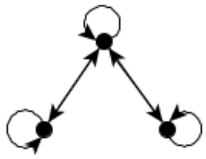
blockmodel compared to the other two initial global network structures (empty network and asymmetric core-periphery blockmodel). Further, “transition” global network structures might appear in some cases. For example, when the initial network is empty, a network structure close to, for example, the cohesive blockmodel, might appear after some iterations. When studying empirical networks, the researcher usually does not know at which stage of the network evolution the data were collected. This confirms that studies such as the one in this chapter are important for the specific context under examination.

The methodology that is applied can be used to study the relationships between other global network structures and local network mechanisms. It is important that they are carefully chosen according to the context of the study, as done, e.g. in Chapter 7 of this dissertation, where the social context of knowledge-flow is taken into account. Furthermore, the proposed algorithm for generating networks could be adjusted or extended by making different assumptions (e.g. in such a way that the mechanism weights are treated as variable in time or across the nodes, or that valued networks or growing networks are generated).

5 Emergence of the symmetric core-cohesive blockmodel

The symmetric case of the core-cohesive blockmodel is studied in this chapter. A symmetric core-cohesive blockmodel (see Figure 5.1) consists of one core cluster of nodes to which all nodes in the network are all linked and where nodes from the core cluster are linked to all other nodes in the network. The other nodes are classified in cohesive clusters. Nodes from each cohesive cluster are internally linked to each other, while the nodes from different non-core clusters are not linked to each other. The model can be extended in such a way that a cluster of nodes which are not linked to each other inside the cluster would also exist. The difference between the symmetric and asymmetric core-cohesive blockmodel type is that with the symmetric core-cohesive blockmodel the nodes from the core cluster and the nodes from cohesive clusters are symmetrically linked.

Figure 5.1: A symmetric core-cohesive blockmodel with three clusters



com	com	com
com	com	null
com	null	com

This chapter consists of two parts. In the first part, the assumption that the proposed global network structure emerges among preschool children is empirically tested. For this purpose, the symmetrized networks previously analysed by Schaefer et al. (2010) are analysed using the blockmodeling approach (Doreian et al., 2005). Even though the focus of this dissertation is more on the local network mechanisms without considering the attributes of the nodes, the results of the current chapter raise very important developmental questions, e.g. how the nodes from the core cluster differ from those from cohesive clusters and which implications (if any) does this hold for their further individual development? Should such global network structure be encouraged or discouraged?

The second part of this chapter indirectly addresses the second research question of this dissertation. By using Monte Carlo simulations, we test whether the symmetric core-cohesive blockmodel type can emerge from the selected local network mechanisms. Here, the same local network mechanisms are assumed as in the previous chapter that studied the emergence of the asymmetric core-cohesive blockmodel.

One of such mechanisms that is included is the mutuality mechanism. The reason for considering this mechanism in the setting of interactional networks is that the social process underlying it is in fact asymmetric. For example, there is always one node i which is an initiator of interaction with node j. Node j can respond in different ways. One way is by engaging in further interactions with node i or avoiding (or rejecting, i.e. running away) interactions with node i. The networks generated by the NEM are later symmetrized because we wish to simulate: (i) the emergence of the global network structure (which can be either symmetric or asymmetric); and (ii) the process of collecting (observing) the network data. The network data (in the empirical data analysed in this chapter) do not provide information on who is the initiator of a given interaction. Therefore, the empirical network is symmetrized and the same is done for the generated networks by the NEM algorithm used in this chapter.

5.1 The empirical case

The hypothesis that a symmetric core-cohesive blockmodel is present in empirical interactional networks is tested in the subsections below. To this end, the empirical data collected among preschool children are analysed using blockmodeling.

5.1.1 Data

The data were collected as part of a bigger longitudinal study of young children's preparedness for school between 2004 and 2006 in Head Start preschools. These data were also analysed in a study by Schaefer et al. (2010). The data are observational in nature, meaning that trained observers present in school classes recorded interactions among the children. Specifically, observers were present for several hours in a classroom two to three days per week. To ensure the greatest validity and reliability, two observers monitored the same children at the same time for 10 seconds. The order in which the observers watched over the children was random. When all children had been observed, the observers waited 5 minutes before repeating their observations (with a randomly reordered list of children). Children were observed engaging in different activities, e.g. free play, talking, aggressive behaviour, and others. The observers coded the type of activity in which a given child was involved and up to five other children with whom the selected child was interacting. Only the free-play data (data collected when children were able to play freely) are analysed in this study. Children had to be observed at least 13 times during the whole school year to be included

in the analysis. Based on the observational data, four complete networks were generated for each class.

Table 5.1: Some descriptive statistics for the undirected interactional networks (from September 2004 to May 2005)

Class	Number of children				Number of observations				Age span in the last period (in months)	Share of males in the last period
	Sep-Oct 2004	Nov-Dec 2004	Feb-Mar 2005	Apr-May 2005	Sep-Oct 2004	Nov-Dec 2004	Feb-Mar 2005	Apr-May 2005		
ID1	21	20	20	19	814	510	484	321	42-58	63
ID2	17	17	15	14	57	95	236	374	48-59	50
ID3	16	17	14	14	75	200	190	184	50-58	50
ID4	17	18	18	16	104	410	525	548	49-55	69
ID5	17	17	14	14	280	406	862	413	37-57	50
ID6	15	15	14	14	202	343	1005	510	46-59	43

Table 5.2: Some descriptive statistics for the undirected interactional networks (from September 2005 to May 2006)

Class	Number of children				Number of observations				Age span in the last period (in months)	Share of males in the last period
	Sep-Oct 2005	Nov-Dec 2005	Feb-Mar 2006	Apr-May 2006	Sep-Oct 2005	Nov-Dec 2005	Feb-Mar 2006	Apr-May 2006		
ID7	21	19	17	16	594	564	196	589	46-60	44
ID8	18	18	16	16	396	432	273	855	43-58	69
ID9	18	18	16	15	663	406	368	1237	37-59	40
ID10	16	15	15	14	931	496	309	1609	39-60	64
ID11	15	16	15	15	172	241	395	574	48-60	47

Each network's construction is based on a two-month period, as presented in Table 5.1 and in Table 5.2. The networks are in matrix form in which each row and each column represents a child. The number of a given child's (in a given row) observed interactions with other children (in a given column) is shown in the corresponding matrix cells. The obtained networks were transformed from directed to undirected and binarized: there is a link between two children if the number of observed interactions is higher than the median (of the number of interactions between all possible pairs in the network which include non-links) divided by two.

The number of children varies between 14 and 21 across all networks. In the last period, the children were aged between 37 and 60 months and the share of males varied between 43% and 69%.

5.1.2 Methodology

Binarized networks are blockmodeled to evaluate the global network structure[26]. Since structural equivalence is considered, the possible block types are null and complete. In order to not constrain the blockmodeling procedure, the relationships between clusters (image matrix) are not pre-specified.

The blockmodeling is performed using the "blockmodeling" package (Žiberna, 2018) for the R programming language. The number of random restarts in the blockmodeling procedure is 500 and 3 clusters are set for all networks[27].

5.1.3 Results: empirical blockmodels

Figure 5.2 gives the matrix representation of the analysed networks. Each matrix corresponds to one network at a given time point. Black dots denote links. Children are ordered by rows and by columns in line with the solution from the blockmodeling. It can be seen that the networks are very dense, which is expected since interactional networks were observed in a closed environment (classroom). Some are almost complete.

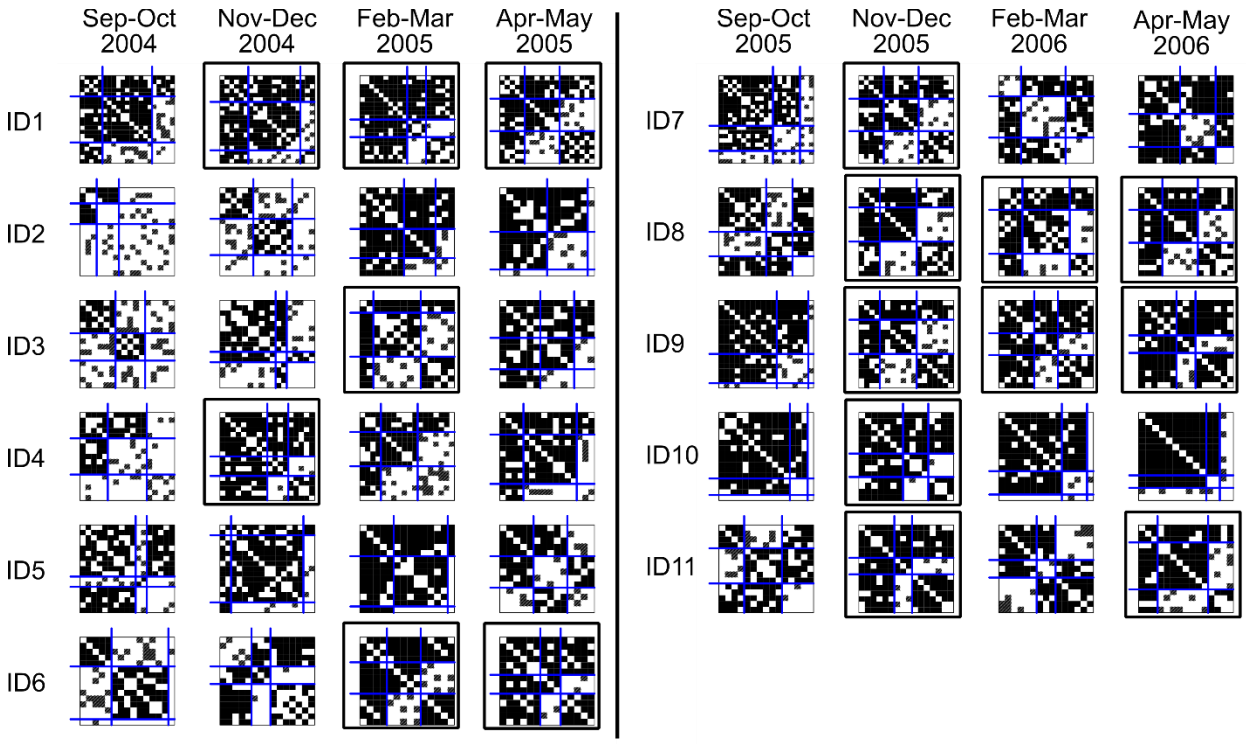
A symmetric core-periphery blockmodel structure (see the framed matrices) appears in almost all classes in at least one time period. It does not appear in just two classes (ID2 and ID5) out of 11 classes. In the other classes, the symmetric core-cohesive blockmodel appears in the 2nd time

[26] The results of binarized networks are shown because the focus in this study is narrowed to the relationship between local network mechanisms and global network structures in binary networks. Nevertheless, the preliminary analysis shows similar results would appear if the valued networks were to be analysed. This is true for these particular empirical analysed networks, but in general it is suggested using blockmodeling for valued networks in order to consider as much information as possible.

[27] The preliminary analysis showed that the blockmodels' fit is the most optimal when the three clusters are specified for most networks, which entails a clear global network structure. Further analysis showed the clusters can be highly explained by gender. Two clusters are usually gender-homogenous while the third cluster is gender-heterogenous. In the case of some networks two clusters would be more appropriate, as seen in Figure 5.2 (e.g. the class with ID 6 from the 1st time point).

period (in 7 classes out of 11) or in the 3rd and 4th time periods (in 5 classes out of 11). The cluster sizes vary – in some cases, the core cluster consists of only 2 children (ID3 in Feb-Mar 2005) while in other cases the core cluster consists of almost half the children (e.g. ID4 in Nov-Dec 2004 and ID43 in Nov-Dec 2005).

Figure 5.2: Obtained blockmodel structures for each preschool class and each time period



Note: The classes are in rows denoted by IDs from ID1 to ID11. Undirected and binarized empirical networks are considered. The obtained symmetric core-cohesive blockmodels are presented in the framed matrices.

Some of the obtained blockmodels are similar to the symmetric core-cohesive blockmodel type but are without links within the core (e.g. ID8 and ID11 in Sep-Oct 2005) or without one cohesive cluster (ID3 in Apr-May 2005).

It has been shown that the proposed symmetric core-cohesive blockmodel type appears in empirical networks; specifically, in interactional networks collected among preschool children. The question of whether the most commonly studied local network mechanisms can lead the global network structure to the symmetric core-cohesive is addressed in the next section. The attributes of the children are not considered in this study.

Although the selected mechanisms are very common in studies of network dynamics among preschool children, it is not assumed that they play a role in forming the global network structures in the empirical networks presented in this section.

5.2 Simulation approach

To evaluate whether the selected local network mechanisms can lead the global network structure to the symmetric core-cohesive blockmodel, a similar methodology is applied as in the last chapter. The differences are described in subsection 5.2.1 while the results are provided in subsection 5.2.2.

5.2.1 Simulation design

Compared to the asymmetric core-cohesive blockmodel, a symmetric core-cohesive blockmodel may be generated in several ways by considering different local network mechanisms. Two distinct approaches are identified with regard to whether symmetric or asymmetric links are generated:

1. **Symmetric (non)links:** here, it is assumed that all asymmetric links are reciprocated immediately. This means that a symmetric tie will exist if at least one of the actors chooses that tie and will not exist if at least one of the actors does not want it. The reciprocity mechanism is not considered in this case.
2. **Asymmetric links:** only asymmetric links can be formed at a time. To achieve symmetric networks:
    a. the reciprocity mechanism must be considered. Here, a symmetric tie will exist if both actors choose the tie and will not exist if neither actor wants it (an asymmetric link will exist if only one chooses the tie); and
    b. the reciprocity mechanism does not necessarily have to be considered, but the networks must be symmetrized before being further analysed. This means that a symmetric tie will exist if at least one actor chooses the tie and will not exist if neither actor wants it.

The observed interactional networks are symmetric by the definition of "interaction", although the process that initiates interactions is asymmetric. In such a process, an ego has to initiate an interaction while an alter can either: (i) accept (and reciprocate), (ii) tolerate, or (iii) reject (i.e. actively avoid) interaction. Even interactions actively rejected by the alter can be observed and

recorded by the observer, although one is more likely to be recorded if it is either accepted or tolerated (because such interactions might, but not necessarily, last for a longer amount of time).

Therefore, the approach where asymmetric ties are formed (by considering the mutuality mechanism) and the network is symmetrized before being further analysed is the closest representation of the emergence of empirical networks. The algorithm for generating networks is therefore the same as that proposed in subsection 4.4.1. Because the generated networks are symmetrized before being further analysed, the global network structures of the generated networks must be re-evaluated (by using the concept of inconsistent blocks and the concept of relative fit) by considering the target symmetric core-cohesive blockmodel. This can result in different selected θs.

The local network mechanisms are operationalized in the same way as in subsection 4.4.2. The considered local network mechanisms are as follows: the mutuality mechanism (M), alter popularity mechanism (P), assortativity mechanism (A), transitivity mechanism (T) and outgoing shared partners mechanism (OSP). The parameter q (the fundamental mechanism which operationalizes the tendency of having a link) is also considered. The value of parameter q is set, with reference to generating networks with an asymmetric core-cohesive blockmodel, to $5/9$. The asymmetric core-cohesive blockmodel is used as a reference because it is assumed that the underlying process of initializing interactions is essentially asymmetric.

To generate the networks using the proposed NEM algorithm and by considering the selected local network mechanisms, the same 300 sets of weights of the local network statistics (θs) are relied on as in the case of the asymmetric core-cohesive blockmodel.

Thirty networks are generated for each θ. Blockmodeling for binary networks (on symmetrized generated networks, structural equivalence is used) is performed after the selected number of iterations of the algorithm. More precisely, the intermediate number of iterations m, at which the global network structure is analysed, is determined as $m_i = m_{i-1} * 1.9$, where $m_1 = 100$. This approach is used since most changes in the structure of the links happen at a lower number of iterations. A total of 116,490 iterations is applied.

Based on the blockmodeling solution, the number of inconsistent blocks is calculated and used as the fit function. Some θs that generate networks with the lowest number of inconsistent blocks are selected and further analysed by the RF function.

5.2.2 Generated networks

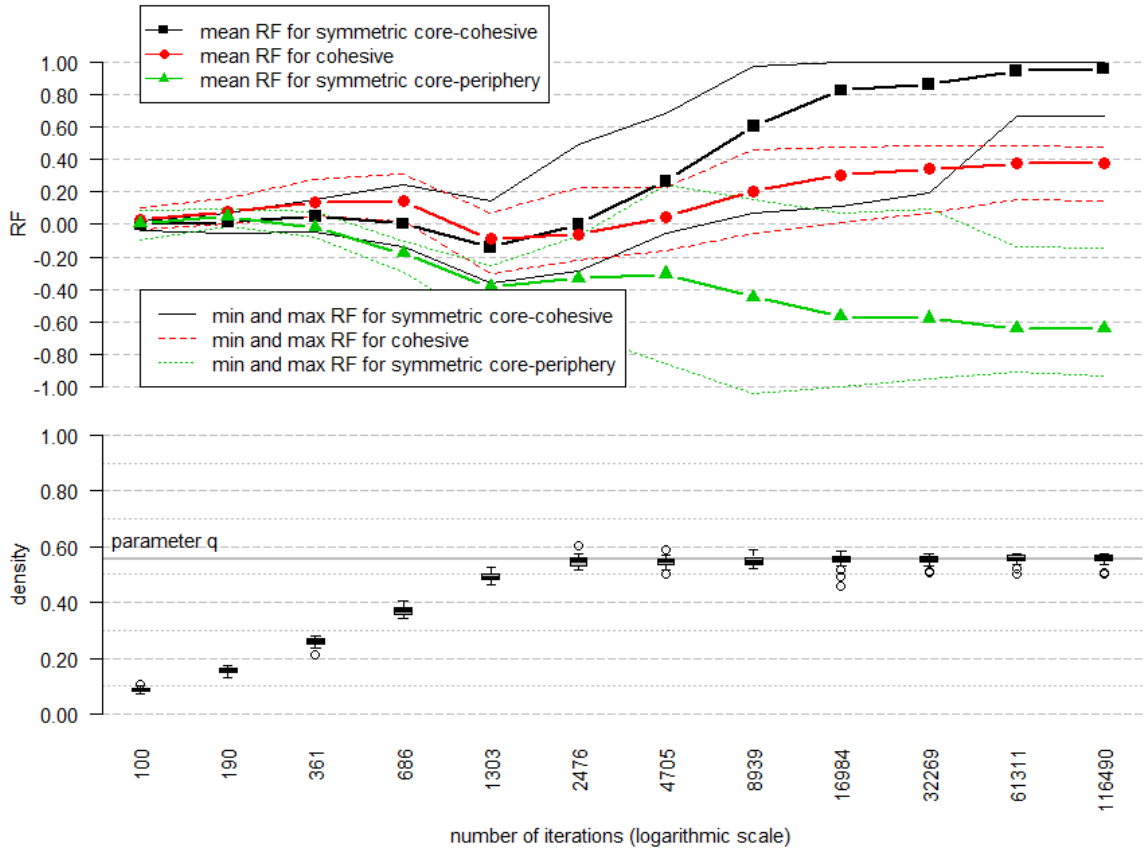
There are 6 different θs generating networks without any inconsistent block at the end of the iterations. Further, 76 different θs generate networks with the mean number of inconsistent blocks less than or equal to 0.5 and 109 different θs generate networks with the mean number of inconsistent blocks less than or equal to 1.

Table 5.3: Mean number of inconsistent blocks for those θs that generated networks with zero (mean number) inconsistent blocks at the end of the iterations (initial is an empty network and target is a symmetric core-cohesive blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS												Mean RF at 116.490 iterations
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269	61.311	116.490	
136	-.18	.74	.37	-.35	.42	4.9	5.0	4.4	3.3	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.96
25	-.43	.27	.66	.25	-.50	4.9	4.8	3.6	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.80
279	.17	-.11	.43	.60	.65	4.7	4.9	3.4	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.50
248	.11	-.58	.49	.78	-.38	4.7	5.0	4.0	1.4	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.48
72	-.57	.68	.04	-.46	.10	4.9	5.0	4.1	3.8	0.7	0.2	0.0	0.0	0.1	0.0	0.0	0.0	0.35
22	-.24	-.51	.21	-.21	-.78	5.0	5.1	4.5	2.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.26

The θs that generate each network with a symmetric core-cohesive blockmodel are shown in Table 5.3 along with the number of inconsistent blocks at a different number of iterations and the mean RF value of the generated networks. Although all the generated networks have the same blockmodel, they differ largely in the level of errors, expressed by the RF.

Figure 5.3: Mean value of the relative fit for each blockmodel type and the distribution of the density of the generated networks (by considering the θ with ID 136)

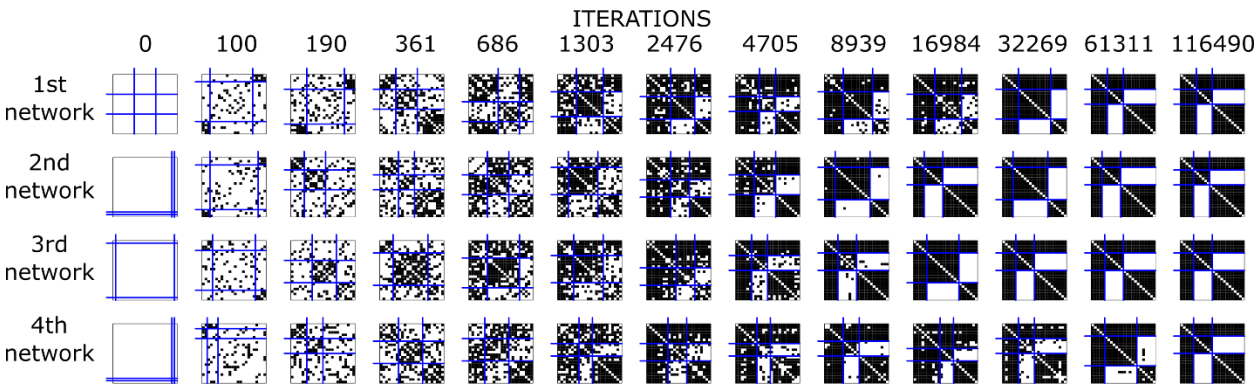


Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = -0.18, P = 0.74, A = 0.37, T = -0.35, OSP = 0.42\}$, $q = 5/9$, $d_0 = 0$. Initial networks are empty networks.

A more detailed insight into RF for a selected θ (with ID 136) is given in Figure 5.3. The mean RF values are calculated for the symmetric core-cohesive blockmodel type, cohesive blockmodel type and symmetric core-periphery blockmodel type. All RF values are close to zero at the first 190 iterations. At such a small number of iterations, there are insufficient links to enable any of the considered blockmodel types to emerge. However, at 361 iterations, a global network structure, close to cohesive, can be visually recognized in the generated networks (Figure 5.4). Since there is a relatively high level of errors in null and complete blocks, the corresponding mean RF is very low. With a higher number of iterations (until 1,303 iterations), the mean RF corresponding to all considered blockmodel types is decreasing. In this step, links are established among different clusters but, in some cases, links within the core nodes are not present at this number of iterations.

Moreover, there is a high level of errors in the null and complete blocks. After 1,303 iterations, the mean RF value for the core-cohesive and cohesive blockmodel only increases up until 61,311 iterations.

Figure 5.4: Some randomly selected generated networks (by considering the θ with ID 136) with a symmetric core-cohesive blockmodel



Note: The networks are generated by considering $\theta = \{M = -0.18, P = 0.74, A = 0.37, T = -0.35, OSP = 0.42\}$, $q = 5/9$. The networks are drawn in line with the blockmodels obtained (non-specified model). The initial network is an empty network.

The mean RF, corresponding to the symmetric core-cohesive blockmodel type, is close to 1 at the end of the iterations, indicating the global network structure is the desirable one with almost no error in null and complete blocks (as confirmed in Figure 5.4). The mean RF for the cohesive blockmodel is lower while the mean RF for the symmetric core-periphery blockmodel type is highly negative, indicating that the randomized networks fit this blockmodel type much more than the networks generated by the proposed algorithm.

5.3 Conclusion

This chapter builds on what it is shown for the asymmetric core-cohesive blockmodel in the previous chapter. Specifically, it is revealed that the selected local network mechanisms can drive the global network structure towards the symmetric core-cohesive blockmodel. The mechanisms were selected based on previous studies on linking and friendship networks on the assumption that such a blockmodel type exists among pre-schoolers.

This chapter proposes the symmetric version of the core-cohesive blockmodel and tests whether the proposed blockmodel appears in interactional networks in kindergarten. Further, by applying

the methodology proposed in Chapter 4, we test the research question of whether the mutuality, popularity, assortativity, OSP and transitivity mechanisms can lead a global network structure towards the symmetric core-cohesive blockmodel.

The existence of the studied blockmodel type is evaluated using empirical data. The data were collected within a larger longitudinal study among preschool children in the USA between 2004 and 2006. The interactions among the children in classrooms were recorded and complete networks were formed. The symmetric core-cohesive blockmodel was found to be present in almost all analysed classes in at least one time period. This proves that the proposed global structure (blockmodel type) is relevant for such data.

To address the research question concerning the emergence of the proposed blockmodel, the algorithm proposed in Chapter 4 was used to generate networks by considering the local network mechanisms. The generated networks were symmetrized before being further analysed, which reflects the nature of the interactions among children and the absence of information on who initiated the interactions. The results of the Monte Carlo simulations confirm that the selected mechanisms are able to generate networks with the symmetric core-cohesive blockmodel. The results do not imply that the global network structures of the empirical preschool networks collected in the 11 classes in the USA emerged due to the studied local network mechanisms. To address this question, a different methodology should be applied.

The study is important in several ways, given that understanding the (emergence of) peer network structure holds important implications for directing adaptive (prosocial) and redirecting maladaptive (bullying) peer network dynamics via intervention and prevention strategies. First, blockmodeling is shown to be an efficient way of describing and analysing empirical interactional network global structures and, second, it is necessary to understand the link between the global network structure and the local network mechanisms in a given context while studying (e.g. modelling) the empirically obtained networks. It was shown that the selected local network mechanisms are important for the formation of the symmetric core-cohesive blockmodel, even without considering any other node attributes. This implies that these local network mechanisms should be considered when empirical networks are being studied.

6 Generating the selected blockmodel types with the proposed NEM algorithm

The focus in the previous chapter (Chapter 5) is on whether the selected local network mechanisms can lead the network to the asymmetric core-cohesive blockmodel. Three different initial global network structures are considered – an empty network, a network with an asymmetric core-periphery blockmodel and a network with a cohesive blockmodel. The effect of the selected mechanisms is discussed in this context.

This chapter addresses the issue of generating networks with other well-known blockmodel types (i.e., cohesive blockmodel, symmetric and asymmetric core-periphery blockmodel, transitivity blockmodel, transitive-cohesive blockmodel, hierarchical blockmodel and hierarchical-cohesive blockmodel) by considering the same set of local network mechanisms as in the case of generating asymmetric core-cohesive blockmodels.

6.1 Methodology

The methodology that is applied is the same as for generating an asymmetric core-cohesive blockmodel (see Section 4.4). The value of parameter q (which expresses the tendency to have a link) is chosen according to the density of an ideal network with a given blockmodel. To compute this density, the number of clusters is set to 2 (core-periphery blockmodel) or 3 (other blockmodels) and all clusters are assumed to be of equal size (the exceptions are symmetric and asymmetric core-periphery blockmodels where the core clusters consist of $1/3$ of all the nodes).

The total number of iterations is set to 32,269 iterations to select the best ten θs according to the mean number of inconsistent blocks. The restriction to only ten θs is applied due to the high computational costs of generating networks and obtaining the RF values.

The intermediate generated networks are analysed at $n_1 = 100$ iterations and $m_i = m_{i-1} * 1.9$ (for $1 < i < 11$) iterations. This approach is used because most changes usually happen at a lower number of iterations (at earlier stages of the network evolution). To estimate the RF values, the number of iterations is increased, $n_i = n_{i-1} + 1.9(n_{10} - n_9)$ (for $10 < i < 15$).

6.2 Networks with a cohesive blockmodel

The number of inconsistent blocks (not shown in this document) approaches 2 in the case of most θs that generate networks with a cohesive structure with the lowest number of inconsistent blocks. This is expected when pre-specified blockmodeling with three clusters is searched for on the network with the blockmodel with two clusters. Visual representations of the generated networks reveal that the global network structures are indeed very close to cohesive not with three but two clusters. This may be confirmed based on the number of inconsistent blocks presented in Table 6.1.

The global network structure converged within 2,478 iterations. The absolute weights of the parameters are not generally comparable, but when looking at the signs some tendencies are revealed. A very obvious one is that the popularity mechanism is negative or close to zero in most cases, while OSP and assortativity are positive in all cases. Transitivity is usually positive but, when the value corresponding to this local network mechanism is low or slightly negative, the values corresponding to other parameters are positive.

Although all the θs shown in Table 6.1 generate networks with a cohesive blockmodel with two clusters, some generated networks (i.e. θs with ID 5 and ID 25) have higher levels of errors as reflected by the RF in both cases – with two and three clusters. In general, the RF values are lower in the case of a cohesive blockmodel with three clusters than with a blockmodel with two clusters. The differences are relatively small which reflects the fact that the third cluster usually consists of only one unit. The difference in the number of inconsistencies (in the case blockmodeling for binary networks is used) between the model with two and the model with three clusters is therefore relatively small and usually depends on the size of the clusters.

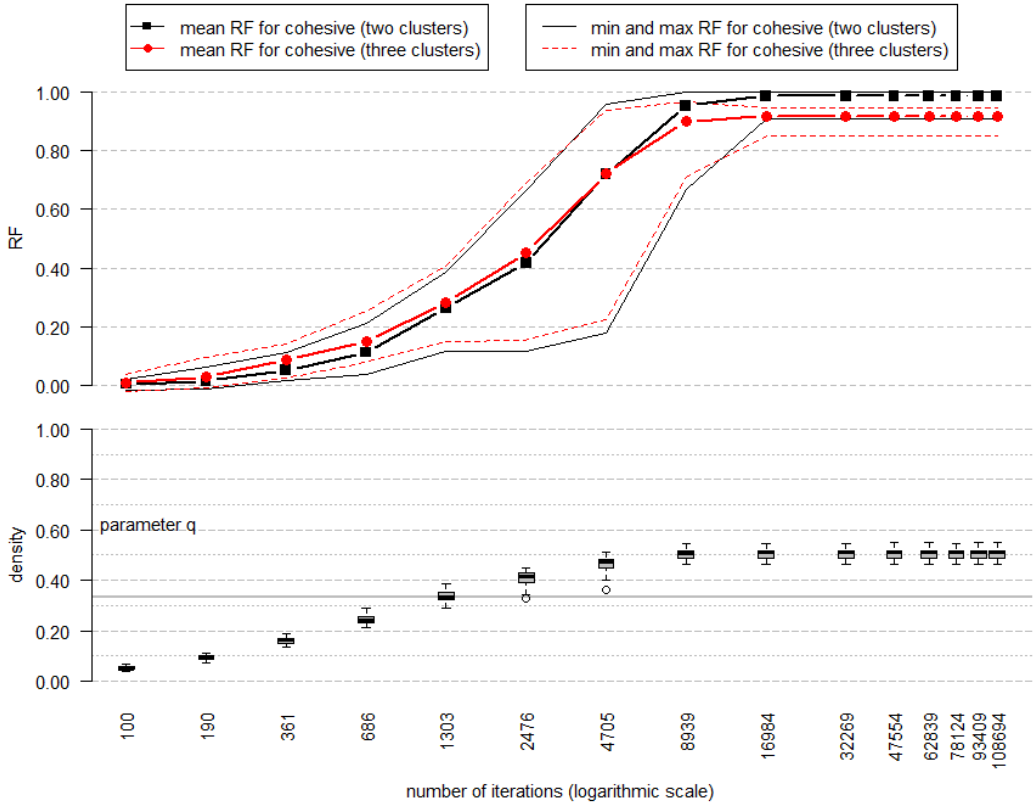
As may be seen in Figure 6.1, the global network structure converges between 8,939 and 16,984 iterations. The density of the generated networks is closer to $1/2$ than to $1/3$. The density of $1/2$ corresponds to an asymmetric core-periphery blockmodel with two clusters of equal size while the density of $1/3$ corresponds to the density of a cohesive blockmodel with three clusters of equal size. Lowering the value of parameter q (which expresses the tendency to create links) does not affect the results.

Table 6.1: The mean number of inconsistent blocks for the selected θs (initial is an empty network and target is a cohesive blockmodel with two or three clusters)

ID of θ	θ					NUMBER OF ITERATIONS (the mean number of inconsistent blocks for cohesive blockmodel with two clusters)										mean RF after 108,694 iterations	
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269	2 clusters	3 clusters
5	-.36	-.37	.86	-.01	.07	1.7	1.2	1.0	1.0	1.0	0.1	0.0	0.0	0.1	0.0	0.49	0.44
8	.11	-.03	.42	.66	.61	1.2	1.0	1.0	1.0	0.6	0.0	0.0	0.0	0.0	0.0	0.75	0.74
15	-.07	-.84	.24	.04	.47	1.9	1.4	1.0	1.0	0.9	0.0	0.0	0.0	0.0	0.0	0.83	0.78
20	.23	-.52	.07	.76	.29	1.1	1.0	1.0	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.89	0.83
25	-.50	-.43	.66	.27	.25	1.9	1.6	1.1	1.0	0.6	0.0	0.0	0.0	0.0	0.0	0.60	0.58
32	.68	.38	.09	-.03	.62	1.1	1.0	1.0	1.0	0.9	0.7	0.6	0.3	0.1	0.0	0.99	0.93
90	.43	-.77	.34	.13	.31	1.2	1.0	1.0	1.0	0.9	0.1	0.0	0.0	0.0	0.0	0.95	0.90
95	.36	.40	.16	-.44	.70	1.2	1.0	1.0	1.0	0.9	0.5	0.2	0.0	0.0	0.0	0.99	0.92
97	.48	.05	.32	.31	.75	1.0	1.0	1.0	1.0	0.6	0.0	0.0	0.0	0.0	0.0	0.96	0.90
98	.40	-.58	.04	.41	.58	1.1	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.96	0.88

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

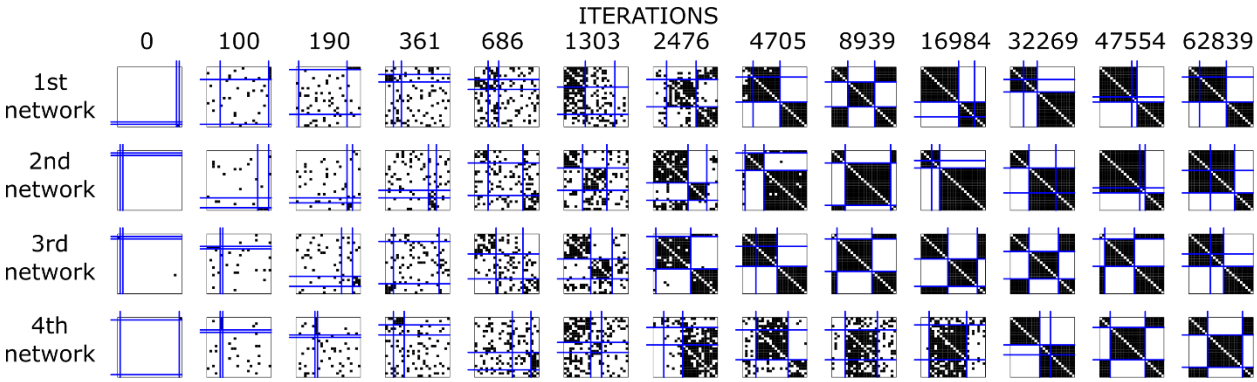
Figure 6.1: Mean value of the relative fit for each blockmodel type and the distribution of the density of the generated networks (by considering the θ with ID 95)



Note: The networks are generated by considering the θ with the highest mean RF value (for the case with two and for the case with three clusters); $\theta = \{M = 0.36, P = 0.40, A = 0.16, T = -0.44, OSP = 0.70\}$, $q = 3/9$, $d_0 = 0$. Initial networks are empty networks.

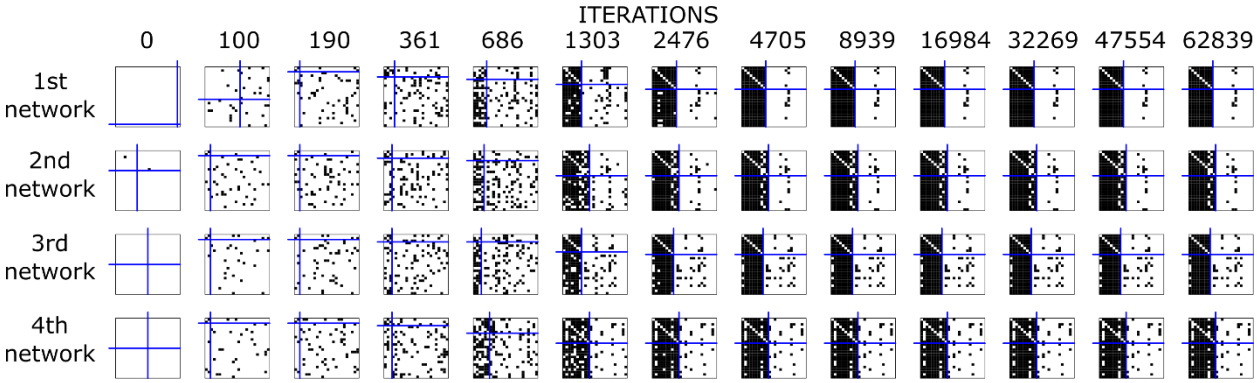
The transitivity type of mechanisms is generally known to be related to the cohesive clusters in the network. It was shown in Section 3 that there is a proportionally higher number of triads of type 300 (and 102) in the cohesive blockmodel. Similarly, Foster et al. (2011) showed that the clustering coefficient, which is defined based on the number of triads and the modularity measure (a higher value of the modularity measure indicates the cohesive clusters are more separated), are positively correlated. However, as Bianconi et al. (2014) pointed out, the correlation does not imply causation. Therefore, they conducted an experiment in which they showed that in growing non-directed networks the closing of triads leads to the emergence of cohesive clusters. In their experiment, the new links were set with probability p if they closed the triad, and with probability $1 - p$ otherwise.

Figure 6.2: Some randomly selected generated networks (by considering only the OSP mechanism) with a cohesive blockmodel with two clusters



Note: The networks are generated by considering only the OSP mechanism: $\theta = \{OSP = 1.00\}, q = 3/9, d_0 = 0$. Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown. The global network structures remain stable at later iterations.

Figure 6.3: Some randomly selected generated networks (by considering only the transitivity mechanism) with an asymmetric core-periphery blockmodel



Note: The networks are generated by considering only the transitivity mechanism; $\theta = \{T = 1.00\}$, $q = 3/9$, $d_0 = 0$. Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown. The global network structures remain stable at later iterations.

In this dissertation, it is shown that networks with a cohesive blockmodel can be generated by considering only the OSP mechanism. Such networks have no errors, although consist of two clusters (Figure 6.2). Another, often discussed transitivity-related mechanism is OTP (a.k.a. transitivity). Generating the networks by considering only transitivity does not result in a cohesive blockmodel as one might assume, but results in an asymmetric core-periphery blockmodel with a very low level of errors (Figure 6.3).

Some other local network effects[28], which are related to transitivity (or to the tendency to form closed triads), are alternating k-triangle and alternating k-two paths (Snijders et al., 2006). The k in the name of the effects refers to the number of shared partners between node i and node j. There is a link between them in the case of the alternating k-triangle while with the alternating k-two paths there is no link between i and j. Both effects are often used in ERGM to analyse non-directed networks.

A positive estimate of the alternating k-triangle parameter in empirical networks is evidence of triangulation in the network and also of the tendency for triangles to occur together in "clumps", of which the core-periphery global network structure might be the outcome (Robins et al., 2009, p. 107). When both effects are included in the model and the alternating k-triangle effect is positive, controlling for the alternating k-two-paths effect, "then there is an evidence that transitivity in this network tends to occur because of the completion of the bases of k-triangles, rather than completion of the sides. In other words, multiple connectivity in the form of independent two-paths tends to lead to network closure" (Robins et al., 2009, p. 109).

The next important effect is an alternating k-star mechanism, which is defined through the number of k-stars and is used to model the degree distribution in the network. A positive estimate (e.g. in ERGM) "suggests a preference for connections between a larger number of low degree nodes and a smaller number of higher degree nodes, akin to a core-periphery structure" (Robins et al., 2009, p. 107).

On the other hand, a combination of the alternating k-triangle effect and the alternating k-stars effect can indicate a "segmented network of multiple (but small) dense regions connected by low-density paths" (Robins et al., 2009, p. 108) when the estimate for the first one is positive and negative for the second one.

The relationship between OSP and OTP and the resulting global network structures is not well known, even though they are defined based on the effects presented above. It is known, however,

[28] The term "effect" is used since this is established terminology in ERGM. These effects are essentially related to the number of different local network structures, but are operationalized as mechanisms in the process of generating the networks and estimating the corresponding model's parameters.

that the path-closure effect (the link from $i \rightarrow j$ is set once the path $i \rightarrow k \rightarrow j$ exists; this is related to the OTP mechanism) is related to the closure of the structural holes (Robins et al., 2009, p. 110). Therefore, the fact that different transitivity-related mechanisms can be related to either the emergence of the core-periphery structure or a cohesive global network structure is extremely relevant.

6.3 Networks with an asymmetric core-periphery blockmodel

Some aspects of the relationship between the transitivity-related mechanisms and the emergence of the asymmetric core-periphery blockmodel were already discussed in Section 6.2. Here, the results of the simulation study regarding the asymmetric core-periphery blockmodel are given.

Table 6.2: The mean number of inconsistent blocks for the selected θs (initial is an empty network and target is an asymmetric core-periphery blockmodel with two clusters)

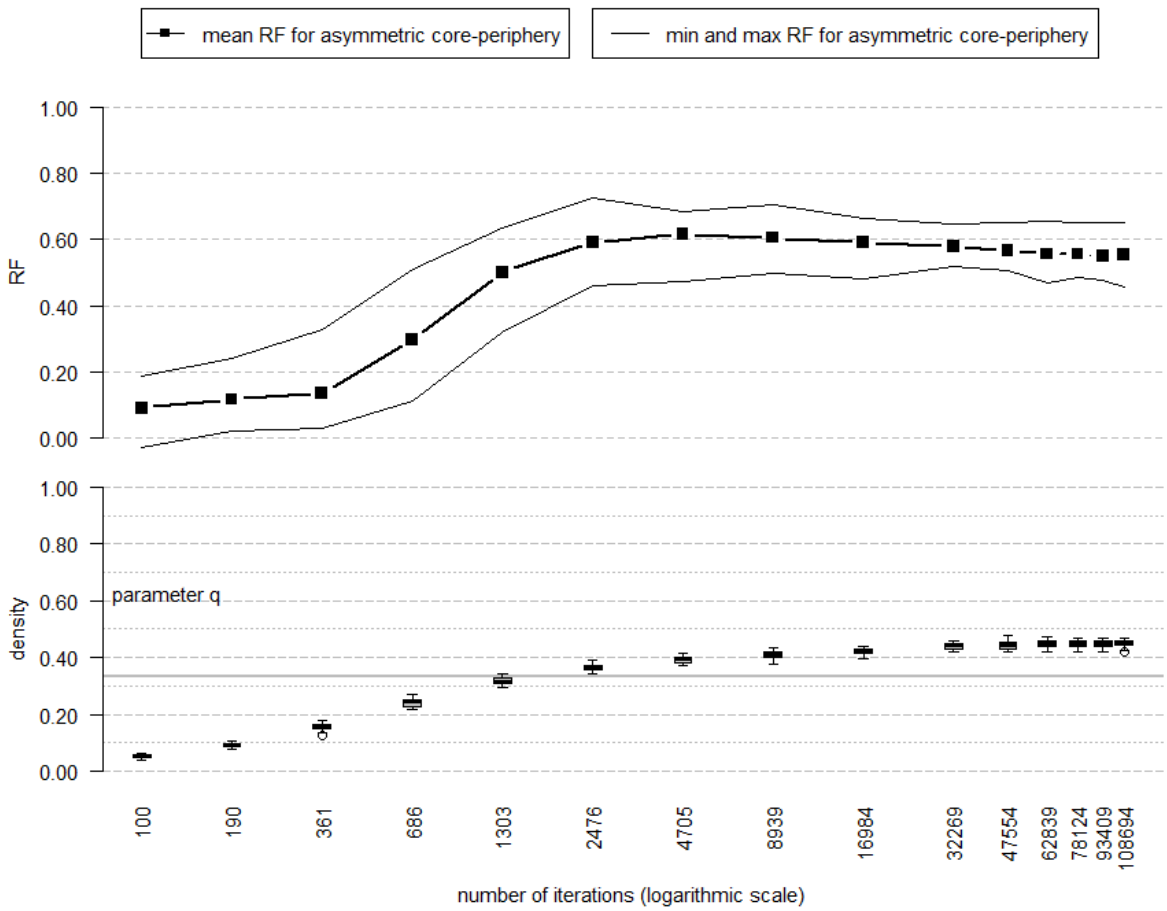
ID of θ	θ					NUMBER OF ITERATIONS										Mean number of IB at 108,694 iterations	mean RF after 108,694 iterations
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269		
3	-.24	.72	-.08	.18	.62	1.67	1.27	0.83	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84
11	-.56	.16	.46	.53	.40	1.83	1.57	1.10	1.00	0.80	0.17	0.00	0.00	0.00	0.00	0.00	0.49
13	-.40	.08	-.17	.66	-.61	1.87	1.57	1.90	0.53	0.10	0.00	0.03	0.03	0.00	0.00	0.00	0.55
17	.25	.03	-.40	.85	-.21	1.17	1.10	1.43	0.87	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.58
19	-.77	-.16	.29	.47	-.26	2.00	1.83	1.27	1.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.51
27	-.66	.15	.29	.62	.27	1.93	1.63	1.07	1.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.48
28	-.06	.23	.59	.28	-.72	1.63	1.17	1.00	1.00	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.56
36	-.77	.06	.34	.53	.14	1.97	1.70	1.07	1.00	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.48
39	-.18	.41	-.38	.36	.72	1.70	1.40	1.47	0.27	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.52
40	-.13	-.06	.49	.85	.10	1.73	1.27	1.00	1.00	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.54

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

Based on the number of inconsistent blocks (Table 6.2), one may conclude that the global network structure converges within 1,303 or 2,705 iterations to the asymmetric core-periphery blockmodel. The target blockmodel appears at the lowest number of iterations when the θ with ID3 (in Table 6.2) is considered. Here, the weight corresponding to the popularity mechanism is the highest compared to other θs. The weights of the popularity mechanism are not so high in the case of other θs, yet the resulting global network structures are still very clearly asymmetric core-periphery. As discussed in Section 4.6, this implies that the mechanisms not seen as directly related to a particular

global network structure (such as popularity for the core-periphery blockmodel or certain transitivity-related mechanisms for the cohesive blockmodel) can produce an unexpected global network structure when combined with other local network mechanisms.

Figure 6.4: Mean value of the relative fit for an asymmetric blockmodel with three clusters and the distribution of the density of the generated networks (by considering the θ with ID 13)



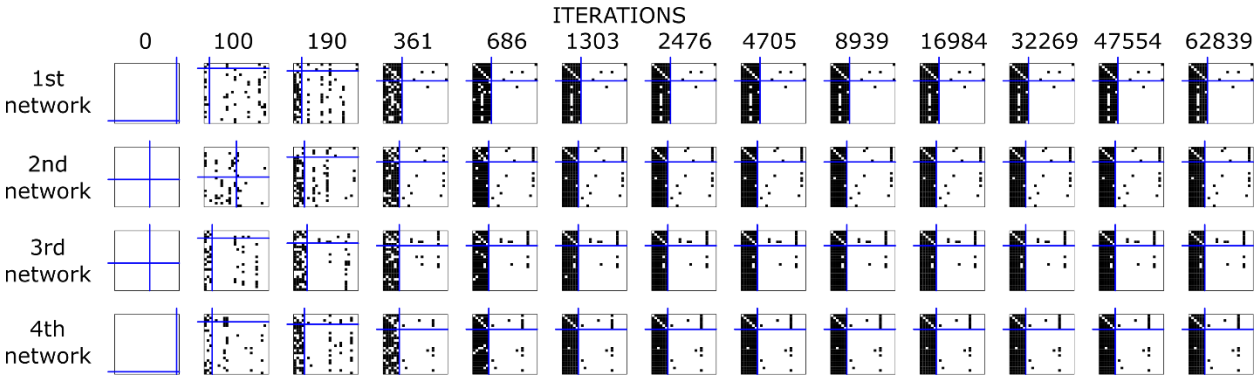
Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = -0.40, P = 0.08, A = -0.17, T = 0.66, OSP = -0.61\}, q = 3/9,\ d_0 = 0$. Initial networks are empty networks.

The mean values of the RF are intermediate to relatively high for the networks generated with the θs shown in Table 6.2. The θ with ID 3 generates networks with an asymmetric core-periphery blockmodel with the highest mean RF. The mean values of the RF converge at around 2,476 iterations (Figure 6.4), but the global network structure can already be visually recognized at 361 iterations (if the networks are drawn in line with the solution of the direct blockmodeling, non-

specified model, see Figure 6.5). The generated blockmodels contain a very small amount of inconsistencies and the weight which corresponds to the popularity mechanism is relatively high compared to the weights of other θs. Solely considering the popularity mechanism would lead a global network structure towards the asymmetric core-periphery blockmodel type.

The lowest values (0.48) correspond to the θs with ID 27 and ID 36. The values are lower due to a higher number of errors in null blocks. More precisely, there is a tendency for a lower level of errors in complete blocks that correspond to the links from peripheral nodes to core nodes compared to the complete blocks which correspond to the links among the core nodes. There is also a larger amount of errors in the block which corresponds to the links between the peripheral nodes (like those generated by considering the θs with ID 19). It turns out that these errors are not random – one or two cohesive clusters tend to emerge in the periphery cluster. This implies that these mechanism weights might lead the global network structure to the asymmetric core-cohesive blockmodel with relatively a large number of errors and with a prevalent core cluster.

Figure 6.5: Some randomly selected generated networks (by considering the θ with ID 3) with an asymmetric core-periphery blockmodel



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = -0.24, P = 0.72, A = -0.08, T = 0.18, OSP = 0.62\}$, $q = 3/9$, $d_0 = 0$. Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown. The global network structures remain stable at later iterations.

6.4 Networks with a symmetric core-periphery blockmodel

The mean number of inconsistent blocks for the best ten θs approach to zero at a relatively low number of iterations (Table 6.4), meaning that the symmetric core-periphery blockmodel can be generated by considering the selected local network mechanisms.

It is worth mentioning that the weights corresponding to the popularity mechanism are negative or close to 0 in most cases. This is expected because very strong positive weights of the popularity mechanism promote links to the most popular ones (i.e. those with a high in-degree) and, therefore, lead the global network structure towards the asymmetric core-periphery blockmodel. However, negative weights of the popularity mechanism promote links being established to those with a lower popularity level (i.e. with a lower in-degree). This can move the global network structure towards a symmetric core-periphery blockmodel.

The weights corresponding to the assortativity mechanism are also negative. This is anticipated since heterogeneity in in-degree is a fundamental characteristic of the core-periphery network structure. A very strong assortativity mechanism might generate links between the peripheral nodes. Moreover, the weights of the mutuality mechanism are positive in most cases, which is expected since all links in the network with a symmetric core-periphery blockmodel are symmetric.

Although the symmetric core-periphery blockmodel appears in the generated networks, the mean RF values are relatively low (Table 6.3). In some cases, the mean RF is close to zero (indicating the analysed networks do not have the selected blockmodel type), while the highest observed value is 0.30. In the latter case, the weights corresponding to mutuality and transitivity are around 0.5 while the weights for assortativity and OSP are around -0.5. The weight of the popularity mechanism is close to 0.

The fact that a relatively high level of errors is found in the generated networks (Figure 6.6) may be due to how θs are generated. Although 300 randomly generated θs are considered, it is possible that 300 is too low and that the θ which would generate the network with a symmetric core-periphery structure still exists.

For example, one could argue that only the popularity and mutuality mechanisms would lead to an asymmetric core-periphery blockmodel (with the logic that an asymmetric core-periphery blockmodel appears because of the popularity mechanism and therefore only asymmetric links must be symmetrized for the transition to a symmetric core-periphery blockmodel).

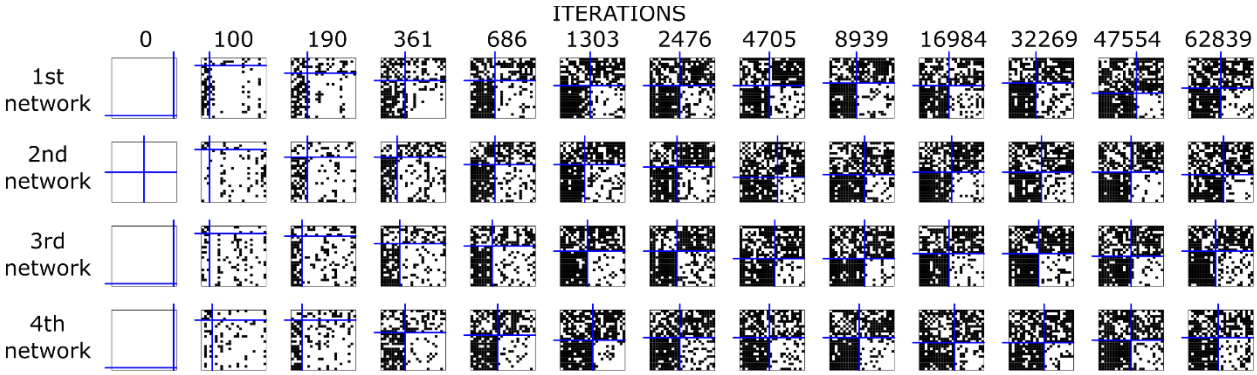
Indeed, the arbitrarily selected $\theta = \{P = 0.25, M = 0.97\}$ generates networks with a symmetric core-periphery blockmodel with almost no errors in null blocks (Figure 6.7). Even though the absolute value for the popularity mechanism is low, the most popular cluster emerges at a very small number of iterations. The nodes from the most popular cluster are not internally linked at a low number of iterations in some cases, which might be due to the high weight of the mutuality mechanism.

Table 6.3: The mean number of inconsistent blocks for the selected θs (initial is an empty network and target is a symmetric core-periphery blockmodel with two clusters)

ID of θ	θ					NUMBER OF ITERATIONS										Mean number of IB at 108,694 iterations	Mean RF at 108,694 iterations
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269		
31	.49	-.08	-.31	-.26	.77	2.0	2.0	2.0	1.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.12
102	.17	-.35	-.48	.64	.46	2.0	2.0	1.9	1.9	2.4	1.4	0.5	0.1	0.1	0.0	0.0	0.11
106	.77	-.56	-.03	.29	-.09	2.0	2.0	2.0	2.7	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.03
110	.43	.16	-.10	.87	-.14	2.0	1.9	1.2	1.0	1.0	0.6	0.1	0.0	0.0	0.0	0.0	0.22
130	.52	-.18	-.66	.47	-.18	2.0	2.0	1.9	1.5	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.12
174	.35	-.40	-.19	-.33	.75	2.0	2.0	2.0	2.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.12
185	.86	.17	-.43	.10	.21	2.0	2.0	1.5	0.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.27
196	.26	.08	-.20	-.02	.94	2.0	2.0	1.8	1.8	1.0	0.0	0.2	0.0	0.0	0.0	0.0	0.17
218	.57	.04	-.54	.44	-.43	2.0	2.0	1.7	1.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.30
229	.08	-.53	-.44	.61	-.38	2.0	2.0	2.0	1.7	0.9	0.2	0.0	0.1	0.0	0.0	0.0	0.06

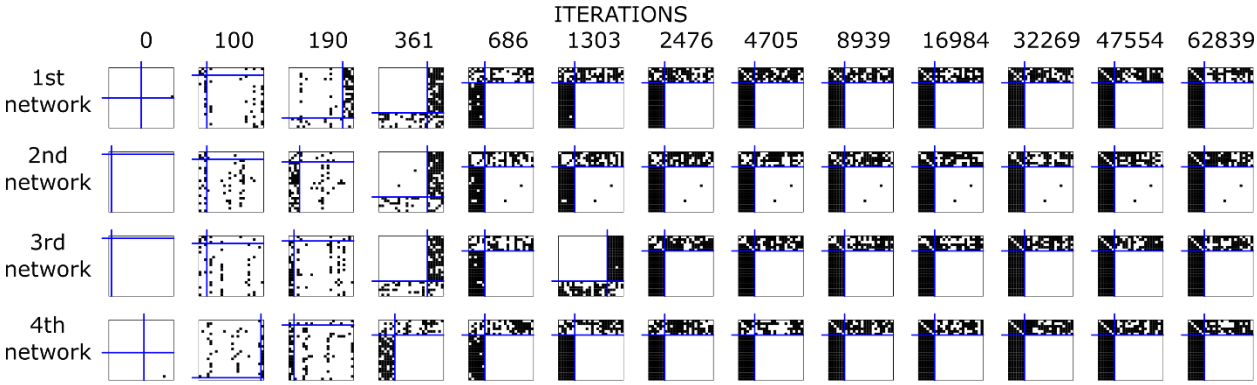
Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

Figure 6.6: Some randomly selected generated networks (by considering the θ with ID 218) with a symmetric core-periphery blockmodel



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = 0.57, P = 0.04, A = -0.54, T = 0.44, OSP = -0.43\}$, $q = 5/9$, $d_0 = 0$. Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown. The global network structures remain stable at later iterations.

Figure 6.7: Some randomly selected generated networks (by considering only the popularity and mutuality mechanisms) with a symmetric core-periphery blockmodel



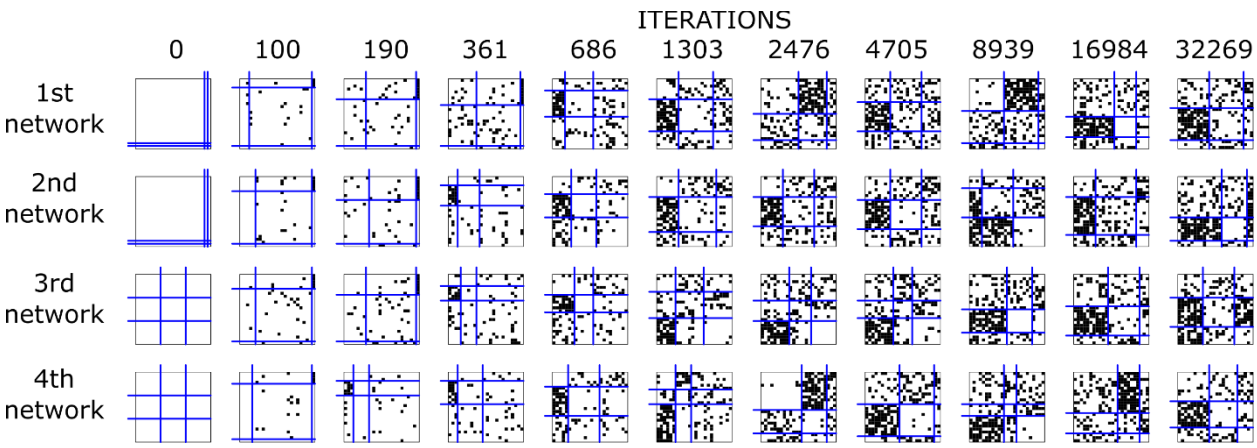
Note: The networks are generated by considering only the popularity and mutuality mechanisms $\theta = \{P = 0.25,\ M = 0.97\}$ and $q = 5/9$, $d_0 = 0$. Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown. The global network structures remain stable at later iterations.

6.5 Networks with a hierarchical blockmodel

Based on the mean number of inconsistent blocks, it may be concluded that none of the best ten θs leads the global network structure to a hierarchical blockmodel (Table 6.4). Evidently, the algorithm for generating networks did not converge at 32,269 iterations in the case of any θ. The

pattern of decreasing and increasing the mean number of inconsistent blocks by iterations indicates the global network structure will not converge to a hierarchical one if the number of iterations is increased. However, based on a visual inspection of the networks so generated it can be seen that many θs lead the global network structure towards the transitivity blockmodel (θs with ID 21, 43, 190, 37, 247, 86 and 91; also see subsection 6.7). The global network structures of the other θs are less clear. They are all similar to those shown in Figure 6.8.

Figure 6.8: Some randomly selected generated networks (by considering the θ with ID 200) with the target hierarchical blockmodel



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = -0.60, P = -0.16, A = -0.73, T = -0.25, OSP = 0.10\}, q = 2/9, \quad d_0 = 0.$ Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown.

Table 6.4: Mean number of inconsistent blocks for the selected θs (initial is an empty network and target is a hierarchical blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS									
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269
21	-.33	-.41	-.78	.35	.02	2.7	2.3	1.6	1.3	1.7	1.7	1.9	1.8	1.6	1.5
43	-.50	.23	-.55	-.04	.63	2.7	2.5	1.9	1.5	2.1	2.3	2.3	1.7	1.8	1.6
190	-.56	-.58	-.18	.42	.38	3.0	2.9	2.3	1.6	1.9	2.2	1.8	1.9	2.0	1.7
208	-.76	-.27	-.39	-.06	.44	2.9	2.7	1.9	1.4	2.1	2.5	1.9	1.5	1.8	1.7
37	-.71	.02	-.69	.10	.05	2.7	2.2	1.4	1.3	2.5	1.9	1.7	1.5	1.7	1.8
247	-.53	-.64	-.22	.43	.26	3.0	2.9	2.8	1.4	1.9	1.6	1.6	2.4	1.8	1.8
290	-.38	-.31	-.60	.27	.57	2.7	2.5	1.8	1.3	1.7	2.4	2.3	2.1	2.3	1.8
86	-.05	.26	-.70	.26	.61	2.9	2.8	1.9	1.6	3.0	2.7	2.2	1.9	1.9	1.8
200	-.60	-.16	-.73	-.25	.10	2.7	2.3	1.6	1.3	2.2	2.2	2.0	2.1	2.0	1.8
91	-.78	-.01	-.58	.06	.24	2.9	2.7	1.5	1.2	1.6	2.1	1.8	2.1	1.7	1.9

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

Table 6.5: Mean number of inconsistent blocks for the selected θs (initial is an empty network and target is a hierarchical-cohesive blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS									
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269
221	.41	-.75	.20	.48	-.10	5.4	3.9	3.4	2.6	3.4	3.5	2.8	2.7	2.0	1.9
153	.84	-.14	.36	32	-.21	5.2	4.6	3.5	3.1	2.7	2.6	2.3	2.4	2.3	2.1
226	.65	-.04	.11	.71	.23	4.5	3.5	3.0	2.3	3.5	3.1	2.4	2.2	2.1	2.1
228	.59	.12	.24	.76	-.08	5.1	4.2	3.2	2.8	3.3	3.0	2.6	2.3	2.2	2.2
136	.41	-.18	.37	.74	-.34	5.1	4.4	3.6	3.2	2.8	2.8	2.6	2.3	2.3	2.2
238	.37	.61	.35	.00	.61	5.3	4.1	3.4	2.6	3.4	3.3	2.5	2.4	2.3	2.2
99	.55	.40	.18	.19	-.69	5.7	5.0	4.7	4.2	3.3	2.9	2.3	2.2	2.2	2.3
224	.45	.10	.77	.44	-.08	5.7	4.3	3.5	3.1	3.3	3.0	2.3	2.2	2.1	2.3
286	.52	.26	.38	.72	.16	4.6	4.1	3.3	3.2	3.2	3.4	2.6	2.4	2.4	2.3
114	.14	-.92	.02	.09	.36	5.1	3.5	3.0	2.2	3.7	4.0	2.9	2.0	2.2	2.3

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

6.6 Networks with a hierarchical-cohesive blockmodel

As with the hierarchical blockmodel, the mean number of inconsistent blocks does not fall below 1.9 inconsistent blocks, indicating a hierarchical-cohesive blockmodel is unable to be generated by using the proposed algorithm and selected mechanisms (Table 6.5). This means that some other or some additional mechanisms should be considered.

6.7 Networks with a transitivity blockmodel

Low values of the mean number of inconsistent blocks at 32,269 iterations indicate that a transitivity blockmodel can be generated by considering the selected local network mechanisms. However, since the mean number of inconsistent blocks is decreasing with iterations, it may be assumed the global network structure did not converge after 32,269 iterations. As mentioned, some values are already low at this number of iterations. Examples are the mean number of inconsistent blocks corresponding to the first and second θs with ID 38 and 43 (Table 6.6). The sizes of the clusters in the networks generated are more homogenous when the θ with ID 4 is used.

Table 6.6: Mean number of inconsistent blocks for the selected θs (initial is an empty network and target is a transitivity blockmodel with three clusters)

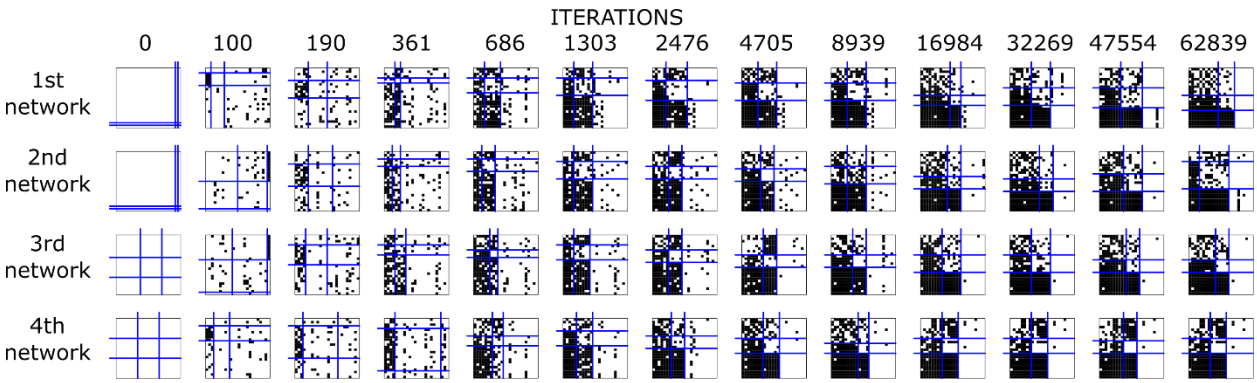
ID of θ	θ					NUMBER OF ITERATIONS										Mean number of IB at 108,694 iteration	mean RF at 108,694 iteration
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269		
38	-.77	.38	-.23	0.0	.45	3.6	3.1	2.2	2.2	1.7	1.2	0.8	0.7	0.7	0.5	0.6	0.48
43	-.50	.23	-.55	-.04	.63	3.7	3.1	2.3	2.4	2.4	1.8	1.2	0.9	0.9	0.5	0.4	0.30
21	-.33	-.41	-.78	.35	.02	3.6	2.6	2.1	2.3	2.1	1.6	1.1	0.6	1.0	0.7	0.4	0.33
244	-.52	-.39	-.25	.31	.65	3.8	3.3	2.3	2.1	2.6	2.9	1.7	1.9	1.9	0.7	0.5	0.23
91	-.78	-.01	-.58	.06	.24	3.4	2.7	2.2	2.3	2.1	2.0	0.9	1.2	1.7	0.7	0.2	0.31
37	-.71	.02	-.69	.10	.05	3.6	2.8	2.1	2.2	1.2	0.7	0.4	0.7	0.8	0.8	0.9	0.47
146	-.60	.30	-.61	-.37	.19	3.7	2.8	2.3	2.3	2.4	2.4	2.0	1.1	1.0	0.8	0.1	0.31
247	-.53	-.64	-.22	.43	.26	3.1	3.6	2.4	3.3	2.6	1.8	1.5	0.9	1.0	0.9	0.6	0.30
13	-.40	.08	-.17	.66	-.61	3.5	3.1	2.4	2.2	2.1	1.6	1.6	1.3	1.3	1.0	1.1	0.61
175	-.33	.07	-.33	.72	.50	3.8	3.2	2.3	2.8	1.8	1.7	1.5	1.4	1.2	1.1	1.3	0.54

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

The mean RF values (Table 6.6) are relatively low, mainly due to the high level of errors in the null and complete blocks. In the case of some θs, the number of inconsistent blocks is relatively

high at 32,269 iterations, but the mean RF after 108,694 iterations is higher compared to some θs, generating networks with a lower mean number of inconsistent blocks at 32,269 iterations.

Figure 6.9: Some randomly selected generated networks (by considering the θ with ID 13) with the target transitivity blockmodel



Note: The networks are generated by considering the θ with the highest mean RF value; $\theta = \{M = -0.40, P = 0.08, A = -0.17, T = 0.66, OSP = -0.61\}, q = 3/9, \quad d_0 = 0.$ Initial networks are empty networks. The networks are drawn in line with the blockmodels obtained (non-specified model). To save space, only networks generated up to the 62,839th iteration are shown. The global network structures remain stable at later iterations.

Table 6.7: Mean number of inconsistent blocks for the selected θs (initial is an empty network and target is a transitive-cohesive blockmodel with three clusters)

ID of θ	θ					NUMBER OF ITERATIONS									
	MUTUALITY	POPULARITY	ASSORTATIVITY	TRANSITIVITY	OSP	100	190	361	686	1.303	2.478	4.705	8.939	16.984	32.269
228	.59	.12	.24	.76	-.08	5.20	4.87	3.97	2.77	2.17	1.90	1.20	1.13	1.30	1.17
286	.52	.26	.33	.72	.16	5.13	4.97	4.10	3.00	2.07	2.00	1.40	1.23	1.30	1.17
40	-.13	-.06	.49	.85	.10	5.27	4.97	4.33	2.87	1.33	1.30	1.27	1.40	1.17	1.20
153	.84	-.14	.36	.32	-.21	5.63	5.17	4.30	2.70	2.10	1.53	1.20	1.23	1.27	1.20
136	.41	-.18	.37	.74	-.34	5.13	4.83	4.17	2.70	2.00	1.90	1.27	1.17	1.30	1.23
48	.44	.15	.26	.79	-.32	5.50	4.70	4.43	3.00	2.03	2.03	1.47	1.67	1.37	1.27
226	.65	-.04	.11	.71	.23	5.03	4.13	3.63	3.33	2.97	2.73	1.83	1.53	1.37	1.27
292	.33	.04	.14	.84	-.42	5.50	4.60	4.03	2.90	1.97	2.00	1.83	1.50	1.30	1.27
224	.45	.10	.77	.44	-.08	5.33	5.27	4.23	2.77	2.23	2.00	1.30	1.23	1.33	1.30
70	.74	.16	.51	.34	-.23	5.20	5.20	4.53	3.03	2.03	1.97	1.30	1.33	1.57	1.37

Note: The results for the best ten θs, according to the mean number of inconsistent blocks, are shown.

The networks with the highest mean RF are generated using the θ with ID 13 (Table 6.6). There is a very small number of errors in complete blocks and a higher number of errors in the first null

block on the diagonal (showing the most popular ones tend to be linked to each other) (Figure 6.9). This global network structure is thus very close to the asymmetric core-periphery blockmodel type. In Section 6.4, the networks generated by considering this θ are evaluated in terms of the asymmetric core-periphery blockmodel. The mean number of inconsistent blocks is very low there.

6.8 Networks with a transitive-cohesive blockmodel

None of the generated θs produced the transitive-cohesive blockmodel. The eight out of ten θs with the lowest number of inconsistent blocks are the same as when generating asymmetric core-cohesive blockmodels. The exceptions are the θ with ID 292 and the θ with ID 70 which were not identified as θs, thereby leading lead the global network structure towards the asymmetric core-cohesive blockmodel but, based on the results provided in Table 6.7 one cannot conclude that they lead the network to a transitive-cohesive blockmodel with three clusters. Instead, the global network structure of such generated networks is reminiscent of the asymmetric core-cohesive blockmodel type except for one cohesive cluster which is mutually linked to the core cluster (the θ with ID 292) or to the asymmetric core-cohesive blockmodel with three clusters of which one cohesive cluster is internally non-linked (the θ with ID 70).

6.9 Conclusion

The aim of this chapter was to test whether one can generate some of the most common blockmodel types by applying the methodology used in Chapter 4 to generate networks with asymmetric core-cohesive blockmodel type. The proposed NEM algorithm was used along with five selected mechanisms: mutuality, popularity, assortativity, transitivity and OTP. The networks were generated by considering different θs. All mechanisms were considered, even though only a few would be theoretically justified in some cases. For example, to generate networks with the asymmetric core-periphery blockmodel, only the popularity mechanism is needed. Yet, the analyses show that combinations of other local network mechanisms can also drive the global network structure towards an asymmetric core-periphery blockmodel.

The five selected mechanisms can also be used to generate the symmetric core-periphery blockmodel, although the number of errors in the networks so generated is relatively high. This is not due to the inability to generate such networks with the selected local network mechanisms, but

because of the relatively low number of θs that were generated (300 different θs). This was shown by considering only the popularity and mutuality mechanisms with arbitrarily selected weights. In this setting, it is possible to generate networks with a symmetric core-periphery blockmodel with a very small number of errors. The parameter's value corresponding to the mutuality mechanism was set to a much higher value than the value for the parameter corresponding to the assortativity mechanism.

A very clear global network structure emerges when generating networks with a cohesive blockmodel, yet the resulting blockmodels consist of two instead of the expected three clusters. A vital observation here is that the transitivity (also known as OTP) mechanism leads the global network structure towards an asymmetric core-periphery while the OSP mechanism leads the global network structure towards a cohesive blockmodel.

Networks with a transitivity blockmodel can be generated using the selected local network mechanisms. These networks have a higher number of errors than networks generated with a cohesive or asymmetric core-periphery blockmodel. In the case of some θs, the generated networks contain errors in one of the null blocks on the diagonal.

By considering the selected mechanisms and their weights, it was impossible to generate networks with a hierarchical blockmodel. The θs that generated the networks with the lowest number of inconsistent blocks in fact produced a global network structure that is closer to a transitive-cohesive blockmodel.

It was also impossible to generate a hierarchical-cohesive blockmodel and a transitive-cohesive blockmodel. The θs that generated the networks with the lowest number of inconsistent blocks actually generated networks whose global network structure was close to an asymmetric core-cohesive blockmodel.

One reason for the inability to generate networks with such a global network structure may be the relatively small number of generated θs (300 in this study), although another reason may be a failure to consider certain other relevant local network mechanisms. Some possible local network mechanisms relevant to the emergence of the hierarchical blockmodel type are discussed in the chapter that follows, where the social context of the knowledge-flow is taken into account.

7 Hierarchical blockmodels in knowledge-flow networks

It is well known that the social network structures in a given company not only influence knowledge creation by determining individuals' opportunity to access and combine knowledge (Nahapiet & Ghoshal, 2000), but also affect their willingness and ability to transfer any more complex knowledge (Reagans & McEvily, 2003). Yet, less is known about which type of global network structure supports optimal knowledge transfer and which local network mechanisms promote the emergence of this global network structure.

Therefore, chapter aims to study which local network mechanisms can lead a network towards the one proposed in the sections appearing below. The type of global network structure is determined according, first, to previous studies that seek to understand how different network properties affect the global network structure and, second, to analysis of the empirical knowledge-flow network. Next, the selection of the local network mechanisms is based on the theory proposed by Nebus (2006) and analyses of the evolution of empirical knowledge-flow networks including selected employee attributes. Like in the previous chapters, a NEM is used to test whether the proposed local network mechanisms can lead the global network structure to the selected one (see subsection 7.1).

It must be emphasized that the formal global network structure can be determined by company policies and the informal global network structure can also be influenced by some personal employee characteristics (such as personality type) which can affect employees' position in the network and thereby the global network structure (Mehra, Kilduff, & Brass, 2001). Formal and informal global network structures are not independent.

7.1 The global network properties and knowledge transfer

The nodes in the network are employees and the links between them operationalize the transfer of knowledge. The latter is defined by »advice seeking« and »learning from«. Empirical studies show that the relationships in advice (Carley & Krackhardt, 1996) and learning (Škerlavaj, Dimovski, & Desouza, 2010) networks tend to be asymmetric or non-reciprocated.

“Advice seeking” and “learning from” are the most often used operationalizations of the flow of knowledge in empirical research, yet measuring different kinds of knowledge (like meta-knowledge, problem reformulation, validation and legitimation) increases the validity of the measurement since different kinds of knowledge can lead to very different global network structures (Cross, Borgatti, & Parker, 2001). The distinction must also be made between advice-giving and advice-receiving because they might differently affect the construct of interest. For example, both advice-giving and advice-receiving are related to job involvement, while only advice-receiving is positively related to commitment to the working group (Zagenczyk & Murrell, 2009).

Many studies have looked at how centrality in the network affects knowledge transfer and performance. In general, a more central individual has a relatively large number of links compared to others in the network, giving them an opportunity to obtain resources from many others. This makes them less dependent on any particular individual (Cook & Emerson, 1978). Sparrowe et al. (2001) showed that those leaders with a higher level of in-degree centrality (i.e. are more popular) estimate their performance higher than leaders with a lower level of in-degree centrality. On the other hand, leaders from groups with higher group centralization (degree centrality was used (Freeman, 1978; Wasserman & Faust, 1994)) estimate the performance of the group as lower. Wong (2008) thought this might relate to the variety of knowledge, which is lower in more centralized groups. She justified the hypothesis by stating:

> Internal advice network centralization, in particular, can foster inequality of member influence when there are increasingly a few group members who are the objects of advice-seeking from other members. /…/ Thus, in a highly centralized internal advice network, there are a few individuals who are most central in providing task knowledge. As such, we can expect their knowledge to become increasingly valued relative to others and they become increasingly influential in decision making (e.g., Bottger (1984), Wittenbaum (1998)). This inequality in influence can lead to increasing deference to the knowledge of more central members (Kirchler & Davis (1986)). When this happens, the opportunity to create new understandings through integrating different viewpoints can be reduced as members become less likely to contradict the perspectives of more central members. In addition, as the knowledge of more central members becomes more valued relative to others, there is the risk of convergence on these ‘valued’ knowledge domains and less emphasis on developing other knowledge domains, thus decreasing knowledge variety over time.

A more limited variety of knowledge is not the only long-term negative outcome of highly centralized groups. Although being central in the advice-giving network can provide a central individual with greater prestige, they can become overloaded by requests for advice from others.

In order to avoid such an overload, the most central individual starts referring the advice-seekers to others in the network. This requires fresh coordination between the most popular ones in order to avoid status competition or conflicts (Lazega, Lemercier, & Mounier, 2006). In addition, maintaining a high number of social ties can lead to lower level of well-being (Rook, 1984), and can become so demanding that work performance is lowered (Burt & Ronchi, 1990; Mehra et al., 2001).

A global network structure can arise as a result of local network processes which may include policies of the company. Studies show that global network structures, when not influenced by company policies, depend on the difficulty of the tasks at hand. Brown and Miller (2000) observed that groups working on more complex problems tend to develop less central communication patterns, while groups working on less complex tasks tend to develop more centralized communication patterns. This may be related to the more efficient knowledge transfer that occurs in less centralized networks.

Sharing knowledge outside the group is especially important when groups are structurally more diverse since members can benefit from different, unique sources of knowledge outside their group (Cummings, 2004). However, group heterogeneity can also bring certain cost. For example, members of different business units can find it difficult to transfer knowledge (Szulanski, 1996). Yet, Cummings (2004) found no difference in group performance between a structurally homogenous and a structurally heterogenous group. A factor more significantly affecting group performance was the extent to which the individuals within a group shared their knowledge.

Establishing links between different groups is associated with structural holes or bridging nodes or groups (Burt, 2009). They are important when it is assumed that different internally highly linked groups of nodes possess different types of knowledge. It is expected that bridging nodes or groups of nodes enable knowledge to be transferred among different groups of nodes. Studies show this depends largely on the knowledge complexity. When knowledge is simple, the presence of a bridge is a necessary and sufficient condition for knowledge transfer, yet more complex knowledge is more likely to transfer (across bridging nodes or groups) when the individuals who bridge either have a strong tie across to both groups or have a diverse network (Reagans & McEvily, 2003). In short, more complex knowledge is more likely to stay embedded in local communities of practice (Reagans & McEvily, 2003). In terms of group productivity – the most productive teams are

internally well connected and have external networks full of structural holes, which connect these teams with external groups (Reagans, Zuckerman, & McEvily, 2004). It is hypothesized that the bridging nodes or groups have a greater absorptive capacity[29] than the others (Reagans & McEvily, 2003).

7.2 Global network structure

Most authors do not define global network structures in a very exact way. Instead, they talk about the presence of hierarchy, cohesiveness or bridging individuals. They also often use terms like "cohesive", "central", "core-periphery" or "hierarchical" to describe a global network structure. These terms are frequently defined and operationalized in very dissimilar ways. Therefore, two very similar global network structures can be described in two very different ways and vice versa.

The terms "central" and "hierarchical" describe two global network properties, which are usually dependent but can be related to very different global network structures from the blockmodeling point of view. For example, the core nodes in the core-periphery blockmodel type are central and hold a higher position in the network hierarchy than the peripheral nodes. On the other hand, the hierarchical blockmodel or the transitivity blockmodel has a high level of hierarchy and also the nodes in the top cluster are the most central. These blockmodels are very different and probably have different impacts on the transfer of knowledge within a company.

It is hence not trivial to choose the most appropriate global network structure, in terms of a blockmodel, for the transfer of knowledge. As shown, this also depends on a company's organizational structure, the task complexity and the absorptive capacity of the employees (Tang, Xi, & Ma, 2006). Lazega (1992) suggested that "smaller knowledge based organizations should have a structure of relationships closer to cohesive groups, while large (mainly manufacturing) systems are supposed to look similar to hierarchical blockmodels". According to the above literature review on this topic, some global network structures can be proposed.

When the tasks are highly complex, one would recommend promoting the establishing of links between those from the same task groups (or business units). Here, the creation of non-formal

[29] Absorptive capacity is related to the capacity to observe new knowledge (Cohen & Levinthal, 2000; Zahra & George, 2002).

relationships is hugely important. In order to further increase the variety of knowledge, management should consider promoting bridging nodes/cores between the groups. In this respect, the whole group or only a single individual can act as the bridge between different groups. However, in order to avoid overload, the number of connections should be limited and thus also the number of bridging nodes/groups. It is suggested that a node or group should bridge only those who are not too different in terms of their knowledge (especially when more complex tasks or knowledge are entailed) since communication between them could be too strenuous (Dougherty, 1992).

Based on the above literature review, it seems that a hierarchical-cohesive or transitive cohesive blockmodels (here, the links among the employees on the same hierarchical levels exist) may be seen as the most efficient for supporting the transfer of knowledge within a given company. In the case of more complex knowledge, the emphasis should be given to links between individuals from the same groups and bridging cores, whereas when the task or knowledge at hand is simpler, the emphasis should be on the links among those from the same groups and also on a hierarchical structure.

7.3 Local network mechanisms

Nebus (2006) proposed the formation theory in which the selection and retention of an advice network formed by a given actor is proposed. Although Nebus considered the case of an ego-network, his well-developed propositions are used in the context of full networks here.

Nebus proposed the theory on the assumption that the ego, who is seeking help (e.g. advice), can have very detailed information about potential contacts (contact-information-rich case) or have no information at all on potential contacts (contact-information-poor case). In the first scenario, the ego can compare the net knowledge value of all potential contacts before choosing one, while in the second scenario the ego does not possess any decision-relevant information. However, as the ego's knowledge increases information about potential contacts (experts) also increases.

In this study, it is assumed that the ego knows all about the global network structure and the tenure of all the other employees. Except for the tenure, the ego is assumed to not have any information

about the nodes' attributes. Instead, the local and global network statistics serve the ego by way of operationalization of those nodes' attributes.

The list of all selected local network mechanisms considered in this study is presented in Table 7.1. The table is separated horizontally into two parts: part one contains the mechanisms related to the perceived value of the alter's advice, while part two contains the mechanisms related to the perceived cost of the alter's advice.

Table 7.1: The considered mechanisms in knowledge-flow networks

		NAME OF THE OPERATIONALIZATION OF THE MECHANISMS	MECHANISMS	OPERATIONALIZATION OF THE MECHANISMS
PERCEIVED VALUE		ALTER-BASED MECHANISMS		
		Hierarchical position of the alter	expertise	how many nodes can reach a given unit
		Tenure of the alter	experiences, skills	tenure (time in the network)
		Popularity level of the alter	willingness to share knowledge, cognitive trust	in-degree
		DYAD-BASED MECHANISMS		
	PERCEIVED COST	Outgoing shared partners	cognitive distance, realizing the alter's knowledge value	number of outgoing shared partners between the ego and the alter
		Difference in hierarchical position between the ego and the alter	social cost, psychological cost, institutional cost, organizational separation	the difference between the number of nodes that can reach the ago and the number of nodes that can reach an alter
		Difference in tenure between the ego and the alter	likelihood of a response, trust	difference in tenure between the ego and the alter
		Distance between the ego and the alter	psychic distance, cognitive distance, geographical distance	geodesic distance between the ego and the alter

7.3.1 Mechanisms related to the perceived value of alter's advice

Four mechanisms (i) hierarchical position of the alter, (ii) tenure of the alter, (iii) popularity level of the alter and (iv) the number of outgoing shared partners between the ego and the alter are assumed to play an important role when an actor is choosing the others he/she will ask for advice. All four mechanisms can be related to the perceived value of the alter (except for the last one, they depend on the alter only and are therefore called "alter-based mechanisms"). For those in a higher hierarchical position it is assumed they possess a higher level of expertise, while those with a higher tenure might have more experience and skills for independent problem-solving. Those with

a higher tenure are more often formal or non-formal mentors to newcomers. Hierarchical position and tenure can be dependent.

The most popular ones are those with the highest in-degree. These are seen (by the ego) as the most active and are thus perceived (by the ego) as the most willing to share their knowledge. Therefore, the cost of obtaining knowledge from such nodes is perceived to be lower. Still, having many requests can pose a burden for the most popular ones. Therefore, the probability the ego will receive actual help of high quality may be lower than perceived.

7.3.2 Mechanisms related to the perceived cost of asking for advice

The mechanisms in this subsection relate to the ego's perception of the cost of advice given by potential alters. The perceived cost not only depend on the alter, like is more the case with the value-related mechanisms, but also on the ego. It can therefore be said that the next four mechanisms describe the relationship between the ego and a potential alter (a group of these mechanisms is also called "dyad-based mechanisms"). These mechanisms are: (i) the difference in hierarchical position between the ego and the alter; (ii) the difference in tenure between the ego and the alter; (iii) the distance between the ego and the alter; and (iv) the outgoing shared partners mechanism.

These mechanisms largely relate to the perceived cost of asking for advice and the probability that the alter will respond to the ego. For example, the probability the selected alter will accept the request for advice depends on the difference in their hierarchical levels. The probability decreases as the absolute difference increases. Responding to those on a much lower hierarchical level can bring a risk of high social cost (e.g. loss of social status) for those in the higher hierarchical position. Also asking for advice from those in a much higher hierarchical position can impose the risk of high psychological cost (e.g. inability to formulate the problem, stress arising from fear of rejection) for those on the lower hierarchical level. For both, the institutional cost (e.g. formal or non-formal feedback can follow after passing formal processes or lines of authority) can be high.

It is reasonable to assume that the distance (e.g. the shortest path) between the ego and alter is negatively associated with the probability the alter will respond to a request for advice. First, the distance between the ego and the alter can be associated with the geographical or psychic (reflecting cultural and institutional differences) distance between them, which increases the cost

and the opportunities of contacts. High network distance can also indicate a higher cognitive distance and, therefore, a lower level of applicability of the alter's knowledge to the ego's current advice requirements.

The number of shared partners between the ego and the alter as defined above is an operationalization of the cognitive distance and the ego's ability to realize the value of the alter's knowledge and, therefore, it is considered as a perceived cost related mechanism and as a perceived value related mechanism.

7.4 Aim

Like in the previous chapters, this one aims to discuss the relationship between the global network structure and the local network mechanisms. The selected social context in this chapter is the flow of knowledge within a company or organization. The possible local network mechanisms and global network structures were already discussed in previous subsections. What follows are analyses of the dynamic and evolution of the empirical global network structures (by considering certain other employee characteristics, such as business unit and tenure). Based on these, the blockmodel to be studied is selected. More specifically, by using the algorithm from the NEM family, we test whether the selected local network mechanisms can drive the global network structure towards the chosen one.

There are some important differences between the NEM proposed in this chapter and the ones proposed in the earlier chapters. In the NEM proposed in this chapter:

- newcomers and outgoers are considered
- links are not dissolved according to the local network mechanisms; instead, the duration of the links (advice-giving or knowledge-flow) is specified
- the number of iterations is not determined arbitrarily but is based on the maximum expected out-degree in the case of a model without any mechanism (null model).

7.5 Empirical case

In this section, real network data collected in a Slovenian medium-sized knowledge-based company are analysed. The aim is to identify a typical global network structure of advice networks, and its evolution.

7.5.1 Company profile and the data collection technique

The data were collected at three points in time (December 2004, July 2006 and April 2007) in a knowledge-based company whose core business is software development, IT and business consulting, maintenance and support (Škerlavaj, 2007).

The company was founded in 1989 with a subsidiary set up in Croatia in 2000 and a joint venture in Serbia in 2003. This was a growing, medium-size company with a total of 93 survey participants in December 2004 and 145 participants in April 2007 (Table 7.2). In this research, the focus is on the employees who were working in Ljubljana since geographical location has a very prominent impact on the global network structure (Pahor, Škerlavaj, & Dimovski, 2008; Škerlavaj, Dimovski, & Desouza, 2010).

Table 7.2: Number of survey participants by geographical location

Time of collecting the data	Total number of employees	Total number of employees who did not participate in the survey	Number of employees participating in the survey by geographical location			
			Slovenia		Croatia	Serbia
			Ljubljana	Maribor	Zagreb	Belgrade
December 2004	93	12	59	0	11	11
July 2006	136	47	60	0	14	15
April 2007	145	0	80	6	26	33

Because no information is available on which employees left the company (outgoers) and which joined the company (newcomers) at the different time points, all outgoers and newcomers are identified based on their participation in the survey. Several employee profiles regarding in which time period they participated in the survey are shown in Table 7.3. Most of the newcomers (44 employees; named "in at 2nd time point" or "in at 3rd time point" in Table 7.3) joined the company in between the second and third points in time. The number of those present in the company at all three time points is also high (30 employees; named "loyal"). The number of those who left after the first time point (named "out at 2nd time point") and the number of those who left after the second time point (named "out at 3rd time point") are nearly the same (13 and 12 employees). Four

employees were not present at only the second time point (named "returning") and only 2 were employed before the second time point (and after the first time point) (named "in at 2nd time point"). Sixteen employees were present in the company at only the second time point. Those who were present at only one point in time might be students (named "out at 2nd time period" and "temporary"). Returning employees (4 employees) might have been present in the company for the whole time, but might not have participated in the survey at the second time point.

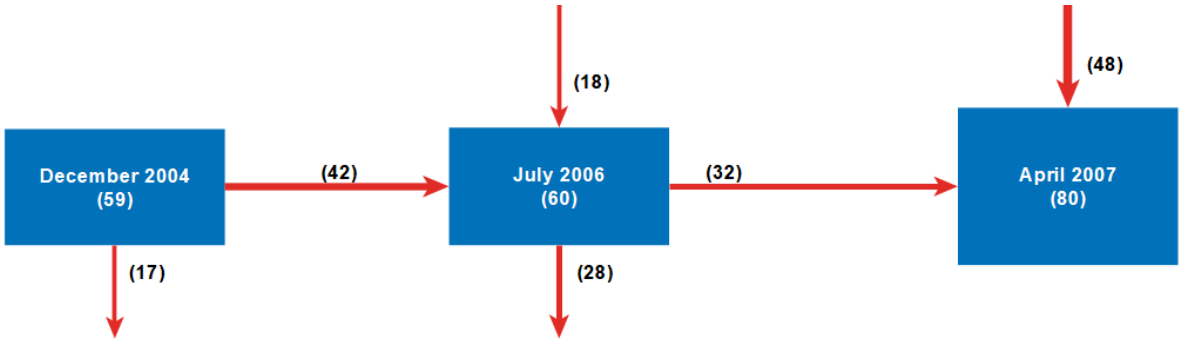
Table 7.3: Different profiles of the employees regarding their survey participation

	time point 1	time point 2	time point 2	frequency
loyal	*participated*	*participated*	*participated*	30
returning	*participated*	not participated	*participated*	4
out at 3rd time point	*participated*	*participated*	not participated	12
out at 2nd time point	*participated*	not participated	not participated	13
in at 2nd time point	not participated	*participated*	*participated*	2
in at 3rd time point	not participated	not participated	*participated*	44
temporary	not participated	*participated*	not participated	16

Note: the total frequencies may differ from those in Table 7.2 because only unique employees are considered in Table 7.3 while in Table 7.2 the same employees may be counted at different time points.

The analyses in which different network statistics were considered (in-degree, out-degree, hierarchic level, transitivity, betweenness centrality, tenure) showed (the results are not included) that the employees with different profiles (regarding their survey participation, Table 7.1) do not differ much in the listed characteristics.

Figure 7.1: Number of employees participating in the survey by years and the number of newcomers and outgoers



The simplified number of employees participating in the survey at a given point in time is visualized by the sizes of the squares in Figure 7.1. The arrows pointing towards the squares are for the newcomers while the arrows pointing away from the squares are for the outgoers. In this figure, some outgoers at the first time point can be seen as newcomers at the third time point (these are "returning" employees).

The company has three business units: (i) Enterprise Resource Planning Solutions (Navision); (ii) Industry Solutions; and (iii) Banking Solutions. The employees in common services and in the directorate are also included in the analysis and considered as business units. The highest number of employees worked in Navision (at all three time points). This is also the business unit in which most new employees were employed.

Table 7.4: Size of the business units in Ljubljana

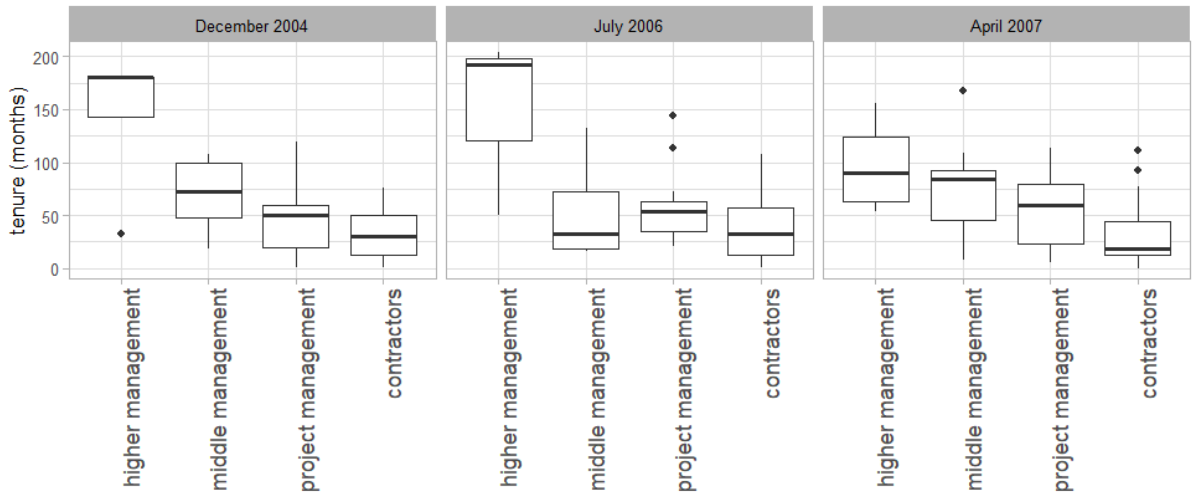
Time of collecting the data		Business Unit					Total
		Common services	Navision	Industry solutions	Banking solutions	Directorate	
December 2004	Frequency	4	22	16	11	2	55
	Share	7	40	29	20	4	100
July 2006	Frequency	4	29	13	12	2	60
	Share	7	48	22	20	3	100
April 2007	Frequency	10	30	22	15	3	80
	Share	13	38	28	19	4	100

Note: the total frequencies might differ from those in Table 7.2 due to some missing values

Approximately 75% of the males were employed at all three time points. The average tenure (the number of months employed at the company) was 48 months with a standard deviation of 42 in December 2004, 53 months with a standard deviation of 42 in July 2006 and 41 months with a standard deviation of 34 in April 2007. Half the employees had worked for 42 months with the company in December 2004 and in July 2006, while in April 2007 half the employees had worked with the company for only 26 months. This is due to the many new employees in between the last two time points of observations.

Most employees were contractors (56% – 64%) or project management (20% – 28%) while the minority was in higher or middle management. In April 2004, there was a very clear relationship between tenure and hierarchical position in the company, while at the latter time points this relationship become less clear (Figure 7.2).

Figure 7.2: Relationship between tenure and hierarchical position



Different name generators were used to measure the flow of knowledge within the company. In this study, only those which were used at all time points are used:

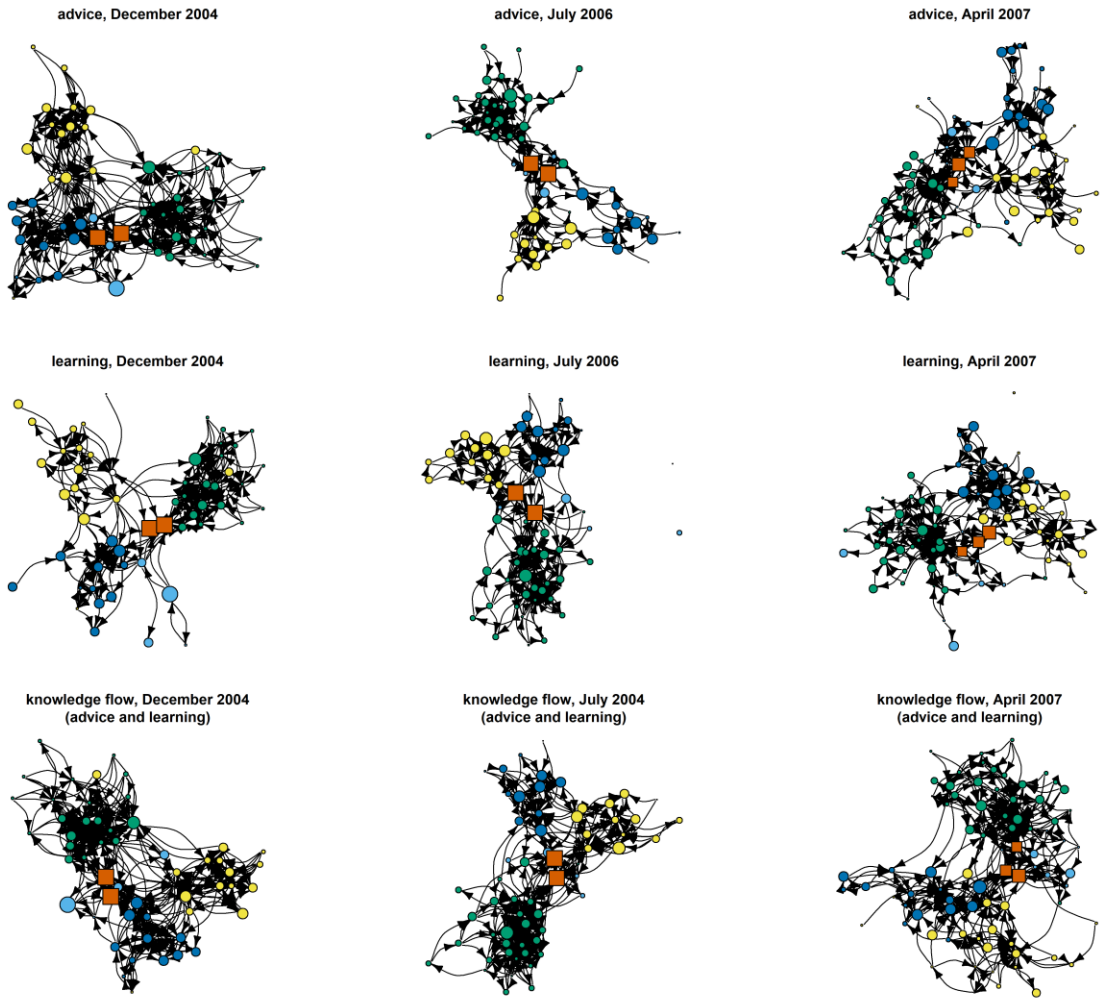
i. "To whom from the company do you ask when you need advice or information related to work?"; and
ii. "Who are the others in the company from which you learn the most?".

The employees were asked to list as many other employees as they wished. Instead of names, they wrote codes assigned to each employee to ensure confidentiality. One or two employees listed all of their co-workers from the same business units. Such answers were considered as valid.

Asymmetric binary complete networks were created based on the information provided (Figure 7.3). These networks are sparse (the density is between 0.03 and 0.10 for all of them). Learning networks are generally sparser than advice networks. This might be a consequence of the perception of the advice and learning. Giving advice can be seen as less formal and less threatening in the sense of losing one's non-formal hierarchical status in the company, compared to learning, which is more status-related. Moreover, the density of the networks is decreasing in both network types, which may be due to the fact that the networks are growing in time. In order to retain the same density as the network grows, the average in-degree or out-degree must be increasing. However, in the empirical networks studied the mean in-degrees and out-degrees are slightly lower

at later points in time. This might reflect the fact that it takes newcomers time to develop their personal learning and advice networks in a company.

Figure 7.3: Binary advice, learning and knowledge-flow networks for different points in time (the nodes are coloured by considering the business unit)



The formation of clusters can be seen in all networks, based on the visualizations in Figure 7.3. The size of the nodes is proportional to the tenure of an employee while the different colours denote different business units. One can see that the clusters are mainly separated by the business units, which is expected since different business units deal with very specific areas of work. Further analysis will reveal if any specific non-cohesive structures appear within clusters from the same business units. Figure 7.3 reveals that two orange coloured nodes with a higher tenure are the bridging nodes. They belong to the directorate.

The global network structures are similar in advice networks and in learning networks because both name generators measure the same dimension, i.e. knowledge flow. Therefore, the advice and learning networks are combined into so-called knowledge-flow networks. A link from node i to node j exists in the knowledge-flow network if it exists in at least one of the networks (in the advice or learning network). The density of the knowledge-flow networks decreases from 0.11 in December 2004 to 0.05 in April 2004.

7.5.2 Prior analyses of this data

The data were already analysed using different approaches such as exploratory data analysis and SAOM. In all cases, the learning networks were used (Pahor et al., 2008; Škerlavaj, Dimovski, & Desouza, 2010; Škerlavaj, Dimovski, Mrvar, & Pahor, 2010). Employees from all geographical locations were included in the analyses.

The results of the exploratory network analysis show the formation of clusters in the network is highly affected by the geographical location much more than by the business units. The most popular employees in terms of transferring knowledge are the most experienced in the field and the most competent.

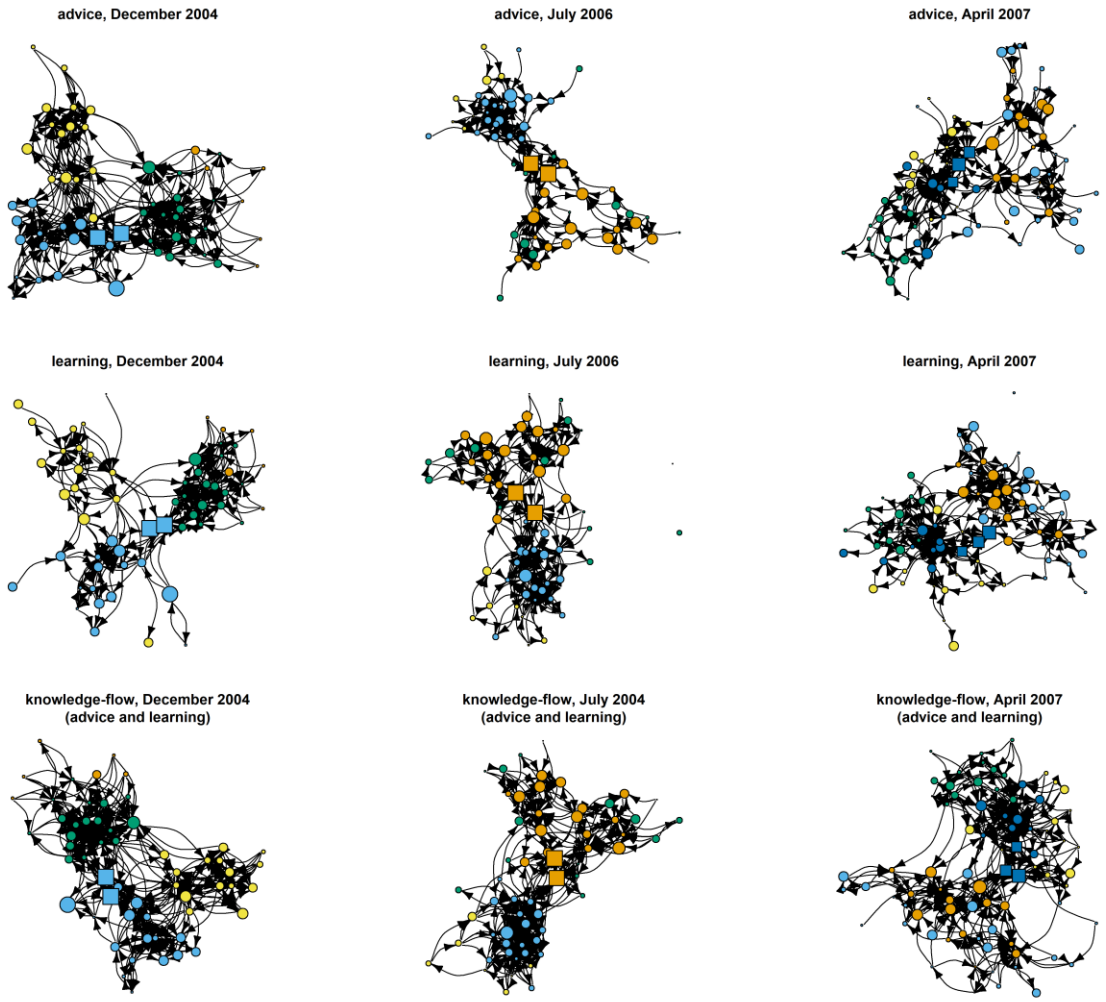
By using Exponential Random Graph Modelling (ERGM), the mentioned researchers confirmed there is a higher probability that knowledge will flow between two employees if they are from the same location and/or business unit. Knowledge will more likely be exchanged between employees of the same gender or with a similar tenure or hierarchical position. Similar applies for working in the same business unit. There is also a greater probability for those higher on the hierarchical level and with higher tenure that they will be recognized as one from whom the others learn. Hierarchical position in the company has a bigger impact on the number of incoming ties than tenure (experience). Knowledge flow is quite an asymmetric relationship.

7.5.3 Methodology

Direct blockmodeling for sparse networks (Žiberna, 2013) is used to obtain a blockmodel. As suggested by Žiberna (2013), the weight $d/(1-d)$ for complete blocks and $1-d$ for null blocks (where d is the density of the whole network) is used.

The number of random restarts in the relocating algorithm is set to 500. The number of clusters is estimated based on dendrograms (Ward's agglomerative clustering method is used on the dissimilarity matrix obtained by the corrected Euclidian distance) and the stability of the blockmodeling solution, which is accessed by a visual examination of the networks and blockmodels.

Figure 7.4 Binary advice, learning and knowledge-flow networks for different points in time (the nodes are coloured by considering the blockmodeling solutions)



7.5.4 Results

The blockmodeling solutions are represented in graphic form in Figure 7.4. The nodes from the same clusters are coloured the same. Compared to Figure 7.3, it can be seen that business units have a relatively high impact on the blockmodeling solution (the clusters are mainly separated by

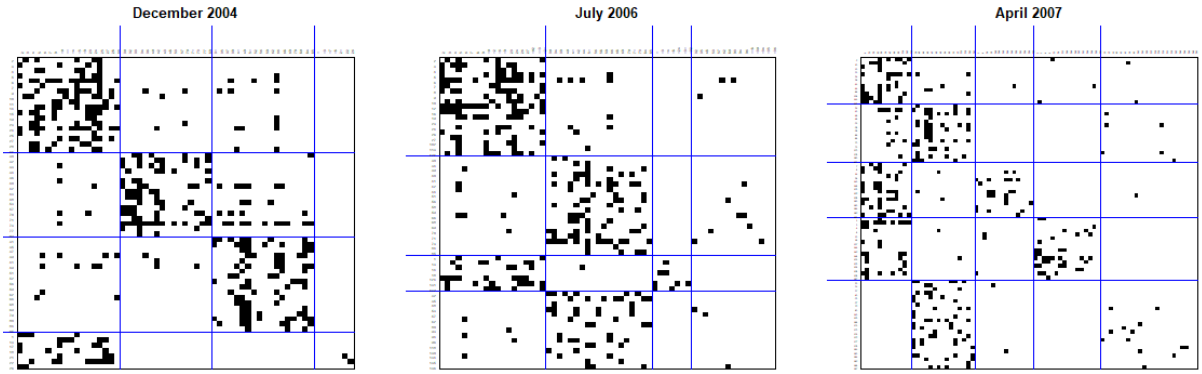
the business units), but some additional sub-structures appear within different clusters of nodes which are mainly from the same business units.

The blockmodeling solutions for the knowledge-flow networks at the three points in time are represented in matrix form in Figure 7.5. The employees are listed by rows and by columns and the order of the rows and the columns is in line with the blockmodeling solution. Each dot represents a link. Horizontal and vertical blue lines denote the clusters that are obtained.

There is a relatively small number of inconsistencies in complete and in null blocks in all three solutions. The global network structure of the first network is close to the cohesive one with three clusters. However, there is a cluster of seven employees who are well linked to the first cohesive cluster, although they are not internally linked to each other. This part of the network expresses the tendency for an asymmetric core-periphery structure in the network. The global network structure of the second network consists of two separate parts where the structure of each part is asymmetric core-periphery (1st and 3rd cluster, 2nd and 4th cluster). The diagonal block, corresponding to the third cluster, is classified as a complete block even though only a few links are present.

The tendency for the presence of a hierarchical structure in the networks is well expressed in the knowledge-flow network, observed at the last time point. Employees from the fifth cluster are linked to those from the second cluster, while those from the second cluster are weakly linked to those in the first cluster. The employees from all clusters, but not the fifth, are internally well linked. The third and fourth clusters are linked to the first cluster.

Figure 7.5: The empirical knowledge-flow networks



Note: The networks are drawn in line with the blockmodel obtained (non-specified blockmodeling.

The similarity or stability of the obtained clusters can be evaluated by using the Rand Index (Rand, 1971) or Modified Rand Index (Cugmas & Ferligoj, 2018). Both indices are defined based on the number of pairs of units that are classified into the same or different clusters in both partitions. The Rand Index requires that both partitions to be compared are obtained from the same set of units. If newcomers and outgoers are present, they must be removed from the data prior to analysis. On the other hand, the Modified Rand Index is defined in such a way that newcomers and outgoers lower the value of the index, along with the merging and splitting of clusters. Non-adjusted measures can take values on the interval between 0 and 1, where a higher value indicates more stable or similar partitions. In general, they are not comparable. Therefore, they are adjusted for chance. In this case, the expected value of the indices equals around zero in the case of two random and independent partitions, and in the case of two identical partitions the values of both indices are equal to one.

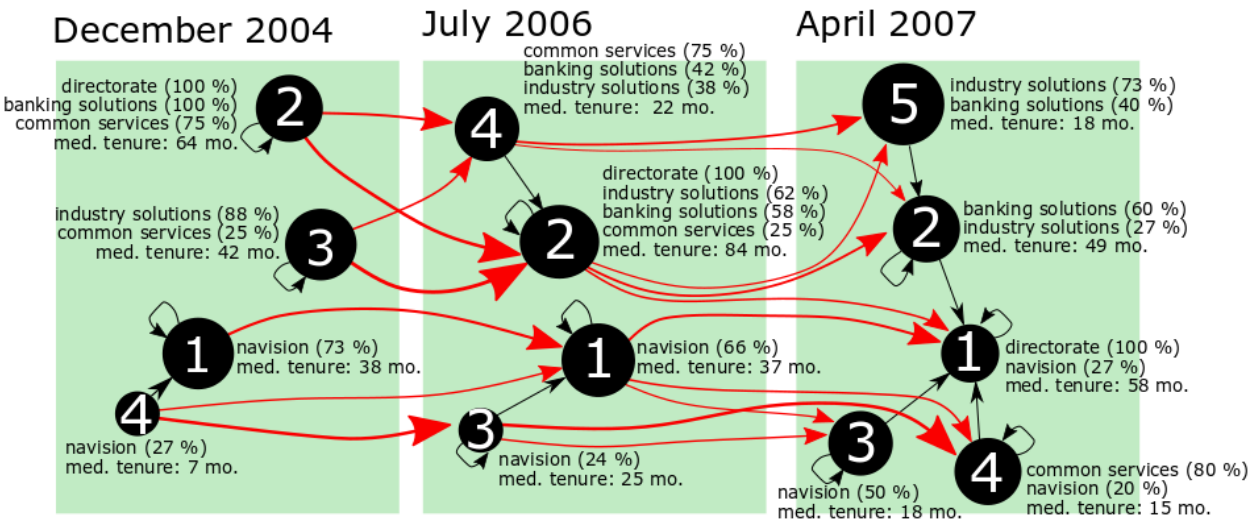
The values of the indices are shown in Table 7.5. The Rand Index indicates the clusters that are obtained are relatively stable when only those present in both time periods are considered. On the other hand, the stability of the obtained clusters is extremely low when newcomers and outgoings are also considered, which is the case with the studied company.

Table 7.5: Stability/similarity of the obtained clusters

	December 2004 vs. July 2006	July 2006 vs. April 2007	December 2004 vs. April 2007
Rand Index	0.43	0.32	0.21
Modified Rand Index	-0.01	-0.08	-0.08

Figure 7.6 visualizes the stability and structure of the clusters that are obtained. The nodes represent clusters. Loops illustrate that the employees within clusters are well linked. The sizes of the nodes are proportional to the number of employees classified in each node (cluster). Black arrows visualize the relationships between the clusters. Red arrows show the selected transitions between clusters at two time points.

Figure 7.6: Stability of the clusters obtained by blockmodeling and their structure according to business units and tenure



The share of employees from each business units and the median number of months working for the company (tenure) are given. These are important differentiators between the various clusters. A more detailed interpretation of the relationship between the business units and the clusters' structures is as follows:

- **December 2004:** all employees from the directorate and banking solutions are classified in cluster 2. Most employees from common services are also classified in this cluster with the highest median tenure. The employees from industry solutions are classified in cluster 3. All employees from Navision are in clusters 1 and 4. The employees in cluster 1 have a higher median tenure than those from cluster 4. They are internally well linked to each other, while those in cluster 4 are not linked to each other.
- **July 2006:** After merging clusters 2 and 3 from December 2004 into cluster 2 in July 2006, most employees from the directorate, industry solutions and banking solutions are in this cluster. Most employees in cluster 4 are newcomers. Most of those from common services are in cluster 4. Clusters 1 and 4 from December 2004 also remain stable in July 2006 (cluster 4 in December 2004 is labelled as cluster 3 in July 2006). However, there are many new employees in cluster 1 in July 2006 (clusters 1 and 3). Probably due to some old well-linked employees, cluster 1 in July 2006 remains internally well linked. On the other hand, a flow of knowledge is established among those who were in cluster 4 in December 2004.

- **April 2007:** All employees from the directorate are in cluster 1 in April 2007. Some employees from Navision are also in cluster 1. The knowledge flow among those from different hierarchical levels (namely the directorate and Navision) might be because these employees are present in the company for a very long time. This is the most central cluster and the employees in this cluster have the highest median tenure. Cluster 2 also remains relatively stable in April 2007. Most employees from clusters 3, 4 and 5 are newcomers with some old employees. All except those in cluster 5 are internally well linked to each other. They also ask for advice or learn from the other employees who are classified in clusters with a similar structure regarding the business units, but have a higher tenure.

Based on the description above, it seems the initial network structure was mostly determined by expertise (i.e. business units). However, those with a lower tenure tend to acquire knowledge from more experienced employees. As time passes, the directorate (along with some "old-timers") become more and more central. The very peripheral clusters (which are internally non-linked at the beginning) mostly consist of newcomers. Many newcomers became outgoers at a later time. However, newcomers tend to ask for advice from those who are more experienced and come from a similar business unit.

7.6 Simulation approach

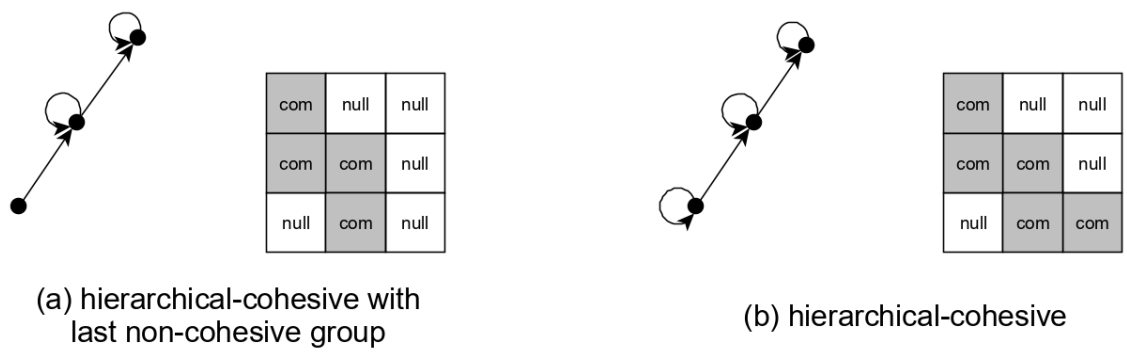
In this section, it is evaluated whether the selected local network mechanisms can drive the global network structure towards ones similar[30] to what is found in the empirical networks. The chosen blockmodels are visualized in Figure 7.7. Both are hierarchical-cohesive, but in the case of the first blockmodel, the last block is a null block. This means there are no links between the nodes on the lowest hierarchical level. Three clusters are assumed in both cases.

The following methodology is applied to study the relationship between the selected local network mechanisms and the chosen blockmodels. First, the local network mechanisms are selected

[30] It is expected that the global network structure found in the empirical data would be even closer to the selected blockmodels if the impact of the business (on the global network structure) were weaker. However, the proposed global network structures are chosen for their simplicity and because the focus of this study is not on the business units.

(see Section 7.3) and operationalized (see Section 7.7). Many vectors of the mechanisms' weights (θs) are then generated in such a way that they are approximately equally distributed in k-dimensional space, where k is the number of dimensions (i.e. the number of local network mechanisms being considered). A very detailed description of how to generate θs is given in Appendix C.

Figure 7.7: A hierarchical-cohesive blockmodel with last non-cohesive group and a hierarchical-cohesive blockmodel



By using the proposed NEM, which considers the selected local network mechanisms and their weights, 30 networks are generated for each θ. The global network structures of these generated networks are evaluated by interpreting the number of inconsistent blocks (see subsection 2.5.1). The θs are also interpreted. For networks containing the selected blockmodel, the RF is calculated.

The NEM algorithm for generating networks is introduced in the next subsection, followed by a definition of the selected local network mechanisms (i.e. network statistics), a more detailed explanations of how global network structures are evaluated and the description of simulation design. The results are presented in Section 7.10.

7.6.1 The algorithm for generating networks

The network is represented in the form of adjacency matrix X with n rows and n columns, both corresponding to the number of nodes in the network. The links have values which are only considered to control the duration of the links. The local network mechanisms and the global network structures are analysed by considering binarized networks. A link in a binarized network exists if the value in a corresponding valued network is higher than 0.

The algorithm allows different initial networks to be to specified, namely either an empty network, random network or a network with a specific global network structure (e.g. a blockmodel). Besides initial network X, parameters λ and κ have to be set. Parameter λ expresses the maximum expected out-degree, while parameter κ relates to the number of iterations. The local network mechanisms must also be provided with the corresponding vector θ which operationalizes the importance (strength) of the selected local network mechanisms.

The algorithm is iterative. At each iteration, one node (ego i) is selected among all the nodes in the network (each node is selected with equal probability)[31]. Considering i and the selected local network mechanisms, the network statistics are calculated and weighted by θ. The weighted network statistics are normalized on the interval between 0 and 1. From among 25% of the nodes with the highest value of the weighted network statistics, one node is randomly selected.

In addition, the tenure is calculated at each iteration and new nodes (newcomers) are added to the network and some existing nodes are removed (outgoers) at selected iterations.

Weighted network statistics

The weighted network statistics are calculated by the function $compute.S(X,\ i,\ M,\ \theta)$ that considers the set of mechanisms M and the weights of the corresponding mechanisms θ. M is the set of operationalized mechanisms defined on the binarized network X and node i.

The computed value of a given mechanism (from the set of mechanisms M) is a vector of length n. Each element of this vector corresponds to one node in the network. When several mechanisms are considered, the vectors can be organized into matrix H with n rows and m columns representing the mechanisms. The matrix that is obtained is weighted as $S = H\theta^T$, resulting in a vector of length n which is returned by the function $compute.S(X,\ i,\ M,\ \theta)$.

[31] The probabilities could vary among the nodes. For example, it could be assumed that those nodes with a lower tenure will have more opportunities to ask for advice. However, whether this is a reasonable assumption depends on the company's policies and organizational culture. To consider the most parsimonious case, it is assumed in this study that all nodes have equal probabilities of asking for advice at any time.

Duration of the links

No specific mechanism that is considered in this study is able to control the duration of a link (i.e. duration of the interaction between the advice-seeker and advice-giver). Instead, it is assumed that all interactions last the same amount of time. One unit of time is defined through the number of iterations at which each node will get (in average) one opportunity to establish a link. The number of iterations depends on the size of the network and the desired maximum expected out-degree (parameter λ).

Let us consider the case without newcomers and outgoers Let us also assume there are n nodes in the network and each node has up to λ opportunities to establish a link. On the assumption the nodes are chosen randomly with equal probability, the number of iterations needed to reach the expected number of opportunities to establish a link is $\alpha = \lambda n$ (this is the length of one unit of time). Over α iterations, each node receives (on average) λ opportunities to establish a link. This implies that some individual nodes can have a higher out-degree, which happens because some receive more opportunities to establish a link than others. In a very unlikely case, each node could receive exactly λ opportunities to create a link. In that case, the maximum out-degree of each node is exactly $min(n,\ \lambda)$ if loops are allowed and if no tie is chosen twice (in a sense, confirmed or reset).

Links last a limited amount of time. Specifically, the duration of links is set to $\alpha + 1$ iterations. When new nodes are added to the network, parameter α must be updated by considering the new number of nodes. This implies the number of iterations between different waves can vary. The algorithm is implemented in such a way that the number of outgoers does not affect the number of iterations.

In order to ensure sufficient iterations so that the mechanisms being considered can affect the global network structure considerably, the number of iterations is multiplied by the constant κ. The value of $\kappa > 1$ increases the expected number of opportunities for each node to establish a link while it affects neither a link's duration nor the (maximal) expected number of links. A higher expected number of opportunities for each node to establish a tie can also makes the structure more stable before the new nodes are added. In other words, a higher number of iterations gives "more time" to the mechanisms to affect the global network structure.

Newcomers

New nodes can be added over one or several waves. The iterations in which new nodes are added to the network can be selected in different ways: (i) a single node can be added at a time; or (ii) several nodes can be added at once. Further, the node (or several nodes) can be added at randomly selected iterations or at predefined iterations, e.g. equally distributed across iterations. In this study, newcomers are added in three waves. The number of newcomers for each wave is represented by vector $\aleph$. The number of iterations between each wave is determined according to the total number of nodes in the network, based on parameter λ and parameter κ.

Algorithm 7.1: The algorithm for generating networks

```
import initial network X (a matrix with n rows and n columns, where n is the number
of units)
import θ (a vector with the mechanisms' weights)
import M (a set of functions which defines the mechanisms)
set λ (the expected maximum out-degree)
set κ (the factor by which the number of iterations must be increased between the
waves)
set ℵ (a vector with the number of newcomers per waves)
set O (a vector with iterations at which the outgoings are to be removed)
set T (tenure, a vector of length n)
compute forTenureCorVec = cumsum(ℵ_A) * λ * κ (the number of iterations between consecutive
waves of newcomers)
compute E = cumsum(forTenureCorVec) (iterations at which the newcomers are added to the
network)
compute k = max(E) (the total number of iterations)
set forTenureCor to first element of forTenureCorVec
for l in 1:k
|__| set T = T + 1/forTenureCor
|__| randomly select a unit i (actor/ego)
|__| calculate S = compute.S(X_l, i, M, θ) (a vector of the weighted network statistics with
the length n)
|__| calculate φ = (S − min(S)) / (max(S) − min(S)) (normalize S, so the min(S) = 0 and max(S) = 1)
|__| if φ ≥ Q_3(φ) then classify a corresponding unit into set C (where Q_3 is the third
quartile)
|__| randomly select unit j among the units from set C
|__| set a link i → j
|__| calculate X = X − 1/(λn + 1)
|__| calculate X = { 0, x ≤ 0 ; x, x > 0 }
| if l ∈ O
|________|randomly select a unit or a group of units to be removed
|________|remove the selected unit(s) and update X and T accordingly
|if l ∈ E and l ≠ k:
|________|add a unit or a group of units and update X and T accordingly
|________|set forTenureCor to next element of forTenureCorVec
return network X
```

Outgoers

The number of outgoers can be selected arbitrarily. They can leave the network in waves just before or after newcomers are added or can leave the network one by one. In this implementation of the algorithm, outgoers leave the network at the selected iterations which are in vector O. The nodes to be removed from the network can be selected based on their personal characteristics (e.g. tenure), network characteristics (e.g. popularity or hierarchical level) or randomly. Here, the nodes to be removed are selected randomly, which is in line with observations from the empirical data. The number of nodes to be removed from the network is 25% of all the nodes in the network, calculated immediately after a wave of newcomers has been added to the network.

7.7 Operationalization of the local network mechanisms

As proposed in Section 4.2, "the term mechanism describes a process that drives concrete actions according to nodes in the networks (e.g. creating a link to a highly popular unit)". These mechanisms are typically operationalized by different statistics that reflect the mechanisms. These statistics are used in the proposed NEM as described in the previous subsections. The network statistics are defined as follows:

- **Tenure of the alter** and **difference in tenure between the ego and the alter:** The relative value of tenure t (which is a vector of length n as the nodes' attribute) is calculated for the i-th node as

$$\boldsymbol{RT(i)} = \frac{\boldsymbol{t_i}}{\frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{l=1}}^{\boldsymbol{n}} \boldsymbol{t_l}} \tag{7.1}$$

  while the difference in tenure between node i and node j is calculated as

$$\boldsymbol{DT(i,j)} = \frac{\boldsymbol{t_j - t_i}}{\frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{l=1}}^{\boldsymbol{n}} (\boldsymbol{t_l - t_i})^{\boldsymbol{2}}} \tag{7.2}$$

- **Hierarchical position of the alter** and **difference in hierarchical position between the ego and the alter:** First, for each unit, prestige h (which is a vector of length n) is calculated as an indicator of a hierarchical level. Prestige is defined as the proportion of other nodes that can reach the selected ego i in two steps by following the directed links. Then, the relative hierarchical position is calculated for the i-th node as

$$\boldsymbol{RT(i)} = \frac{\boldsymbol{h_i}}{\frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{l=1}}^{\boldsymbol{n}} \boldsymbol{h_l}} \tag{7.3}$$

while the difference in the hierarchical position between node i and node j is calculated as

$$\boldsymbol{DH(i,j)} = \frac{\boldsymbol{h_j - h_i}}{\frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{l=1}}^{\boldsymbol{n}} (\boldsymbol{h_l - h_i})^{\boldsymbol{2}}} \tag{7.4}$$

- **Popularity level of the alter:** The alter popularity mechanism (below referred to as the "popularity mechanism") reflects the tendency to create links to the most popular ones. The popularity statistic (P) is calculated for the i-th node as the ratio between the in-degree of the i-th node and the total number of links in the network:

$$\boldsymbol{P(i)} = \frac{\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{x_{ij}}}{\sum_{\boldsymbol{l=1}}^{\boldsymbol{n}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{x_{ij}}} \tag{7.5}$$

- **Outgoing shared partners (OSP):** This mechanism is defined through the number of nodes k which are shared partners of the ordered pair (i, j) if $i \rightarrow k$ and $j \rightarrow k$. To compute the statistics associated with this mechanism on the selected pair of nodes i and j, one must identify the other nodes (not i and not j) that are linked with i and j (shared partners) in a given way:

$$\boldsymbol{OSP(i,j)} = \sum_{\boldsymbol{k \neq i,j}} \boldsymbol{x_{jk}} * \boldsymbol{x_{ik}} \tag{7.6}$$

$OSP(i, j)$ gives the number of partners shared by i and j. By fixing node i, one can obtain vector V with n elements where each value stands for the number of friends common to node i and all the other nodes. The i-th value of vector V can be normalized as $V_i / \sum_{l=1}^{n} V_l$. Such normalized statistics are used to operationalize the OSP mechanism.

- **Distance between the ego and the alter:** The distance between the ego and the alter is defined by $\mathrm{G}(i, j)$, which is the minimum number of links to reach node j from node i, following the directed links. If node j cannot be reached, then $\mathrm{G}(i, j)$ returns the maximum distance between node i to all other (connected to i) nodes in a network increased by 1. The distance is normalized

$$DD(i,j) = \frac{G(i,j)}{\frac{1}{n}\sum_{l=1}^{n} G(i,l)} \tag{7.7}$$

7.8 Evaluation of the global network structures

The global network structures are evaluated after being generated by the algorithm proposed in the previous section. According to observations from the empirical data and the proposed theory, not only is the emergence of the chosen blockmodel types required, but the average tenure must also be in line with hierarchical levels of the clusters. Therefore, a two-step evaluation procedure is used.

In step one, the number of inconsistent blocks is calculated to evaluate the fit of a global network structure to the chosen one. In the second step, the average tenure is obtained for each cluster in the networks that does not have any inconsistent block. The network is deemed to have the chosen global structure when both conditions are satisfied.

The mean RF value is calculated for those networks with the chosen global network structure to quantify the amount of inconsistencies in the generated networks.

7.9 Simulation design

To generate networks by considering different local network mechanisms, 2,000 θs are generated, with 30 networks being generated for each θ. Initial networks are random networks generated by the model $G(n = 20, p = 0.25)$, $\lambda = 5$ and $\kappa = 4$. λ is estimated based on the empirical networks (see Section 7.5), while parameter κ is determined as a compromise between the convergence of the global network structure before the new nodes are added to the network and the (computational) time needed to generate the networks.

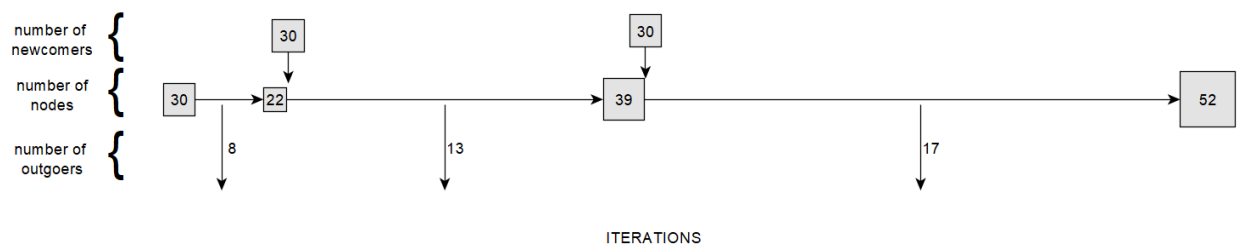
The analyses unfold over several parts, depending on: (i) the local network mechanisms that are included (only tenure-related mechanisms, all but tenure-related mechanisms, and all mechanisms); (ii) the presence of newcomers and outgoers; and (iii) whether the constraints on the signs of the mechanisms' weights are considered (Table 7.2).

Two sets of new nodes are added to the network, $\aleph = \{30, 30\}$ and the following numbers of outgoers (when considered) are removed from the network between different waves of newcomers: $O = \{9,12,18\}$ (see Figure 7.8).

Table 7.6: Different settings for generating networks

Mechanisms	Newcomers / outgoers	Constraints on the signs of the mechanisms' weights
only tenure-related mechanisms	newcomers only	Yes
	newcomers and outgoers	Yes
all but tenure-related mechanisms	newcomers only	Yes
	newcomers and outgoers	Yes
all mechanisms	newcomers only	Yes
	newcomers and outgoers	Yes
all mechanisms	newcomers and outgoers	No

Figure 7.8: The size of networks at a different number of iterations, the numbers of newcomers and outgoers (for the case when both newcomers and outgoers are present)



7.10 Results

This section includes several subsections. In the first subsection, the results concerning the generating of networks with a hierarchical-cohesive blockmodel with last non-cohesive group are presented. In the next subsection, the results are given with respect to generating networks with a hierarchical-cohesive blockmodel. In both cases, the newcomers and outgoers are considered, which is more realistic than a case with only newcomers. The θs are generated in such a way that the signs of the mechanisms- weights are in line with the operationalization shown in Table 7.1. In both subsections, different sets of mechanisms are considered (the set of only tenure-related mechanisms, the set of all but tenure-related mechanisms, and the set of all mechanisms).

A discussion on generating networks by considering only newcomers and another on generating networks by considering the mechanisms' weights without constraints on their signs are found in the last subsection.

7.10.1 Hierarchical-cohesive blockmodel with last non-cohesive group

Here, it is assumed that the nodes on the lowest hierarchical level are not linked to each other as is the case of the 5th cluster at the 3rd time point on the empirical networks (see Figure 7.6). It is also the case in the empirical networks that the employees who form the cluster that correspond to the last (null) block on the diagonal have (on average) the lowest tenure compared to employees from other clusters. This observation is also considered while evaluating the generated global network structures.

Since tenure is closely related to the global network structure, the simulation study-s results are shown separately for the cases when only tenure-related mechanisms are considered, when all but not tenure-related mechanisms are considered, and when all the local network mechanisms are considered.

Only tenure-related mechanisms

By considering only tenure-related mechanisms, none of the θs generated at least 20 networks with the chosen blockmodel. This does not necessarily mean the chosen global network structures cannot emerge when solely tenure-related mechanisms are considered. Instead, it can happen that such (a non-found) θ still exists and would generate networks with the chosen global network structure.

All but tenure-related mechanisms

When all but tenure-related mechanisms are considered, one of the θs generated all the networks with the chosen blockmodel (see Table 7.7). The mean RF is 0.54, which is relatively high considering the constraints on out-degree. The simulation study estimates (see Appendix E) that the highest RF value (for final networks) is approx. 0.59 when the out-degree of each node is 5 (the mean out-degree in the generated networks is 4.84).

Even though the values of the mechanisms' weights are not generally comparable, it is reasonable to interpret that the weights, which correspond to the mechanisms popularity level of the alter

(0.008) and outgoing shared partners mechanism (0.011), are much lower than the weight corresponding to the three other mechanisms (hierarchical position of the alter (0.35), difference in hierarchical position between the ego and the alter (-0.53) and distance between the ego and the alter (-0.78)). This means that the positive effect of the hierarchical position of the alter and the negative effect of the difference in hierarchical position between the ego and the alter and distance between the ego and the alter can lead the global network structure towards a hierarchical-cohesive one with last non-cohesive group.

Table 7.7: The selected θ for generating networks with a hierarchical-cohesive blockmodel with last non-cohesive group

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter		
374	0.351	Fixed to 0.	0.008	0.011	-0.526	Fixed to 0.	-0.775	0.54	0.035

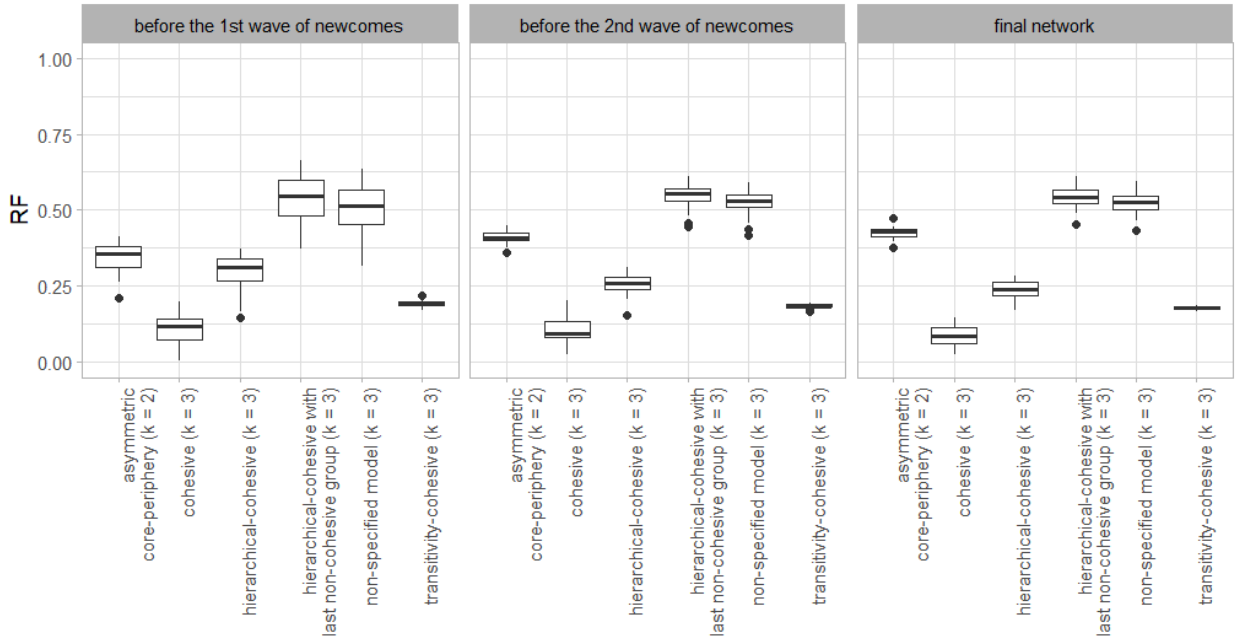
Note: Both newcomers and outgoers are possible. The signs on the mechanisms' weights are fixed. All but tenure-related local network mechanisms are considered. The mean number of inconsistent blocks is 0 in all cases.

Comparing the RF values which are calculated for different assumed blockmodel types at different stages of the network evolution (i.e. at different iterations) can reveal deeper insights into the generated networks' structures (Figure 7.9). In this study, the RF values are calculated after each wave of newcomers and at the end of the iterations. Different blockmodels are assumed: cohesive with three clusters, asymmetric core-periphery with two clusters, hierarchical-cohesive with three clusters, and transitive-cohesive with three clusters. The cohesive and asymmetric core-periphery blockmodel types are selected because they were found to be present at earlier points in time in the empirical networks while a transitive-cohesive blockmodel is selected since it only differs in one or two blocks from the chosen hierarchical blockmodels.

One can see that the RF values corresponding to all of the considered blockmodel types are more variable in the case of the networks observed before the 1st wave of newcomers joined the network compared to networks observed at later iterations. This indicates the global network structure is less clear at the start. Further, the RF values corresponding to the chosen blockmodel type are the

highest compared to other blockmodel types. The values for the non-specified model are slightly lower than those for the chosen blockmodel type, which is expected (see Appendix E).

Figure 7.9: Mean relative fit values for different blockmodel types obtained at different time points (for the θ with ID 374)



All mechanisms

The 9 generated θs (out of 2,000) produced networks with the chosen blockmodel type (all 30 generated networks contained the chosen blockmodel type) (see Table 7.8). The mean RF is relatively high (between 0.42 and 0.52) for most θs but not for the θ with ID 1910 where the mean RF is 0.38. The standard deviations of all RF values are low. The lowest is the one corresponding to the θ with the highest mean RF.

It can be seen in Figure 7.10 (for networks generated by using the θ with ID 1861) that the chosen blockmodel type already appears before the first wave of newcomers is added to the network, although the structure is less clear. Especially high are the RF values corresponding to the hierarchical-cohesive blockmodel (as seen in Figure 7.11, first generated network). Later, the RF values that correspond to the hierarchical-cohesive blockmodel with last non-cohesive group and the non-specified blockmodel increases. The interpretation of the mean RF values for the other θs

in Table 7.8 are similar (the mean RF values are visualized in Figure 1Figure in Appendix D). The findings are consistent with the visual representation of the generated networks in Figure 7.11.

Table 7.8: The selected θs for generating networks with a hierarchical-cohesive blockmodel with last non-cohesive group

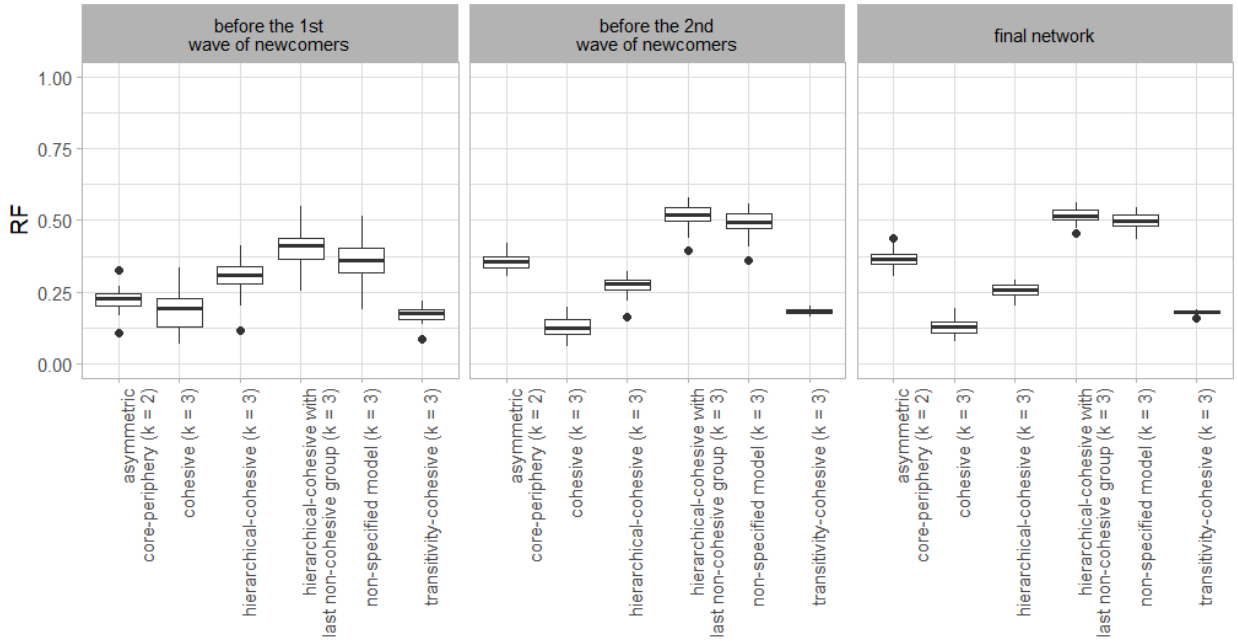
ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter		
1861	0.281	0.380	0.123	-0.016	-0.737	-0.005	-0.468	0.52	0.026
798	0.308	0.703	0.104	-0.023	-0.621	-0.003	-0.121	0.50	0.032
483	0.001	0.794	0.049	0.085	-0.413	-0.153	-0.406	0.50	0.033
1222	0.396	0.557	0.075	0.095	-0.636	-0.015	-0.337	0.50	0.035
1814	0.041	0.732	0.082	-0.011	-0.612	-0.060	-0.280	0.45	0.039
446	0.137	0.546	0.178	-0.118	-0.675	-0.027	-0.425	0.43	0.035
147	0.428	0.377	0.002	-0.121	-0.756	-0.026	-0.297	0.43	0.038
1301	0.039	0.590	0.004	-0.096	-0.483	-0.103	-0.630	0.42	0.035
1910	0.376	0.182	0.047	0.032	-0.827	-0.002	-0.371	0.38	0.031

Note: Both newcomers and outgoers are possible. The signs of the mechanisms' weights are fixed. All local network mechanisms are considered. The mean number of inconsistent blocks is 0 in all cases.

Although the mechanisms' weights are not comparable, some general conclusions can be drawn by considering the most extreme values. It can be seen that the weights for the mechanisms popularity level of the alter, outgoing shared partners mechanism and difference in tenure between the ego and the alter are all generally low.

The weights corresponding to the mechanism tenure of the alter are generally high, although the weights corresponding to the mechanism hierarchical position of the alter are higher in some cases and lower in others. When the weights of the mechanism hierarchical position are high, the weights of the mechanism difference in hierarchical position between the ego and the alter are generally higher as an absolute value while the weight of the mechanism distance between the ego and the alter and the mechanism tenure are generally lower as an absolute value.

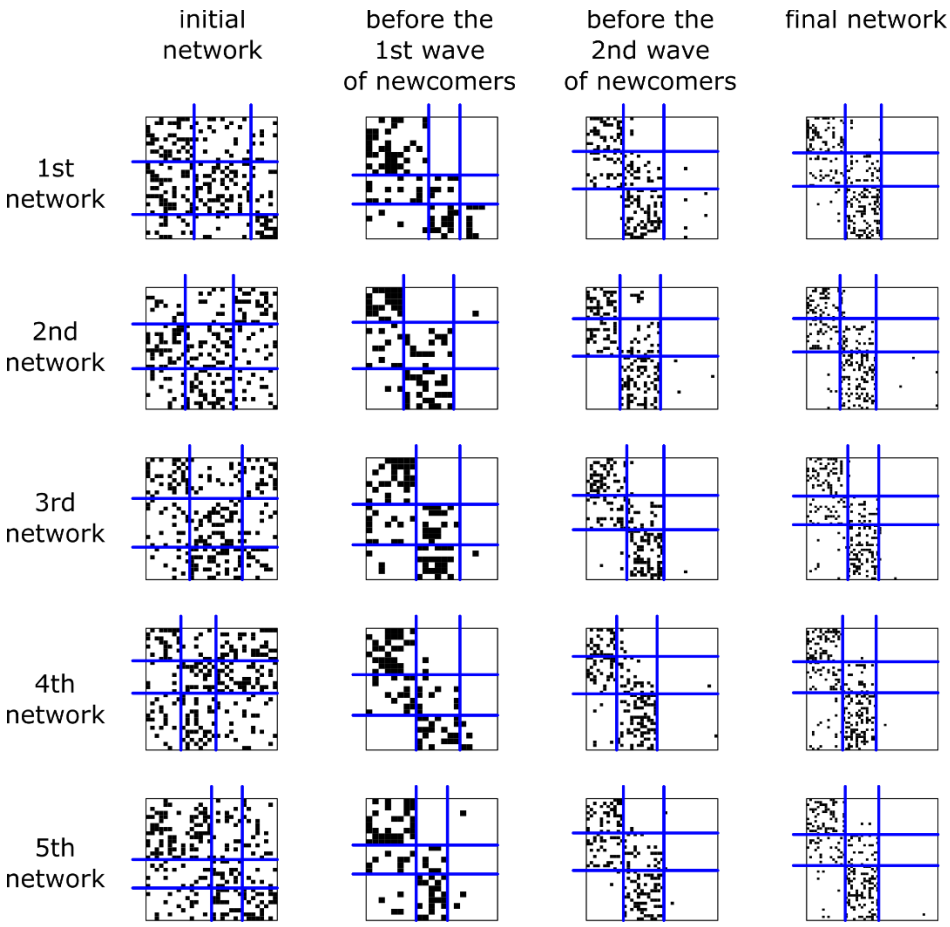
Figure 7.10: Mean relative fit values for different blockmodel types obtained at different time points (for the θ with ID 1861)



While positive weights of the mechanism hierarchical position promote links from the nodes on lower levels to the nodes on higher levels, the mechanism difference in hierarchical position between the ego and the alter prevent links being established from those with a very large difference in hierarchical position (e.g. from those in the lowest hierarchical position to those in the highest hierarchical position). Therefore, in order to prevent the emergence of, e.g. the transitivity blockmodel, both of the mentioned mechanisms must be considered.

The association between the weights of the mechanism hierarchical position and the mechanism distance between the ego and the alter indicates that only one or the other can be sufficient for the chosen blockmodel type to emerge (considering all the other local network mechanisms that are included).

Figure 7.11: Some randomly selected generated networks (by considering the θ with ID 1861) with a hierarchical-cohesive blockmodel last non-cohesive group



Note: Both newcomers and outgoers are possible. The signs of the mechanisms' weights are fixed. All local network mechanisms are considered. Networks are drawn in line with the blockmodeling solution for sparse networks (non-specified model).

7.10.2 Hierarchical-cohesive blockmodel

In the case of a hierarchical-cohesive blockmodel, all the nodes on the same hierarchical level are linked. Such a blockmodel might be less common in empirical knowledge-flow networks, especially because those from the cluster on the lowest hierarchical level are also those with the lowest average tenure. However, the flow of knowledge among those on the lowest hierarchical level can be encouraged by introducing appropriate policies.

By considering only tenure-related mechanisms or all but tenure-related mechanisms none of the θs generated more than 25 networks with the chosen blockmodel type. Yet, there are three θs for the case when all mechanisms are considered which generated 25 out of 30 networks containing

the chosen global network structure (Table 7.9). The RF values are lower than those reported for the hierarchical-cohesive blockmodel with last non-cohesive group but, considering the restriction of the out-degree values above 0.32 are considered as relatively high (see Appendix E). One must also take into account that all of the networks (including those with inconsistent blocks) are considered when computing the mean RF value.

Table 7.9: The selected θs for generating networks with a hierarchical-cohesive blockmodel

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter		
35	0.035	0.635	0.025	-0.038	-0.740	-0.186	-0.100	0.27	0.022
503	0.005	0.128	0.042	-0.113	-0.916	-0.003	-0.362	0.19	0.044
997	0.028	0.453	0.008	-0.001	-0.842	-0.002	-0.290	0.27	0.020

Note: Both newcomers and outgoers are possible. The signs of the mechanisms' weights are fixed. All local network mechanisms are considered. The number of generated networks with the chosen blockmodel is 25 out of 30 in all cases.

7.10.3 Other settings

The results presented in the previous subsections concern the case when both newcomers and outgoers are considered and when the signs of the mechanisms' weights are set in line with the previous knowledge on the local network mechanisms in knowledge-flow networks.

The results for the situation where only newcomers are possible (with constraints on the signs of the mechanisms' weights) are presented first. The sets of θs used here are the same as in the previous section.

What follows the results for the case when newcomers and outgoers are possible, but there are no constraints on the signs of the mechanisms' weights. Because the space of all possible θs is much bigger when no restriction is placed on the signs, the probability that none of the θs would generate networks with the chosen blockmodel is higher than in the case involving constraints on the mechanisms' weights[32]. This does not necessarily mean that it is impossible to generate networks

[32] If the θs are represented as points in k-dimensional space, the distance between points increases with the number of dimensions k, given a fixed number of θs.

with the chosen blockmodel by considering the selected local network mechanisms (in fact, it was shown in the previous subsections that this is possible), but it can be due to an insufficient number of θs being considered.

Considering only newcomers (with fixed signs of the mechanisms' weights)

It is assumed that considering only newcomers does not greatly affect the ability to generate networks with the chosen blockmodels since the outgoers are selected randomly. Further, the chosen blockmodels can emerge relatively early on – before the first wave of newcomers is added to the network.

<u>A hierarchical-cohesive blockmodel with last non-cohesive group:</u> there are many θs which generate networks with the chosen blockmodel in all 3 sets of local network mechanisms. The RF values are highest (around 0.50) when all local network mechanisms are considered (Table 3 in Appendix D). When only the tenure-related local network mechanisms are considered, the mean RF is the same (0.31) for all θs. This is expected since all the θs are extremely similar (Table 1 in Appendix D).

<u>A hierarchical-cohesive blockmodel:</u> no θ generates all the networks with the chosen blockmodel type while considering all three sets of local network mechanisms. In the case where all local network mechanisms are considered, there are 2 θs that generated 28 out of 30 networks with the chosen blockmodel (Table 6 in Appendix D). The mean RF values are 0.21 (for both θs), which is intermediate given the restriction on the maximum out-degree.

Generating networks with non-fixed mechanism weights

Among the generated θs, there is no such θ that generates all the networks with the chosen blockmodel. Yet some θs generated almost all the networks with the chosen blockmodel types.

<u>A hierarchical-cohesive blockmodel with last non-cohesive group:</u> there is one $\boldsymbol{\theta}$ which generates 29 out of 30 networks with the chosen blockmodel type when only newcomers are possible (Table 4 in Appendix D), and one $\boldsymbol{\theta}$ which generated 29 out of 30 networks with the chosen blockmodel for the case where both newcomers and outgoers are possible (Table 5 in Appendix D). The signs of the mechanisms' weights are (for both θs) as hypothesized in this chapter. The mean RF values are lower than observed for the case when the signs of the mechanisms' weights are fixed.

A hierarchical-cohesive blockmodel: there is one θ which generated 29 out of 30 networks with the chosen blockmodel type for the instance when only newcomers are possible (Table in Appendix D), and one θ which generated 26 out of 30 networks for the case when both newcomers and outgoers are possible (Table 5 in Appendix D). The signs of the mechanisms' weights are not as hypothesized in either case. The sign corresponding to the mechanism popularity level of the alter is negative, whereas the sign corresponding to the mechanism distance between the ego and the alter is positive. In addition, in the case of the θ with ID 1456 (for the case with newcomers and outgoers), the sign of the mechanism difference in hierarchical position between the ego and the alter is positive (and the value of the weight of the mechanism hierarchical position of the alter is low and negative: -0.03).

7.11 Conclusion

Understanding how knowledge flows among employees in a company can bring benefits in the employees' higher personal and professional development and thus in a greater competitive advantage for a company. By knowing the link between the local network mechanisms and the global network structure of knowledge-flow networks, a company is able to adopt policies that encourage different kinds of communication patterns among the employees, thereby promoting the most appropriate global network structure for knowledge flow. The decision on the most appropriate global network structure depends on the type of knowledge to be transferred (e.g. tacit vs. complex knowledge) as well as on the type of organization and its size.

In this chapter, the evolution of the global network structure in a given company was studied. It is an international medium-sized company dealing with software development, IT and business consulting, and maintenance and support. The knowledge-flow networks were measured by two name generators ("asking for advice" and "learning from") at three points in time (December 2002, July 2006 and April 2007). In order to avoid the impact of geographical location on the global network structure, only the data collected in Ljubljana (Slovenia) were analysed in this study.

The global network structures for each time point of the observed networks were analysed using direct blockmodeling. The blockmodel obtained at the first time point is close to cohesive, while at the second time point the global network structure consists of two core-peripheries which develop into the hierarchical global network structure at the third point in time. The global network

structure is highly influenced by the business unit and tenure. In such a hierarchical global network structure, the average tenure decreases with hierarchical position, indicating that knowledge flows from those with a higher tenure in higher hierarchical position to those with a lower tenure in a lower hierarchical position. Knowledge is exchanged among employees holding positions on the same hierarchical levels which is not the case for those in the lowest hierarchical position with the lowest average tenure. Based on these empirical results, the hierarchical-cohesive blockmodel with last non-cohesive group (with three clusters) and the hierarchical-cohesive blockmodel are chosen to be further studied. This does not imply that such a global network structure is the one that is desired for this or any other company. However, it is one of the structures that appears in the real world and thus warrants scholarly attention.

Once the global network structure is chosen, the local network mechanisms that might be related to how knowledge flows among the employees are identified. They are mostly chosen based on the theory proposed by Nebus (2006), which primarily addressed the mechanisms of advice-giving. The main research question is whether the selected local network mechanisms are able to drive the global network structure towards the proposed one. The selected local network mechanisms consider only tenure and not the other nodes' attributes. Therefore, the local network mechanisms are related to the popularity level of the nodes, the hierarchical level of the nodes, the (geodesic) distance between the nodes, the number of shared partners, and tenure. Networks were generated by considering these mechanisms by using the proposed algorithm from the family of network evolution models.

The main focus was paid to generating networks with the chosen blockmodels by considering newcomers and outgoers and by considering only the theoretically assumed signs of the weights of the mechanisms. However, the case when only newcomers are possible was also analysed along with considering positive and negative signs of the mechanisms' weights. It was also evaluated whether the chosen blockmodel emerges in the settings listed above.

The results of the Monte Carlo simulations show a hierarchical-cohesive blockmodel with last non-cohesive group can appear when considering all the non-tenure-related mechanisms or all of the considered mechanisms, including tenure. When only newcomers are possible (but not also outgoers), the chosen blockmodel also emerges when only tenure-related mechanisms are considered. There is a relatively low level of errors in the networks that are generated. It turns out

that probably the most important mechanisms are those relating to the tenure of the employees, the hierarchical position of the employees, and the distance between two employees. This is expected since these mechanisms are operationalizations of many important social constructs, including different kinds of cost (social cost, psychological cost, institutional cost) and distances (psychic distance, cognitive distance and geographical distance).

The chosen blockmodel type also emerges when the signs of the mechanisms are not defined in advance. It happens that, even in this case, the signs of the mechanisms which generate networks with the chosen blockmodel are completely (for a hierarchical-cohesive blockmodel with last non-cohesive group) or mostly (for a hierarchical-cohesive blockmodel) in line with the hypothesized ones.

The results are particularly valuable by confirming that the studied global network structures can emerge due to some local network mechanisms that are not related to the employees' attributes (except tenure). The latter holds practical implications because it indicates one can introduce some general policies to promote the emergence of the desired global network structures.

8 Discussion

One of the key attempts in both sociology and psychology is to reveal the (social) mechanisms responsible for a given (social) output. When the relationships among individuals are studied, the social output can be a social network. Social network analysis brings different approaches to studying the social mechanisms underlying a given network. The main focus of earlier studies was on social mechanisms in the context of empirical networks while less attention was paid to the social mechanisms in the context of specific global network structures. Therefore, the general goal of this study was to identify and understand the fundamental social mechanisms that are able to drive the formation of a global network structure. Here, the global network structure is operationalized by different blockmodel types. A blockmodel is defined as a network where the nodes are clusters of equivalent nodes from the studied network. The term block refers to a submatrix of an adjacency matrix and shows the relationship between two clusters or within a cluster (Doreian et al., 2005).

The most common blockmodel types are cohesive, core-periphery, transitive and hierarchical, although many other blockmodel types are possible. There are also many known local network mechanisms. Probably the best known are the mutuality, popularity, transitivity-related and assortativity-related mechanisms. Different local network mechanisms can contribute to the emergence of very different blockmodel types. Therefore, to narrow the scope of this study, only a few blockmodel types and a few local network mechanisms were considered. The attributes of the nodes are not meant to be considered in this study.

The blockmodels and local network mechanisms were selected within the chosen context. As mentioned, while considering the context it is helpful to narrow the number of all possible blockmodel types and to select the most appropriate local network mechanism. Taking the (social) context into account can also increase the quality of the study.

Two of such social contexts were considered. The first one relates to the environment of a preschool class (friendship/liking networks and interactional network are analysed in this setting) while the second one relates to the work environment of a medium-sized knowledge-based company (knowledge-flow networks are analysed in this context). The analyses address one of the

main research questions posed in this dissertation, namely: “Which mechanisms (or combination of several mechanisms) affect a change in blockmodel type?”.

The study begins by analysing the ability to generate the most common blockmodel types by considering only different triad types. This is important since such ability indicates the existence of local network mechanisms that would drive the global network structure towards the selected blockmodel (without considering the nodes’ attributes). This part of the study addresses another research question, which is: “Is it possible to generate networks with a given blockmodel type considering only the number of different types of triads?”.

8.1 The triad types

The relationship between the different triad types and the global network structures was extensively studied by the social network pioneers (Davis & Leinhardt, 1967; Holland & Leinhardt, 1970; Johnsen, 1985). Different triad types were used to test the existence of different global network structures in empirical networks (Holland & Leinhardt, 1970). This was done by comparing the number of different triad types in an observed network with the distribution of the number of triad types in random networks.

Nowadays, the term motif (Milo et al., 2002) is often used to study different aspects of global network structures. They are defined as “patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks” (Milo et al., 2002). The triad types can be considered a subset of all possible patterns. Different patterns with three or four nodes are typically used because considering patterns of size two would be insufficient, while considering patterns of a higher size might be computationally very intensive and less-informative in terms of global network structures. The triad types are seen as the smallest sociological unit from which the dynamic of a multi-person relationship can be observed (Davis & Leinhardt, 1967).

Although different triad types have often been used to describe the global network structures, they have yet to be systematically studied in the context of the most common blockmodel types. It is also known that the distribution of different triad types can be related to a given global network structure, although there is a lack of understanding of whether the distribution of different triad

types can cause the selected global network structure to emerge. This issue is addressed in Chapter 3.

Different triad types cannot be seen as “mechanisms” or “local network mechanisms” as defined in Section 1.1 because the local network structures cannot be indirectly used to explain how the behaviour of the individuals affects the global network structure. However, triad types can be used as a form of help while contemplating the possible relationship between local network mechanisms and global network structures, as often occurs in the context of Exponential Random Graph Modelling.

The most common blockmodels were considered in this study, namely cohesive blockmodel, symmetric and asymmetric core-periphery blockmodel, transitivity blockmodel, transitive-cohesive blockmodel, hierarchical blockmodel and hierarchical-cohesive blockmodel. The three clusters were assumed in all cases but in the asymmetric and symmetric core-periphery blockmodel types only two clusters are possible by definition. For each studied blockmodel type, different triad types were classified in the sets of allowed and forbidden triad types. The classification was obtained by considering the networks containing the selected ideal blockmodel (without any inconsistency). The set of allowed triad types consists of those triad types that appear in such network structures while the set of forbidden triad types is made up of those triad types that do not appear in such network structures. It turned out that one can distinguish different blockmodel types by only looking at the sets of allowed and forbidden triad types.

By considering these sets of triad types, two different algorithms were used to generate the networks in order to increase the reliability of the results. The first algorithm was the proposed Relocation of links algorithm, whereas the second one was the MCMC algorithm implemented in the “ergm” package for the R computer language.

All of the studied blockmodels can be generated by considering different sets of triad types. In general, the fit of the generated global network structures to the ideal global network structures is not significantly worse when the set of forbidden triad types is used as opposed to the case when the set of allowed triad types is used, although a very important difference arises while generating networks by considering one set or another.

When the networks are generated by the Relocation of links algorithm and the set of allowed triad types is used, the distribution of the triad types must be given. The information on the number of clusters and their sizes could be embedded in such a distribution of different triad types. On the other hand, considering the set of forbidden triad types only gives information on which types of triads are not allowed to appear in the network. This can still contain some information on the number of clusters (two vs. more than two clusters), but the amount of information the researcher needs to provide is much smaller.

Nevertheless, the generated networks with the target hierarchical blockmodel contained a greater amount of inconsistencies than the generated networks with other target blockmodel types, especially when the networks were generated using the MCMC algorithm. When using the RL algorithm by considering all triad types, the blockmodel was successfully generated, but the cluster of the nodes on the highest and the cluster of the nodes on the lowest hierarchical level were very small. Considering paths-of-length-three as an additional local network structure led to generated networks containing a very clear hierarchical blockmodel.

The main finding of this chapter is that the selected global network structures are able to emerge due to the selected local network structures even when the nodes' attributes are not considered. This is a good indicator that more complex local network mechanisms that produce a given global network structures might exist. Such local network mechanisms are addressed in the later chapters.

8.2 Emergence of symmetric and asymmetric core-cohesive blockmodels

Two versions of the core-cohesive blockmodel type were proposed. The asymmetric core-cohesive blockmodel type was proposed and analysed in Chapter 4 while the symmetric one was proposed and analysed in Chapter 5.

The proposed blockmodel type entails a combination of the cohesive and core-periphery blockmodel types. It consists of at least three clusters of nodes. In the asymmetric case, there is one cluster of nodes (called a core cluster) to which all the other nodes in the network are linked. The other clusters (called cohesive clusters) are internally well linked, while the nodes from different clusters are not linked to each other. In the symmetric case, there are symmetric links between the nodes from the core cluster and the nodes from the cohesive clusters.

Based on previous studies, it was assumed that such global network structures might appear in friendship-nomination networks or in liking networks (for the asymmetric case) and in interactional networks among pre-schoolers (for the symmetric case). The extensive literature review revealed the most commonly studied local network mechanisms in such networks. These are mutuality, popularity, (in-degree-related) assortativity and different transitivity-related mechanisms. Other very commonly studied mechanisms in this context are different kinds of homophiles, which are related to various types of nodes' attributes. Due to the limitation to only consider network-related characteristics of the nodes in this study, such local network mechanisms were not taken into account.

To address the research question of whether the selected local network mechanisms can drive the global network structure towards the symmetric or asymmetric core-cohesive blockmodel, an algorithm from the family of the network evolution algorithms was proposed. The algorithm is iterative. One node is randomly selected at each iteration. Then, the selected node assigns a link to another randomly selected node with the highest weighted network statistics. At the same time, the selected node assigns a non-link to a randomly selected node with the lowest value of the weighted network statistics. The network statistics are calculated by the linear combination of the mechanisms. The weights of the local network statistics are assigned by the researcher (or randomly generated) and reflect the importance of a given local network mechanism. Even though some effort was made to standardize the local network statistics, the weights assigned to the mechanisms are not generally comparable since the local network mechanisms are dependent. A similar problem of comparability of the coefficient is also present within ERGM and SAOM. Although several researchers have addressed this issue (Indlekofer & Brandes, 2013; Snijders, 2004; Snijders, Van de Bunt, et al., 2010), no generally acceptable solution is available. However, the signs of the mechanisms' weights can be compared. In addition, it is assumed that one can roughly compare the mechanisms' importance by looking at the extremely large differences in their weights.

The latter is important because the mechanisms' weights were generated randomly. Many networks were then generated and their global network structures evaluated. Those sets of mechanisms' weights that generated the networks as close to the chosen blockmodel type as possible were interpreted.

8.2.1 The asymmetric case

The results show that an asymmetric core-cohesive blockmodel can emerge due to the mutuality, popularity, assortativity, and two types of transitivity mechanisms (namely, outgoing-two paths mechanisms also referred to as the transitivity mechanism and outgoing shared partners mechanisms). This is true for all of the initial global network structures that were considered: empty network, asymmetric core-periphery blockmodel and cohesive blockmodel.

For each blockmodel type, the ten best (based on the mean number of inconsistent blocks) sets of mechanism weights were chosen and further analysed for each initial blockmodel type. It appears that similar sets of weights of mechanisms were chosen for the case when the initial network is empty and when the initial network has a cohesive blockmodel. The set of mechanism weights is different, to some extent, when the initial global network structure was the asymmetric core-periphery blockmodel. In this case, the weights of the transitivity-related mechanisms were higher. Higher weights of these mechanisms are expected since only the cohesive clusters have to emerge in the network with the asymmetric core-periphery blockmodel to end up with the asymmetric core-cohesive blockmodel.

However, this is not to argue that the weights of the selected local network mechanisms are unable to drive the network towards the chosen one from the initial network with any blockmodel type. In some cases, only the number of iterations must be increased to reach the chosen blockmodel structure.

8.2.2 The symmetric case

The results for the symmetric core-cohesive blockmodel are given in Chapter 4. Along with the main research question on the emergence of a symmetric core-cohesive blockmodel, the presence of the proposed blockmodel type in real-life networks was addressed. To this end, interactional data collected in Head Start preschools in the USA were analysed using the blockmodeling approach. The proposed blockmodel type was found in almost all classes that were studied.

The simulation part provided similar results as for the asymmetric core-cohesive blockmodel – a symmetric core-cohesive blockmodel can emerge as consequence of the mutuality, popularity, assortativity and transitivity-related mechanisms.

The results were expected for the following main reasons. First, friendships (or liking) among pre-schoolers are assumed to initiate interactions. This is also why similar local network mechanisms were considered in both cases (symmetric and asymmetric). Second, the same algorithm was used to generate the networks in both cases, yet several possible approaches for modelling symmetric networks exist (as discussed in the corresponding chapter). However, the one used in this study is the closest representation of the assumed emergence in the empirical networks.

8.2.3 Other blockmodel types

Chapter 6 sets out the results concerning whether it is possible to generate the most common blockmodel types (other than the symmetric and asymmetric core-cohesive blockmodel) by considering the same local network mechanisms as with the asymmetric core-cohesive blockmodel.

One of the most important observations is that the popularity mechanism leads the global network structure toward the asymmetric core-periphery while a combination of popularity and mutuality leads the global network structure towards the symmetric core-periphery blockmodel. The transitivity mechanism (also called the "closing triads" mechanism) is often related to the emergence of cohesive clusters in the case of undirected networks.

There are several types of transitivity-related mechanisms in the case of directed networks. The outgoing-two-paths (also called "transitivity") mechanism and outgoing-shared-partners mechanisms were considered in this study. Although the outgoing-two-paths mechanism looks more like the transitivity mechanism, it promotes the emergence of the asymmetric core-periphery blockmodel and not a cohesive blockmodel as one would expect. On the contrary, the outgoing-shared-partners mechanism leads the global network structure towards the cohesive one.

The other blockmodel types cannot be generated by considering the selected local network mechanisms (one at a time). Further, by considering all of the selected local network mechanisms we were unable to generate clear hierarchical blockmodels, hierarchical-cohesive blockmodels, transitivity blockmodels and transitive-cohesive blockmodels.

8.3 Emergence of a hierarchical blockmodel in knowledge-flow networks

The knowledge-flow concept was used in Chapter 5 to study the emergence of two types of hierarchical-cohesive blockmodels. A link in a knowledge-flow network operationalizes the flow of knowledge while the nodes are the employees in an organization/company.

One of the studied blockmodels is hierarchical-cohesive, whereas another is hierarchical-cohesive blockmodel with last non-cohesive group. The latter means that nodes on the same hierarchical level are linked to each other, while those on the lowest hierarchical level are not linked to each other. There are empirical evidences that the latter blockmodel might appear in knowledge-flow networks.

The local network mechanisms that might drive the global network structure to the selected one were chosen based on previous studies on advice-giving networks, learning networks and other kinds of networks that are used to operationalize the flow of knowledge in a company/organization. However, the main set of local network mechanisms was selected based on the theory proposed by Nebus (2006) who considered the advice-giving relations (which form part of knowledge flow). He assumed that the employees consider the value of the advice obtained from a given other and the cost of obtaining advice from a given other. Therefore, the selected local network mechanisms can be classified in two sets of mechanisms (Nebus, 2006). The first set (value-related mechanisms) contains the following local network mechanisms: popularity of the alter, hierarchical position of the alter, tenure of the alter, and outgoing shared partners. The second set (cost-related mechanisms) contains: difference in tenure between the ego and the alter, distance between the ego and the alter, difference in hierarchical position between the ego and the alter and outgoing shared partners (the latter can be considered as a cost-related mechanism as well as a value-related mechanism). All of these mechanisms are operationalizations of different sociological and psychological constructs. For example, the distance between the ego and the alter is an operationalization of psychic distance, cognitive distance and geographical distance, while higher popularity of the alter can be an indicator of his willingness to share knowledge and can increase the level of cognitive trust from the ego to the alter.

To study the research question regarding whether the selected local network mechanisms can drive the global network structure towards the chosen blockmodel, a similar methodology was used as

in the previous chapters. Yet, the proposed algorithm for generating networks is more complex because it considers many characteristics of the knowledge flow in a company.

Compared to the algorithms used in other chapters, the algorithm proposed in Chapter 6 is the only one to consider newcomers and outgoers. Moreover, the dissolution of links is not related to any local network mechanism. Instead, the duration of links (flow of knowledge) is time-limited and related to the chosen out-degree of a unit.

The general research question was broken down into several sub-questions. For example, the case where only tenure-related mechanisms and the case where all but tenure-related mechanisms were considered. In addition, the signs of the weights of the mechanisms were fixed in some cases and were set free in others. The signs of the mechanisms' weights are related to how the mechanisms can be interpreted. For example, a positive sign of the mechanism popularity of the alter means the tendency to ask for advice from those with a higher level of popularity, while a negative sign means the ego would avoid asking those with higher levels of popularity. Considering non-constrained signs of the mechanisms' weights increases the computational cost, but can reveal non-expected weights of the mechanisms that can drive the global network structure towards the proposed blockmodel type.

Considering newcomers and outgoers, generating networks with any of the two chosen blockmodel types was only successful when all the selected local network mechanisms were considered. The global network structures were clearer in the case of a hierarchical-cohesive blockmodel with last non-cohesive group. This confirms the studied hierarchical structures can emerge when subjected to the selected local network mechanisms. Here, it has to be noted that the networks were generated by considering fixed signs of the mechanisms' weights. However, the signs are also consistent with the theory when there are no constraints on the weights (for a hierarchical-cohesive blockmodel with last non-cohesive group).

8.4 Relevance of the study

This study is the first to attempt analysis of the link between the local network mechanisms and the emergence of blockmodel types without considering the nodes' attributes. The results clearly show that some blockmodels are able to emerge because of the local network mechanisms, without

considering the nodes' attributes. This means that general policies which promote the emergence of the chosen global network structures can be introduced.

The results are especially valuable by bringing a better understanding of how the considered blockmodels emerge. A very important observation in this study is that a researcher should not make conclusions with respect to the local network mechanisms based on the global network structure of one empirical network. This is also because different intermediate global network structures can emerge during its evolutionary process and the researcher usually does not know at which step they are observing an empirical network. Therefore, it is necessary to consider several observations and the social context of the study. However, even this (observing networks at several points in time) does not guarantee the global network structure being observed will not change later in time (under the influence of the same local network mechanisms and their strengths). The whole evolutionary process of the emergence of a given blockmodel type, when subjected to the selected local network mechanisms, can be used to better understand and predict the dynamics of global network structures described by blockmodel types.

The proposed network evolution models can also be used to generate networks with the chosen blockmodel with some inconsistencies (errors). Compared to other approaches for generating random networks with a chosen blockmodel, the proposed one generates network structures by considering the selected local network mechanisms, meaning that the errors in the blockmodels that are obtained are not random but are in line with the selected local network mechanisms. Networks at different stages of the evolution of the global network structure can be taken and further analysed.

This study proposes and evaluates a methodology for studying the relationship between local network mechanisms and global network structures. This includes an approach to generate networks by considering local network mechanisms as well as an approach to evaluate the global network structures of the networks so generated. Although this was not the main aim of this study, some important steps were made in development of normalizing the criterion function which can be used to compare two blockmodels, select the number of clusters or select a blockmodel type in empirical data.

8.5 The main limitations

Even though local network mechanisms are often a central interest in empirical studies on social networks, the question about the relationship between different local network mechanisms and different blockmodels has not yet been systematically addressed. One of the reasons might be that it is not trivial to address such a question without considering the (social) context, particularly because the number of existing local network mechanisms is enormous.

Therefore, this study may be seen as a very small step towards better understanding the link between the local network mechanisms and global network structures described by blockmodels, which are very precise representations of the global network structure. Given that the research question is very wide, several assumptions and restrictions were applied in this research.

8.5.1 Limitations related to local network mechanisms

One of the strongest assumptions is that local network mechanisms affect individuals' behaviour and, through that, the global network structure, while the global network structure does not directly affect the importance (or strength) of the local network mechanisms (Doreian & Conti 2012). Closely related to this is the assumption that the local network mechanisms hold the same importance for all of the nodes. To our knowledge, both assumptions are rarely explicitly addressed in empirical studies. In the case of ERGM and SAOM, it is possible to estimate weights for the mechanisms (or effects, terms) separately for different clusters of nodes, where clusters are determined based on some nodes' attributes. These can be related to any personal characteristics (such as gender) or to the location in the network (e.g. being part of the core or the periphery). Based on meta-analysis of friendship networks, Block (2015) argues that embeddedness in a transitive triad is more likely to preserve unreciprocated ties because transitive triads provide a forum for social interactions that would not otherwise exist. While for interactional preschool networks, Schaefer et al. (2010) proposed that some less complex local network mechanisms might be more common for younger children while more complex local network mechanisms emerge with their psychological development and are therefore more common among older children.

While different mechanisms can be of different importance to different nodes, one could also consider the case when the weights of different local network mechanisms can change over time.

For example, Schaefer et al. (2010) analysed interactional networks collected among preschool children and argued that popularity and triadic closure are becoming increasingly important over the course of the school year. To apply the methodology out forward in this dissertation, one should propose the way to normalize the weights of the mechanisms so that are comparable among each other and in time. Both SAOM and (temporal)ERGM, which are probably the most commonly used to study the dynamics of the links in empirical networks, assume that the mechanisms' strengths are constant over time.

All of the proposed network evolution models assume all the nodes in the network have equal probabilities of having a chance to establish a link. This is a reasonable assumption, especially because the proposed models are defined such that is it possible (for a given unit) not to change any link when receiving an opportunity to do so. Yet, it would be beneficial to take also into account the case where some nodes would receive fewer opportunities to change a link. For example, in knowledge-flow networks one could consider the situation when the nodes with a lower tenure would be more keen to ask for advice than those with a higher tenure. In case of preschool networks, those with a higher number of mutual ties would receive fewer opportunities to establish new ties (Daniel et al., 2019). Such constraints could be considered within the model (as part of the definition of the algorithm for generating the networks) or be operationalized by different local network mechanisms.

8.5.2 Limitations related to blockmodel

In this study, the focus was given to blockmodels obtained by considering the structural equivalence, since that is very frequently used in real studies. However, some further focus should be paid to studying the relationships between different local network mechanisms and blockmodels with other equivalences.

The current study largely focuses on blockmodels with three clusters. One should verify whether the results can be generalized to blockmodels with a higher number of clusters. The main concern here is whether only the strengths of the considered local network mechanisms should be adapted or whether some additional local network mechanisms are needed.

The very early scientists (Cartwright & Harary, 1956; Moreno, 1934) in the field of social network analysis considered positive and negative ties. Soon, more focus was paid to only the positive ties.

Yet, some researchers argue today that considering negative ties (or repulsion-type mechanisms) is extremely important when studying the evolution of the global network structure. Stadtfeld et al. (2018) argue that positive relations are insufficient to explain the emergence of groups, while Doreian & Mrvar (2014) showed that considering only positive ties can bring spurious results. This is because global network structures are the result of several social processes, which include liking-related and disliking-related social mechanisms (they considered structural balance, differential popularity, differential dislike, and mutual hostility within subgroups larger than dyads). Therefore, studies on the link between local network mechanisms and global network structures should consider both positive and negative ties. Such research results would provide very important practical considerations with respect to the importance of collecting information on negative ties in empirical studies. For example, in the case of a preschool social context, negative ties are particularly important since they can be used to operationalize social behaviour such as bullying. Bullying is a repetitive and intentionally negative behaviour against a victim who finds it difficult to defend him or herself (Olweus, 1994). One of the purposes of bullying is to increase the perpetrator's social status (Caravita, Di Blasio, & Salmivalli, 2009; De Bruyn, Cillessen, & Wissink, 2010; Salmivalli, Lagerspetz, Björkqvist, Österman, & Kaukiainen, 1996). However, there are differences between different age groups in how they react to bullies. While younger children sanction bullying, older children reward the proponents of bullying with a higher social status (Van der Ploeg, Steglich, & Veenstra, in press).

8.6 Recommendations for future research work

Some important directions for future studies are given in the previous subsection that also discussed the study's limitations. Here, some additional recommendations for future research work are given, which do not relate directly to those limitations.

One very important methodological issue in this study was how to evaluate the fit of the empirical network structures to the chosen blockmodels. A very general insight was obtained by calculating the number of inconsistent blocks. The number of inconsistent blocks was used to evaluate how similar the empirical blockmodel and selected ideal blockmodel type are. To evaluate the amount of errors in the empirical network, the normalized value of the criterion function (called relative fit, RF) was defined. The RF takes a value of up to 1 (an ideal blockmodel). The expected value

in the case of a random network with the same density as in an empirical network is 0 (meaning that negative values are also possible). This approach is appropriate when it is reasonable to assume there are no constraints preventing the emergence of the ideal blockmodel without errors. Such an example was provided in Chapter 5, where the out-degree of the nodes was constrained by the algorithm. Because the obtained networks were sparse, the direct blockmodeling approach for sparse networks was applied. By using this approach, the complete blocks are usually sparser than in the case of blockmodeling for "regular networks" (another concern is that in this approach the errors in null and complete blocks are weighted differently, usually based on the network density). As a result, the real value of RF is below 1 and the guidelines for interpreting the RF given in subsection 2.5.4 are too conservative when the blockmodeling approach for sparse networks is used. Therefore, the maximum RF value for the generated networks was estimated by simulations in this study. Still, a more detailed and systematic study on the behaviour of the criterion function and the RF is needed[33].

In this study, several algorithms for generating networks by considering selected local network mechanisms were proposed. They essentially all come from the same family of network evolution models. Some differences between them and SAOM and ERGM are described in Chapter 2. During this study, it was shown that ERGM and SAOM often fail to estimate the models which would generate networks with a very clear blockmodel structure. This occurs even when the set of local network mechanisms (or terms or effect) which generated the global network structures are included in the model. Such global network structures can be generated by considering the same local network mechanisms with the proposed NEM algorithms. The inability to generate networks with SAOM and ERGM could be due to the very unnatural degree distribution. In the case of data from real life, such distributions of the degrees usually arise by virtue of certain constraints on the creating of links. This indicates the need to further develop the already existing models (e.g. ERGM and SAOM) such that the blockmodel structure would be considered along with the already existing terms or effects. Another possibility would be to develop additional models (based on the proposed NEM) that would enable networks to be created with a very clear blockmodel structure

[33] There are also other concerns about the behaviour of the criterion function related to the number of groups, the pre-specification of the model (non-specified vs. specified model) etc.

or to estimate the model's parameters based on an empirical network with a very clear blockmodel structure.

The current research provides the methodology for studying the relationship between local network mechanisms and global network structures. This (extended) methodology could be used to propose a typology of selected local network mechanisms that drive the networks towards the specific blockmodels. This could be used as a framework for analysing empirical networks and also for generating random networks with a given blockmodel, which is useful when testing dynamic blockmodeling algorithms (Matias & Miele, 2015; Xing, Fu, Song, & others, 2010; Xu & Hero, 2014).

In this study, only binary (binarized) networks were analysed. Therefore, the research could also be generalized for valued networks and other blockmodel types that were not considered in this dissertation.

9 Bibliography

1. Adams, R. G., & Torr, R. (1998). Factors underlying the structure of older adult friendship networks. *Social Networks*, *20*(1), 51–61.
2. Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, *9*(1), 1981–2014.
3. Alderson, A. S., & Beckfield, J. (2004). Power and Position in the World City System. *American Journal of Sociology*, *109*(4), 811–851.
4. Anchuri, P., & Magdon-Ismail, M. (2012). Communities and balance in signed networks: A spectral approach. *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining* (pp. 235–242). USA: IEEE Computer Society. Retrieved from https://ieeexplore.ieee.org/document/6425757
5. Anderson, C. J., Wasserman, S., & Faust, K. (1992). Building stochastic blockmodels. *Social Networks*, *14*(1–2), 137–161.
6. Artzy-Randrup, Y., Fleishman, S. J., Ben-Tal, N., & Stone, L. (2004). Comment on "Network motifs: Simple building blocks of complex networks" and "Superfamilies of evolved and designed networks." *Science*, *305*(5687), 1107–1107.
7. Barnett, G. A., & Danowski, J. A. (1992). The Structure of Communication: A Network Analysis of the International Communication Association. *Human Communication Research*, *19*(2), 264–285.
8. Batagelj, V. (1997). Notes on blockmodeling. *Social Networks*, *19*(2), 143–155.
9. Batagelj, V., Doreian, P., & Ferligoj, A. (1992). An optimizational approach to regular equivalence. *Social Networks*, *14*(1–2), 121–135.
10. Batagelj, V., Doreian, P., Ferligoj, A., & Kejžar, N. (2014). Understanding large temporal networks and spatial networks: Exploration, pattern searching, visualization and network evolution (Vol. 2). John Wiley & Sons.
11. Batagelj, V., Ferligoj, A., & Doreian, P. (1992). Direct and indirect methods for structural equivalence. *Social Networks*, *14*(1–2), 63–90.
12. Batagelj, V., Ferligoj, A., & Doreian, P. (1998). Fitting pre-specified blockmodels. Hayashi, C., Ohsumi, N., Yajima, K., Tanaka, Y., Bock, H., & Baba, Y. (Eds.) *Data science, classification, and related methods* (pp. 199–206). Japan: Springer.

13. Batagelj, V., Mrvar, A., Ferligoj, A., & Doreian, P. (2004). Generalized blockmodeling with Pajek. *Metodološki Zvezki*, *1*(2), 455–467.
14. Bettencourt, L. M., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, *3*(3), 210–221.
15. Bianchi, F., & Squazzoni, F. (2015). Agent-based models in sociology: Agent-based models in sociology. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(4), 284–306.
16. Bianconi, G., Darst, R. K., Iacovacci, J., & Fortunato, S. (2014). Triadic closure as a basic generating mechanism of communities in complex networks. *Physical Review E*, *90*(4), 042806.
17. Bickel, P. J., & Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*(1), 253–273.
18. Block, P. (2015). Reciprocity, transitivity, and the mysterious three-cycle. *Social Networks*, *40*(1), 163–173.
19. Block, P., Stadtfeld, C., & Snijders, T. (2016). Forms of Dependence Comparing SAOMs and ERGMs From Basic Principles. *Sociological Methods & Research*, *48*(1), 202–239.
20. Borgatti, S. P., & Everett, M. G. (1992). Notions of position in social network analysis. *Sociological Methodology*, *22*(1), 1–35.
21. Borgatti, S. P., & Everett, M. G. (1993). Two algorithms for computing regular equivalence. *Social Networks*, *15*(4), 361–376.
22. Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, *21*(4), 375–395.
23. Bottger, P. C. (1984). Expertise and air time as bases of actual and perceived influence in problem-solving groups. *Journal of Applied Psychology*, *69*(2), 214–221.
24. Boudon, R. (1991). What middle-range theories are. *Contemporary Sociology*, *20*(4), 519–522.
25. Boyd, J. P. (2002). Finding and testing regular equivalence. *Social Networks*, *24*(4), 315–331.
26. Boyd, J. P., Fitzgerald, W. J., & Beck, R. J. (2006). Computing core/periphery structures and permutation tests for social relations data. *Social Networks*, *28*(2), 165–178.

27. Boyd, J. P., & Jonas, K. J. (2001). Are social equivalences ever regular? Permutation and exact tests. *Social Networks*, *23*(2), 87–123.
28. Brown, T. M., & Miller, C. E. (2000). Communication networks in task-performing groups: Effects of task complexity, time pressure, and interpersonal dominance. *Small Group Research*, *31*(2), 131–157.
29. Brusco, M. J., & Doreian, P. (2019). Partitioning signed networks using relocation heuristics, tabu search, and variable neighborhood search. *Social Networks*, *56*(1), 70–80.
30. Burt, R. (2009). *Structural holes: The social structure of competition*. England: Harvard University Press.
31. Burt, R. S. & Ronchi, D. (1990). Contested control in a large manufacturing plant. Weesie, J. & Flap, H. D. (Eds.), *Social Networks through Time* (pp. 121–157). Utrecht: Isor.
32. Caravita, S. C., Di Blasio, P., & Salmivalli, C. (2009). Unique and interactive effects of empathy and social status on involvement in bullying. *Social Development*, *18*(1), 140–163.
33. Carley, K. M., & Krackhardt, D. (1996). Cognitive inconsistencies and non-symmetric friendship. *Social Networks*, *18*(1), 1–27.
34. Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review*, *63*(5), 277–293.
35. Cattani, G., & Ferriani, S. (2008). A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry. *Organization Science*, *19*(6), 824–844.
36. Chase, I. D. (1982). Dynamics of hierarchy formation: The sequential development of dominance relationships. *Behaviour*, *80*(3), 218–239.
37. Chen, K., & Lei, J. (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, *113*(521), 241–251.
38. Chen, Y., Wang, X. L., Yuan, B., & Tang, B. Z. (2014). Overlapping community detection in networks with positive and negative links. *Journal of Statistical Mechanics: Theory and Experiment*, *2014*(3), P03021.

39. Chinchilla-Rodríguez, Z., Ferligoj, A., Miguel, S., Kronegger, L., & De Moya-Anegón, F. (2012). Blockmodeling of co-authorship networks in library and information science in Argentina: A case study. *Scientometrics*, *93*(3), 699–717.
40. Cillessen, A. H., & Rose, A. J. (2005). Understanding popularity in the peer system. *Current Directions in Psychological Science*, *14*(2), 102–105.
41. Clauset, A., Moore, C., & Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, *453*(7191), 98–101.
42. Cohen, W. M., & Levinthal, D. A. (2000). Absorptive capacity: A new perspective on learning and innovation. Cross, R. & Israelit, S. (Eds.) *Strategic Learning in a Knowledge Economy* (pp. 39–67). USA: Elsevier.
43. Coleman, S. J. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology*, *91*(6), 1309–1335.
44. Coleman, S. J. (1994). *Foundations of social theory*. England: Harvard University Press.
45. Cook, K. S., & Emerson, R. M. (1978). Power, equity and commitment in exchange networks. *American Sociological Review*, *43*(5), 721–739.
46. Corander, J., Dahmström, K., & Dahmström, P. (1998). *Maximum likelihood estimation for Markov graphs*. Sweden: University of Stockholm, Department of Statistics.
47. Crick, F. (1989). *What mad pursuit: A personal view of scientific discovery*. England: Pengiun Books.
48. Crockett, L., Losoff, M., & Petersen, A. C. (1984). Perceptions of the peer group and friendship in early adolescence. *The Journal of Early Adolescence*, *4*(2), 155–181.
49. Cross, R., Borgatti, S. P., & Parker, A. (2001). Beyond answers: Dimensions of the advice network. *Social Networks*, *23*(3), 215–235.
50. Csermely, P., London, A., Wu, L.-Y., & Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, *1*(2), 93–123.
51. Cugmas, M., & Ferligoj, A. (2018). Comparing two partitions of non-equal sets of units. *Metodološki Zvezki*, *15*(1), 1–21.
52. Cugmas, M., Ferligoj, A., & Kronegger, L. (2016). The stability of co-authorship structures. *Scientometrics*, *106*(1), 163–186.
53. Cummings, J. N. (2004). Work groups, structural diversity, and knowledge sharing in a global organization. *Management Science*, *50*(3), 352–364.

54. Da Silva, M. R., Ma, H., & Zeng, A.-P. (2008). Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. *Proceedings of the IEEE*, *96*(8), 1411–1420.
55. Daniel, J. R., Santos, A. J., Fernandes, C., & Vaughn, B. E. (2019). Network dynamics of affiliative ties in preschool peer groups. *Social Networks*, *57*(1), 63–69.
56. Daniel, J. R., Santos, A. J., Peceguina, I., & Vaughn, B. E. (2013). Exponential random graph models of preschool affiliative networks. *Social Networks*, *35*(1), 25–30.
57. Daudin, J.-J., Picard, F., & Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing*, *18*(2), 173–183.
58. Davis, J. A. (1967). Clustering and structural balance in graphs. *Human Relations*, *20*(2), 181–187.
59. Davis, J. A., & Leinhardt, S. (1967). *The structure of positive interpersonal relations in small groups*. ERIC. Retrieved from https://eric.ed.gov/?id=ED024086
60. De Bruyn, E. H., Cillessen, A. H., & Wissink, I. B. (2010). Associations of peer acceptance and perceived popularity with bullying and victimization in early adolescence. *The Journal of Early Adolescence*, *30*(4), 543–566.
61. De Nooy, W., Mrvar, A., & Batagelj, V. (2018). *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software.* England: Cambridge University Press.
62. DeLay, D., Hanish, L. D., Martin, C. L., & Fabes, R. A. (2016). Peer Effects on Head Start Children's Preschool Competency. *Developmental Psychology*, *52*(1), 58–70.
63. Desmarais, B. A., & Cranmer, S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and Its Applications*, *391*(4), 1865–1876.
64. Dijkstra, J. K., Cillessen, A. H., & Borch, C. (2013). Popularity and adolescent friendship networks: Selection and influence dynamics. *Developmental Psychology*, *49*(7), 1242–1252.
65. Doreian, P. (2008). A multiple indicator approach to blockmodeling signed networks. *Social Networks*, *30*(3), 247–258.
66. Doreian, P., Batagelj, V., & Ferligoj, A. (1994). Partitioning networks based on generalized concepts of equivalence. *Journal of Mathematical Sociology*, *19*(1), 1–27.

67. Doreian, P., Batagelj, V., & Ferligoj, A. (2005). *Generalized blockmodeling*. England: Cambridge University Press.
68. Doreian, P., & Conti, N. (2012). Social context, spatial structure and social network structure. *Social Networks*, *34*(1), 32–46.
69. Doreian, P., & Mrvar, A. (2009). Partitioning signed social networks. *Social Networks*, *31*(1), 1–11.
70. Doreian, P., & Mrvar, A. (2014). Testing Two Theories for Generating Signed Networks Using Real Data. *Advances in Methodology and Statistics*, *11*(1), 31–63.
71. Doreian, P., & Mrvar, A. (2016). Identifying fragments in networks for structural balance and tracking the levels of balance over time. *Connections*, *35*(2), 6–18.
72. Dougherty, D. (1992). Interpretive barriers to successful product innovation in large firms. *Organization Science*, *3*(2), 179–202.
73. Dunbar, R. I. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493.
74. Engle, J. M., McElwain, N. L., & Lasky, N. (2011). Presence and quality of kindergarten children's friendships: Concurrent and longitudinal associations with child adjustment in the early school years. *Infant and Child Development*, *20*(4), 365–386.
75. Epstein, J. M. (2006). Generative social science: Studies in agent-based computational modeling. Princeton University Press.
76. Erdős, P., & Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae (Debrecen)*, *6*(1), 290–297.
77. Estell, D. B. (2007). Aggression, social status, and affiliation in kindergarten children: A preliminary study. *Education and Treatment of Children*, *30*(2), 53–72.
78. Everett, M., Boyd, J., & Borgatti, S. (1990). Ego-Centered and Local Roles: A Graph Theoretic Approach. *Journal of Mathematical Sociology*, *15*(3/4), 163–172.
79. Fararo, T. J., & Skvoretz, J. (1986). E-state structuralism: A theoretical method. *American Sociological Review*, *51*(5), 591–602.
80. Faust, K. (2006). Comparing social networks: Size, density, and local structure. *Metodološki Zvezki*, *3*(2), 185.
81. Faust, K. (2007). Very local structure in social networks. *Sociological Methodology*, *37*(1), 209–256.

82. Faust, K. (2010). A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks*, *32*(3), 221–233.
83. Foster, D. V., Foster, J. G., Grassberger, P., & Paczuski, M. (2011). Clustering drives assortativity and community structure in ensembles of networks. *Physical Review E*, *84*(6), 066117.
84. Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*(395), 832–842.
85. Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.
86. Gambetta, D. (1998). Concatenations of mechanisms. Elster, J. & McPherson, S. (Eds.) *Social mechanisms: An analytical approach to social theory* (pp. 102–124). England: Cambridge University Press.
87. Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, *30*(4), 1141–1144.
88. Gilbert, N., & Troitzsch, K. (1999). *Simulation for the social scientist*. England: Open University Press.
89. Glückler, J., & Panitz, R. (2016). Unpacking Social Divisions of Labor in Markets: Generalized Blockmodeling and the Network Boom in Stock Photography. *Social Networks*, *47*(1), 156–166.
90. Goldtborpe, J. H. (1996). The Quantitative Analysis of Large-Scale Data-sets and Rational Action Theory: For a Sociological Alliance. *European Sociological Review*, *12*(2), 109–126.
91. Goldthorpe, J. H. (2001). Causation, Statistics, and Sociology. *European Sociological Review*, *17*(1), 1–20.
92. González, M. C., Herrmann, H. J., Kertész, J., & Vicsek, T. (2007). Community structure and ethnic preferences in school friendship networks. *Physica A: Statistical Mechanics and Its Applications*, *379*(1), 307–316.
93. Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks*, *29*(2), 231–248.

94. Guimerà, R., & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, *106*(52), 22073–22078.

95. Hallinan, M. T. (1978). The process of friendship formation. *Social Networks*, *1*(2), 193–210.

96. Handcock, M., Robins, G., Snijders, T., Moody, J., & Besag, J. (2003). Assessing degeneracy in statistical models of social networks. *Journal of the American Statistical Association*, *76*(1), 33–50.

97. Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, *4*(1), 585–605.

98. Harré, R. (1970). *The principles of scientific thinking*. Chicago: University of Chicago Press.

99. Hartup, W. W. (1993). Adolescents and their friends. *New Directions for Child and Adolescent Development*, *1993*(60), 3–22.

100. Hartup, W. W., & Stevens, N. (1997). Friendships and adaptation in the life course. *Psychological Bulletin*, *121*(3), 355–370.

101. Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

102. Hedström, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. England: Cambridge University Press.

103. Hedström, P., & Bearman, P. (2009). What is analytical sociology all about? An introductory essay. Hedström, P. & Bearman, P. (Eds.), *The Oxford Handbook of Analytical Sociology* (pp. 3–24). England: Oxford University Press.

104. Hedström, P., & Manzo, G. (2015). Recent trends in agent-based computational research: A brief introduction. *Sociological Methods & Research*, *44*(2), 179–185.

105. Hedström, P., & Swedberg, R. (1996). Social mechanisms. *Acta Sociologica*, *39*(3), 281–308.

106. Hedström, P., & Swedberg, R. (1998). Social mechanisms: An introductory essay. Elster, J. & McPherson, S. (Eds.) *Social mechanisms: An analytical approach to social theory* (pp. 1–31). England: Cambridge University Press.

107. Hedström, P., & Wennberg, K. (2017). Causal mechanisms in organization and innovation studies. *Innovation*, *19*(1), 91–102.

108. Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, *36*(1), 49–67.

109. Hedström, P., & Ylikoski, P. (2011). Analytical sociology. Jarvie, I., Zamora-Bonilla, J. (Eds.), *The SAGE Handbook of the Philosophy of Social Sciences* (pp. 386–398). England: SAGE Publications.

110. Hedström, Peter. (2005). *Dissecting the social: On the principles of analytical sociology*. England: Cambridge University Press.

111. Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, *21*(1), 107–112.

112. Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137.

113. Holland, P. W., & Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, *76*(3): 492–513.

114. Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, *2*(2), 107–124.

115. Homans, G. C. (1950). *The human group*. USA: Harcourt, Brace & World.

116. Hu, C., & Racherla, P. (2008). Visual representation of knowledge networks: A social network analysis of hospitality research domain. *International Journal of Hospitality Management*, *27*(2), 302–312.

117. Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2013). Ergm. Userterms: A Template Package for Extending statnet. *Journal of Statistical Software*, *52*(2), i02.

118. Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), nihpa54860.

119. Hyvarinen, A. (2007). Connections between score matching, contrastive divergence, and pseudolikelihood for continuous-valued variables. *IEEE Transactions on Neural Networks*, *18*(5), 1529–1531.

120. Indlekofer, N., & Brandes, U. (2013). Relative importance of effects in stochastic actor-oriented models. *Network Science*, *1*(3), 278–304.

121. Jansson, I. (1997). Clique structure in school class data. *Social Networks*, *19*(3), 285–301.

122. Jansson, I. (2000). Popularity structure in friendship networks. *Social Networks*, *21*(4), 339–359.

123. Johnsen, E. C. (1985). Network macrostructure models for the Davis-Leinhardt set of empirical sociomatrices. *Social Networks*, *7*(3), 203–224.

124. Johnson, J. C., Ironsmith, M., Whitcher, A. L., Poteat, G. M., Snow, C. W., & Mumford, S. (1997). The development of social networks in preschool children. *Early Education and Development*, *8*(4), 389–405.

125. Jonasson, J. (1999). The random triangle model. *Journal of Applied Probability*, *36*(03), 852–867.

126. Jordan, N. (1953). Behavioral Forces That are a Function of Attitudes and of Cognitive Organization. *Human Relations*, *6*(3), 273–287.

127. Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, *84*(2), 427–436.

128. Kejžar, N. (2007). *Evolution of networks*. Slovenia: Faculty of Mathematics and Physics, University of Ljubljana.

129. Kejžar, N., Nikoloski, Z., & Batagelj, V. (2008). Probabilistic inductive classes of graphs. *Journal of Mathematical Sociology*, *32*(2), 85–109.

130. Kerns, K. (2000). Types of preschool friendships. *Personal Relationships*, *7*(3), 311–324.

131. Kirchler, E., & Davis, J. H. (1986). The influence of member status differences and task type on group consensus and member position change. *Journal of Personality and Social Psychology*, *51*(1), 83.

132. Kogut, B. (2000). The network as knowledge: Generative rules and the emergence of structure. *Strategic Management Journal*, *21*(3), 405–425.

133. Koskinen, J., & Snijders, T. (2007). Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference*, *137*(12), 3930–3938.

134. Koskinen, J., & Snijders, T. (2013). Simulation, estimation, and goodness of fit. Lusher, D., Koskinen, J. & Robins, G. (Eds.), *Exponential Random Graph Models for Social Networks* (pp. 141–166). England: Cambridge University Press.

135. Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 29–46.

136. Kronegger, L., Ferligoj, A., & Doreian, P. (2011). On the Dynamics of National Scientific Systems. *Quality & Quantity*, *45*(5), 989–1015.

137. Kronegger, L., Mali, F., Ferligoj, A., & Doreian, P. (2012). Collaboration structures in Slovenian scientific communities. *Scientometrics*, *90*(2), 631–647.

138. Kumpula, J. M., Onnela, J.-P., Saramäki, J., Kaski, K., & Kertész, J. (2007). Emergence of communities in weighted networks. *Physical Review Letters*, *99*(22), 228701.

139. Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., De Luca, E. W., & Albayrak, S. (2010). Spectral analysis of signed graphs for clustering, prediction and visualization. Parthasarathy, S., Liu, B., Goethals, B., Pei, J. & Kamath, C. (Eds.), *Proceedings of the 2010 SIAM International Conference on Data Mining* (pp. 559–570). SIAM.

140. LaFontana, K. M., & Cillessen, A. H. (2002). Children's perceptions of popular and unpopular peers: A multimethod assessment. *Developmental Psychology*, *38*(5), 635.

141. Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, *80*(5), 056117.

142. Latouche, P., Birmelé, E., & Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, *12*(1), 93–115.

143. Lazega, E. (1992). *The micropolitics of knowledge: Communication and indirect control in workgroups*. New York: A. De Gruyter.

144. Lazega, E., Lemercier, C., & Mounier, U. (2006). A Spinning top model of formal organization and informal behavior: Dynamics of advice networks among judges in a commercial court. *European Management Review*, *3*(2), 113–122.

145. Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, *44*(1), 401–424.

146. Leinhardt, S. (1973). The development of transitive structure in children's interpersonal relations. *Behavioral Science*, *18*(4), 260–271.

147. Lewin, K. (1936). *Principles of topological psychology*. England: McGraw-Hill Book Company.

148. Lorrain, F., & White, H. (1971). Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology*, *1*(1), 49–80.

149. Lubbers, M. J. (2003). Group composition and network structure in school classes: A multilevel application of the p∗ model. *Social Networks*, *25*(4), 309–332.

150. Lutter, M. (2013*). Is there a closure penalty? Cohesive network structures, diversity, and gender inequalities in career advancement*. Germany: Max Planck Institute for the Study of Societies.

151. Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3), 151–162.

152. Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in childhood. *Advances in child development and behavior*, *20*(1), 239–287.

153. , D., & Batagelj, V. (in press). Social network analysis as a field of invasions: bibliographic approach to study SNA development. *Scientometrics*. Retrieved from https://doi.org/10.1007/s11192-019-03193-x

154. Manzo, G. (2007). Variables, Mechanisms, and Simulations: Can the Three Methods Be Synthesized ? *Revue Francaise de Sociologie*, *48*(5), 35–71.

155. Marsaglia, G. (1972). Choosing a point from the surface of a sphere. *The Annals of Mathematical Statistics*, *43*(2), 645–646.

156. Matias, C., & Miele, V. (2015). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of Royal Statistical Society*, *79*(4), 1119–1141.

157. McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444.

158. Mehra, A., Kilduff, M., & Brass, D. J. (2001). The social networks of high and low self-monitors: Implications for workplace performance. *Administrative Science Quarterly*, *46*(1), 121–146.

159. Merton, R. K. (1949). On sociological theories of the middle range. Calhoun, C., Gerteis, J., Moody, J., Pfaff, S. & Virk, I. (Eds.), *Classical sociological theory* (pp. 448–459). Malden: Blackwell.

160. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

161. Michaelson, A., & Contractor, N. S. (1992). Structural position and perceived similarity. *Social Psychology Quarterly*, *55*(3), 300–310.

162. Miklós, I., & Podani, J. (2004). Randomization of Presence–Absence Matrices: Comments and New Algorithms. *Ecology*, *85*(1), 86–92.

163. Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., … Alon, U. (2004). Superfamilies of evolved and designed networks. *Science*, *303*(5663), 1538–1542.

164. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824–827.

165. Mizruchi, M. S. (1993). Cohesion, equivalence, and similarity of behavior: A theoretical and empirical assessment. *Social Networks*, *15*(3), 275–307.

166. Moreno, J. L. (1934). *Who shall survive: Foundations of Sociometry Group Psychotherapy and Sociodrama*. New York: Beacon House Inc.

167. Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software*, *24*(4), 1548–7660.

168. Muller, M. E. (1959). A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM*, *2*(4), 19–20.

169. Nadel, S. F. (1957). *The Theory of Social Structure*. England: Routledge.

170. Nahapiet, J., & Ghoshal, S. (2000). Social capital, intellectual capital, and the organizational advantage. Lesser, L. (Ed.), *Knowledge and social capital* (pp. 119–157), England: Elsevier.

171. Nebus, J. (2006). Building collegial information networks: A theory of advice network generation. *Academy of Management Review*, *31*(3), 615–637.

172. Nordlund, C. (2018). Power-relational core–periphery structures: Peripheral dependency and core dominance in binary and valued networks. *Network Science*, *6*(3), 348–369.

173. Oberg, A., & Walgenbach, P. (2008). Hierarchical structures of communication in a network organization. *Scandinavian Journal of Management*, *24*(3), 183–198.

174. Olweus, D. (1994). Bullying at School. Huesmann, L. R. (Ed.), *Aggressive Behavior: Current Perspectives* (pp. 97–130), USA: Springer.

175. Opp, K.-D. (2007). Peter Hedström: Dissecting the Social. On the Principles of Analytical Sociology. *European Sociological Review*, *23*(1), 115–122.

176. Pahor, M., Škerlavaj, M., & Dimovski, V. (2008). Evidence for the network perspective on organizational learning. *Journal of the American Society for Information Science and Technology*, *59*(12), 1985–1994.

177. Pattison, P., & Robins, G. (2002). Neighborhood–based models for social networks. *Sociological Methodology*, *32*(1), 301–337.

178. Peixoto, T. (2015). Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups. *Physical Review X*, *5*(1): 011033.

179. Peixoto, T. (in press). Bayesian stohastic blockmodeling. Doreian, P., Batagelj, V., Ferligoj, A. (Eds.) *Advances in network clustering and blockmodeling* (pp. 281–324). USA: Wiley-Interscience.

180. Prota, L., & Doreian, P. (2016). Finding Roles in Sparse Economic Hierarchies: Going beyond Regular Equivalence. *Social Networks*, *45*(1), 1–17.

181. Prota, L., D'Esposito, M. R., De Stefano, D., Giordano, G., & Vitale, M. P. (2013). *Modeling Cooperative Behaviors in Innovation Networks: An Empirical Analysis*. Italy: Taylor & Francis Group.

182. Proulx, M.-F., & Poulin, F. (2013). Stability and change in kindergartners' friendships: Examination of links with social functioning. *Social Development*, *22*(1), 111–125.

183. Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850.

184. Rao, A. R., Jana, R., & Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyā: The Indian Journal of Statistics, Series A*, *58*(2), 225–242.

185. Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, *48*(2), 240–267.

186. Reagans, R., Zuckerman, E., & McEvily, B. (2004). How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly*, *49*(1), 101–133.

187. Ripley, R., Snijders, T., Boda, Z., Vörös, A. & Preciado, P. (2019). *Manual for RSiena*. Retrieved from http://www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf

188. Robins, G. (2011). Exponential random graph models for social networks. Scott, J. & Carrington, P. (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 484–500). USA: Sage.

189. Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191.

190. Robins, G., Pattison, P., & Wang, P. (2009). Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. *Social Networks*, *31*(2), 105–117.

191. Rombach, P., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2017). Core-Periphery Structure in Networks (Revisited). *SIAM Review*, *59*(3), 619–646.

192. Rook, K. S. (1984). The negative side of social interaction: Impact on psychological well-being. *Journal of Personality and Social Psychology*, *46*(5), 1097–1108.

193. Rubineau, B., Lim, Y., & Neblo, M. (2019). Low status rejection: How status hierarchies influence negative tie formation. *Social Networks*, *56*(1), 33–44.

194. Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *22*(1), 1–15.

195. Sanderson, J. G., Moulton, M. P., & Selfridge, R. G. (1998). Null matrices and the analysis of species co-occurrences. *Oecologia*, *116*(1), 275–283.

196. Schaefer, D., Adams, J., & Haas, S. (2013). Social Networks and Smoking: Exploring the Effects of Peer Influence and Smoker Popularity Through Simulations. *Health Education & Behavior*, *40*(1), 24–32.

197. Schaefer, D., Light, J., Fabes, R., Hanish, L., & Martin, C. L. (2010). Fundamental principles of network formation among preschool children. *Social Networks*, *32*(1), 61–71.

198. Schaefer, D. R., Light, J. M., Fabes, R. A., Hanish, L. D., & Martin, C. L. (2010). Fundamental principles of network formation among preschool children. *Social Networks*, *32*(1), 61–71.

199. Schelling, T. (1998). Social mechanisms and social dynamics. Elster, J. & McPherson, S. (Eds.) *Social mechanisms: An analytical approach to social theory* (pp. 32–44). England: Cambridge University Press.

200. Shen, K., Hutchison, R. M., Bezgin, G., Everling, S., & McIntosh, A. R. (2015). Network structure shapes spontaneous functional connectivity dynamics. *The Journal of Neuroscience*, *35*(14), 5579–5588.

201. Shrum, W., & Cheek, N. H. (1987). Social structure during the school years: Onset of the degrouping process. *American Sociological Review*, *52*(2), 218–223.

202. Škerlavaj, M. (2007). *The network perspective and performance of organizational learning: Theoretical and empirical analysis*. Slovenia: University of Ljubljana.

203. Škerlavaj, M., Dimovski, V., & Desouza, K. C. (2010). Patterns and structures of intra-organizational learning networks within a knowledge-intensive organization. *Journal of Information Technology*, *25*(2), 189–204.

204. Škerlavaj, M., Dimovski, V., Mrvar, A., & Pahor, M. (2010). Intra-organizational learning networks within knowledge-intensive learning environments. *Interactive Learning Environments*, *18*(1), 39–63.

205. Snijders, T. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, *21*(1–2), 149–172.

206. Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*(1), 361–395.

207. Snijders, T. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, *3*(2), 1–40.

208. Snijders, T. (2004). Explained variation in dynamic network models. *Mathématiques et Sciences Humaines*, *168*(4), 5–15.

209. Snijders, T. (2011). Network dynamics. Scott, J. & Carrington, P. (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 501–513). USA: Sage.

210. Snijders, T., Koskinen, J., & Schweinberger, M. (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, *4*(2), 567.

211. Snijders, T., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, *14*(1), 75–100.

212. Snijders, T., Pattison, P. E., & Handcock, M. (2006). New Specifications for Exponential Random Graph Models. *Sociological Methodology*, *36*(1), 99–153.

213. Snijders, T., & Steglich, C. (2015). Representing micro–macro linkages by actor-based dynamic network models. *Sociological Methods & Research*, *44*(2), 222–271.

214. Snijders, T., Van de Bunt, G., & Steglich, C. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*(1), 44–60.

215. Snyder, J., West, L., Stockemer, V., Gibbons, S., & Almquist-Parks, L. (1996). A social learning model of peer choice in the natural environment. *Journal of Applied Developmental Psychology*, *17*(2), 215–237.

216. Sohn, Y., Choi, J.-K., & Ahn, T. K. (2019). Core-periphery segregation in evolving prisoner's dilemma networks. *Complex Networks, cnz021, 1–9*.

217. Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal*, *44*(2), 316–325.

218. Stadtfeld, C. (2018). *The Micro-Macro Link in Social Networks*. Retrieved from: https://papers.ssrn.com/abstract=3211795

219. Stadtfeld, C., Takács, K., & Vörös, A. (2018). *The Emergence and Stability of Groups in Social Networks*. Retrieved from: https://ssrn.com/abstract=3232958

220. Steglich, C., Zander, L., Henke, T., & Baumert, J. (2019). Holding Micro-Level Network Mechanisms Accountable for Observed Macro-Level Group Segregation. The Case of Language-Delineated Student Groups in Bilingual Education. Conaldi, G. & Palloti, F. (Eds.), *4th European Conference on Social Networks* (p. 275). Switzerland: ETH Zürich.

221. Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, *85*(409), 204–212.

222. Strona, G., Nappo, D., Boccacci, F., Fattorini, S., & San-Miguel-Ayanz, J. (2014). A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*, *5*(1), 4114.

223. Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, *17*(S2), 27–43.

224. Tang, F., Xi, Y., & Ma, J. (2006). Estimating the effect of organizational structure on knowledge transfer: A neural network approach. *Expert Systems with Applications*, *30*(4), 796–800.

225. Team, R. C. (2000). *R language definition*. Austria: R Foundation for Statistical Computing.

226. Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J.-P., Saramäki, J., & Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks*, *31*(4), 240–254.

227. Van der Leij, M., In't Veld, D., & Hommes, C. H. (2016). *The formation of a core periphery structure in heterogeneous financial networks.* Netherlands: Universiteit van Amsterdam, Center for Nonlinear Dynamics in Economics and Finance.

228. Van der Ploeg, R., Steglich, C., & Veenstra, R. (in press). The way bullying works: How new ties facilitate the mutual reinforcement of status and bullying in elementary schools. *Social Networks*, in press. Retrieved from: https://doi.org/10.1016/j.socnet.2018.12.006

229. Van Duijn, M. A., Gile, K. J., & Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, *31*(1), 52–62.

230. Vaughn, B. E., Colvin, T. N., Azria, M. R., Caya, L., & Krzysik, L. (2001). Dyadic analyses of friendship in a sample of preschool-age children attending Head Start: Correspondence between measures and implications for social competence. *Child Development*, *72*(3), 862–878.

231. Wang, S., Gong, M., Du, H., Ma, L., Miao, Q., Du, W. (2016). Optimizing dynamical changes of structural balance in signed network based on memetic algorithm. *Social Networks*, *44*(1), 64–73.

232. Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236–244.

233. Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236–244.

234. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. England: Cambridge University Press.

235. Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, *61*(3), 401–425.

236. White, D. (1985). *Doug White's Regular Equivalence Program*. Retrieved from http://eclectic.ss.uci.edu/~drwhite/REGGE/REGGE.FOR

237. White, D., & Reitz, K. (1983). Graph and semigroup homomorphisms on networks of relations. *Social Networks*, *5*(2), 193-234.

238. Wittenbaum, G. M. (1998). Information sampling in decision-making groups: The impact of members' task-relevant status. *Small Group Research*, *29*(1), 57–84.

239. Wong, S.-S. (2008). Task knowledge overlap and knowledge variety: The role of advice network structures and impact on group effectiveness. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *29*(5), 591–614.

240. Xing, E. P., Fu, W., Song, L., & others. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, *4*(2), 535–566.

241. Xu, K. S., & Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, *8*(4), 552–562.

242. Yang, B., Liu, X., Li, Y., & Zhao, X. (2017). Stochastic Blockmodeling and Variational Bayes Learning for Signed Network Analysis. *IEEE Transactions on Knowledge and Data Engineering*, *29*(9), 2026–2039.

243. Yang, Bo, Cheung, W., & Liu, J. (2007). Community mining from signed social networks. *IEEE Transactions on Knowledge and Data Engineering*, *19*(10), 1333–1348.

244. Ylikoski, P. (2012). Micro, macro, and mechanisms. Kincaid, H. (Ed.), *The Oxford Handbook of Philosophy of the Social Sciences* (pp. 21–45). England: Oxford University Press.

245. Yu, H., & Gerstein, M. (2006). Genomic analysis of the hierarchical structure of regulatory networks. *Proceedings of the National Academy of Sciences*, *103*(40), 14724–14731.

246. Zagenczyk, T. J., & Murrell, A. J. (2009). It is better to receive than to give: Advice network effects on job and work-unit attachment. *Journal of Business and Psychology*, *24*(2), 139–152.

247. Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review*, *27*(2), 185–203.

248. Zaman, A., & Simberloff, D. (2002). Random binary matrices in biogeographical ecology—Instituting a good neighbor policy. *Environmental and Ecological Statistics*, *9*(4), 405–421.

249. Zeggelink, E. (1994). Dynamics of structure: An individual oriented approach∗. *Social Networks*, *16*(4), 295–333.

250. Žiberna, A. (2007). Generalized blockmodeling of valued networks. *Social Networks*, *29*(1), 105–126.

251. Žiberna, A. (2008). Direct and indirect approaches to blockmodeling of valued networks in terms of regular equivalence. *Journal of Mathematical Sociology*, *32*(1), 57–84.

252. Žiberna, A. (2013). Generalized blockmodeling of sparse networks. *Metodološki Zvezki*, *10*(2), 99–119.

253. Žiberna, A. (2018). *Generalized and Classical Blockmodeling of Valued Networks*. Retrieved from https://cran.r-project.org/web/packages/blockmodeling/blockmodeling.pdf

254. Žiberna, A., & Lazega, E. (2016). Role Sets and Division of Work at Two Levels of Collective Agency: The Case of Blockmodeling a Multilevel (Inter-Individual and Inter-Organizational) Network. Lazega, E. & Snijders, T. (Eds.), *Multilevel Network Analysis for the Social Sciences* (pp. 173–209). Springer International Publishing.

255. Žnidaršič, A. (2012). *Stability of blockmodeling*. Slovenia: University of Ljubljana.

256. Žnidaršič, A., Ferligoj, A., & Doreian, P. (2012). Non-response in social networks: The impact of different non-response treatments on the stability of blockmodels. *Social Networks*, *34*(4), 438–450.

257. Žnidaršič, A., Ferligoj, A., & Doreian, P. (2017). Actor non-response in valued social networks: The impact of different non-response treatments on the stability of blockmodels. *Social Networks*, *48*(1), 46–56.

258. Žnidaršič, A., Ferligoj, A., & Doreian, P. (2018). Stability of centrality measures in valued networks regarding different actor non-response treatments and macro-network structures. *Network Science*, *6*(1), 1–33.

10 Subject and author index

C

D

E

J

K

L

M

N

O

P

R

S

T

V

W

Y

Z

Appendix A: Mean RF for networks generated by considering different θs and different initial networks

Figure 1: Mean relative fit values for different blockmodel types generated by different θs (the initial network is empty)

Figure 2: Mean relative fit values for different blockmodel types generated by different θs (initial is an asymmetric core-periphery blockmodel)

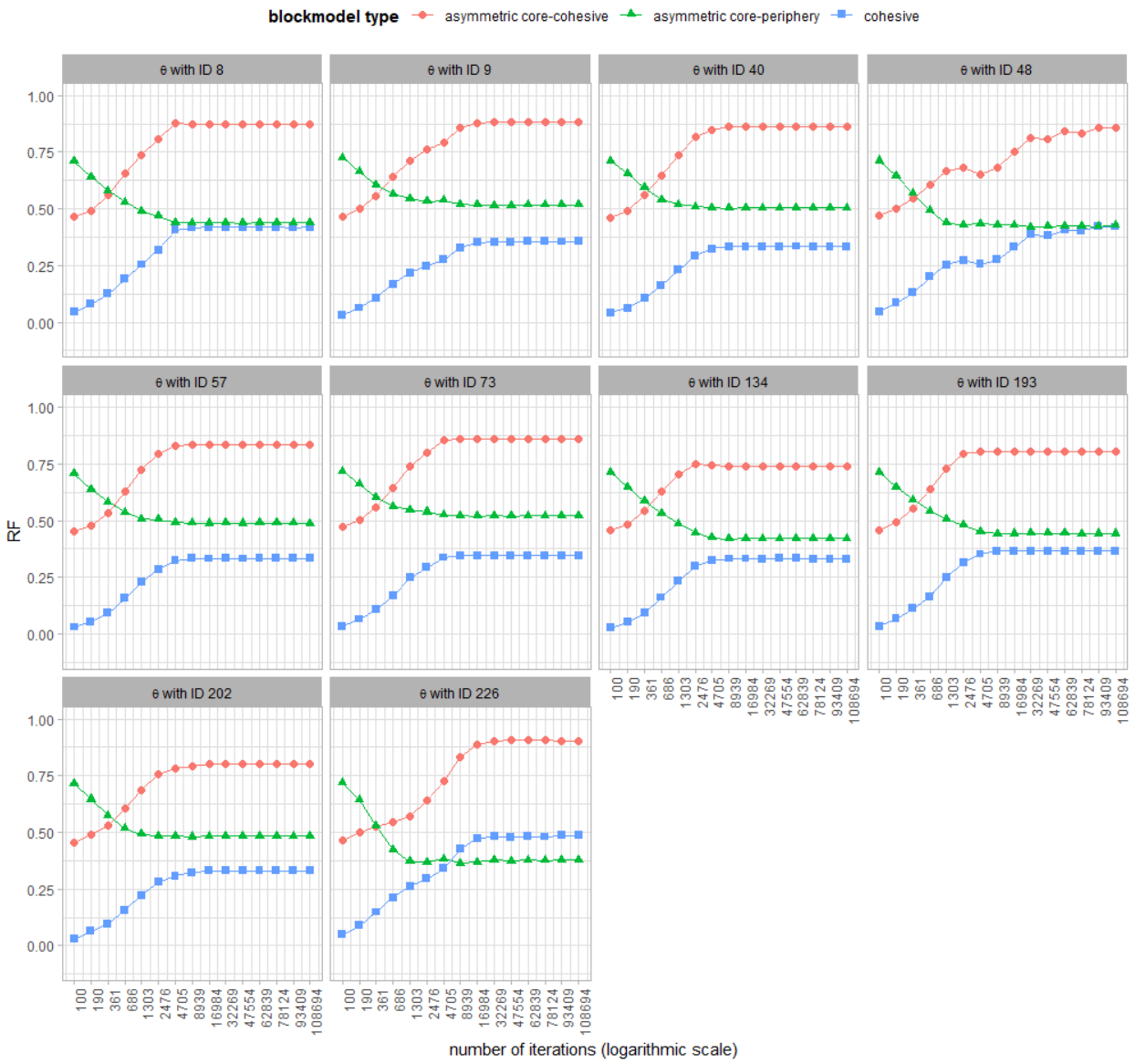


Figure 3: Mean relative fit values for different blockmodel types generated by different θs (initial is a cohesive blockmodel)

Appendix B: The distributions of improvement values

Figure 1: The distribution of the improvement values for the networks generated by different triad types by the RL algorithm

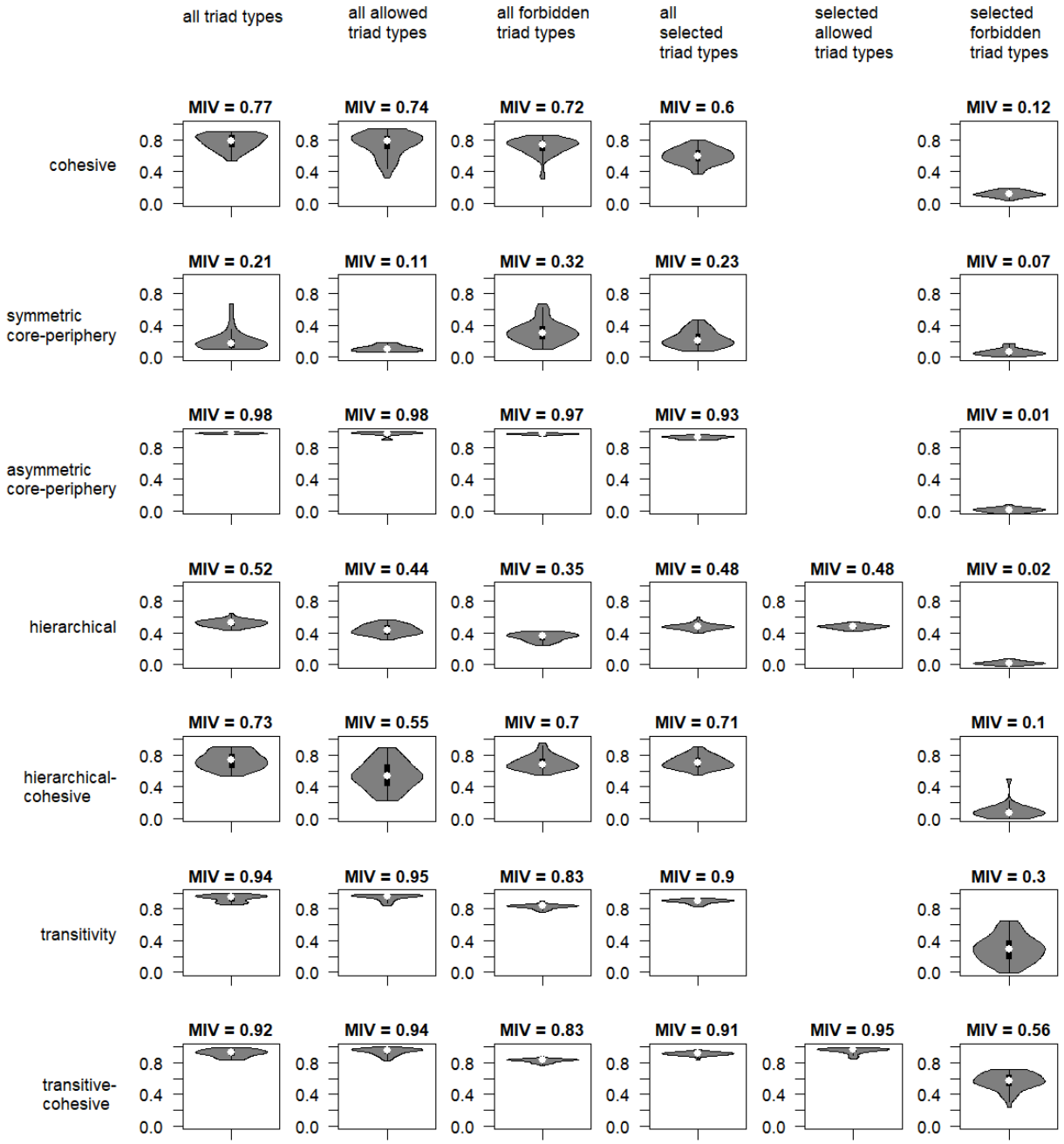


Figure 2: The distribution of the improvement values for the networks generated by different triad types by the MCMC algorithm (fixed density)

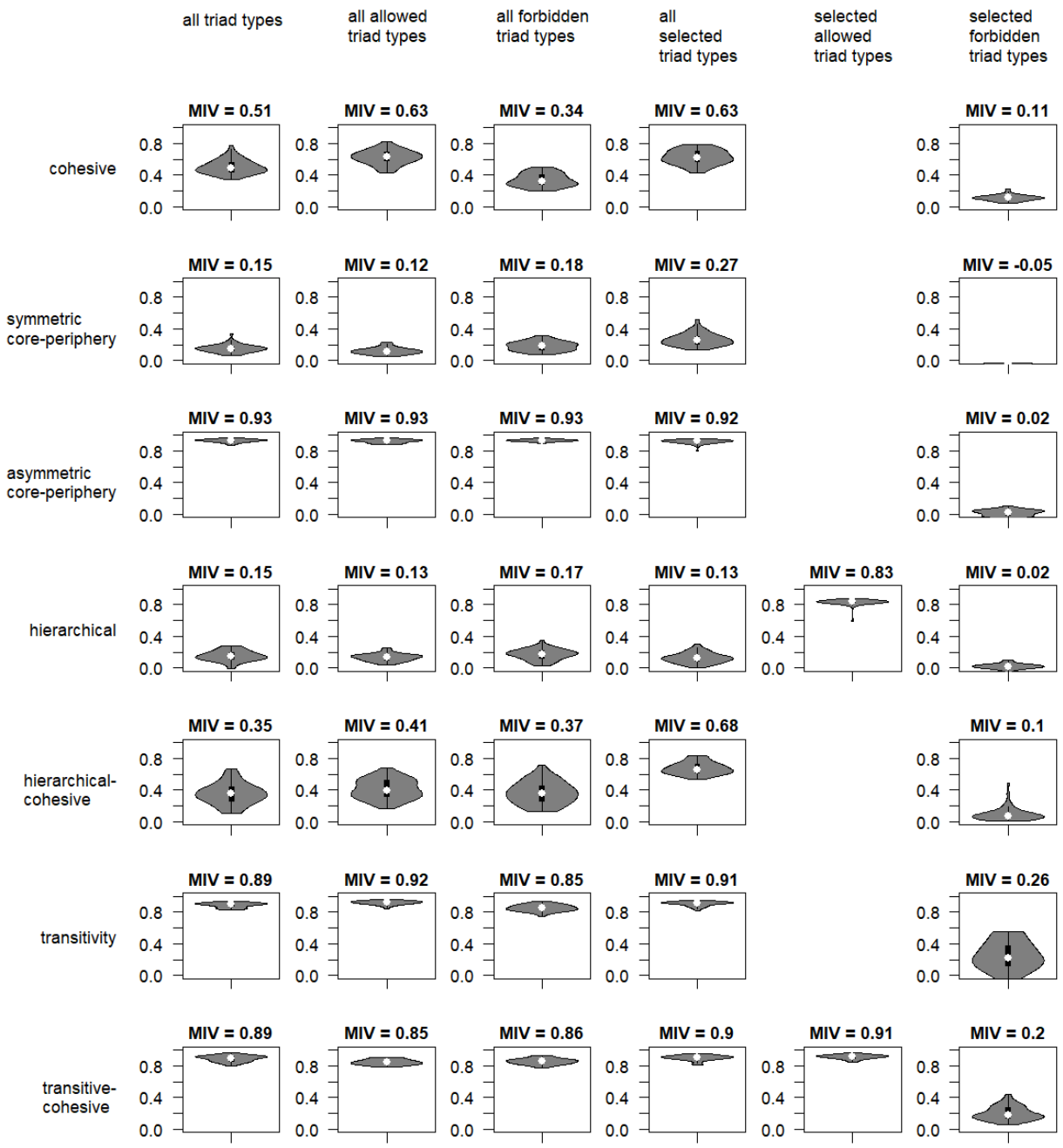
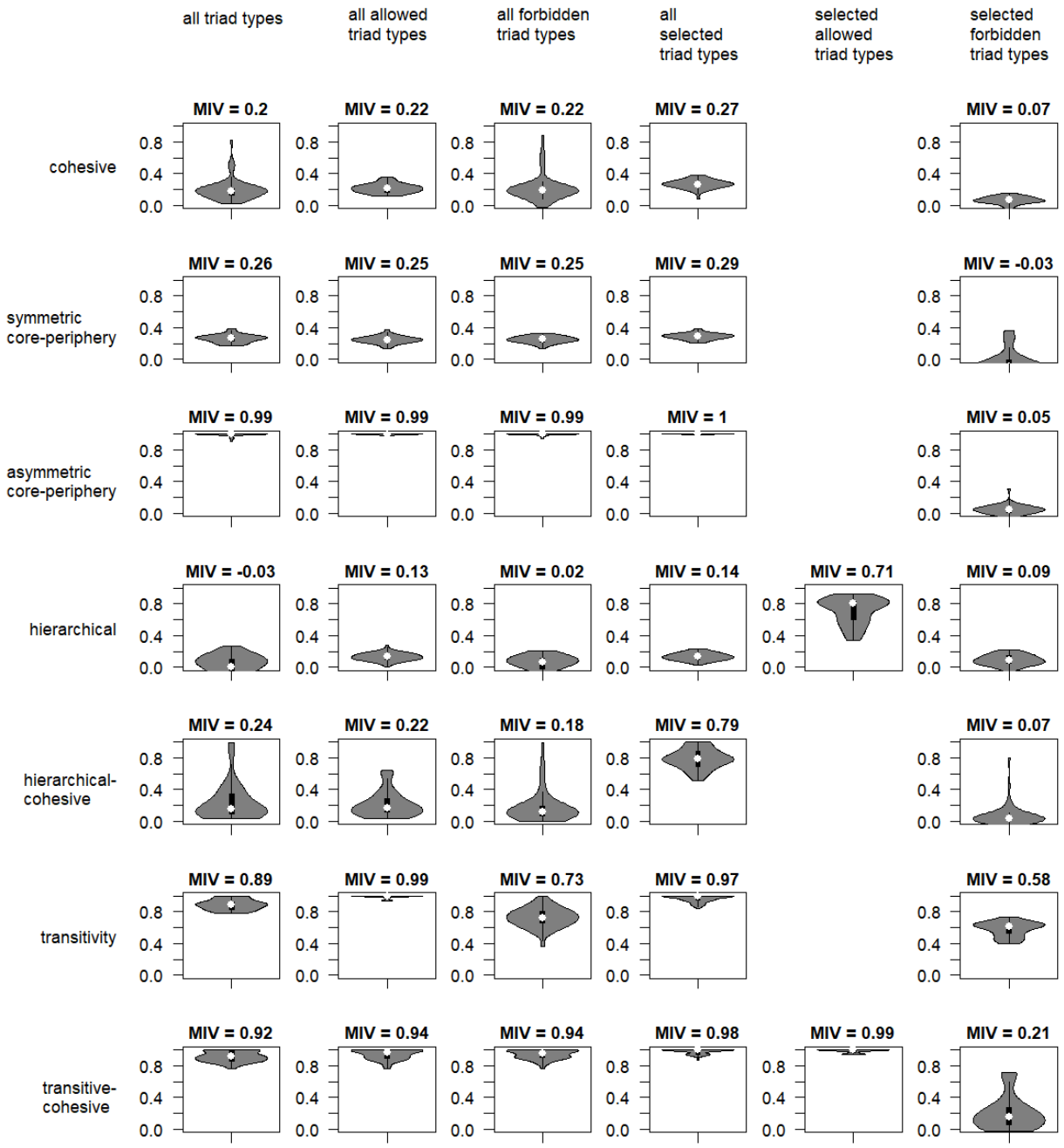


Figure 3: The distribution of the improvement values for the networks generated by different triad types by the MCMC algorithm (non-fixed density)



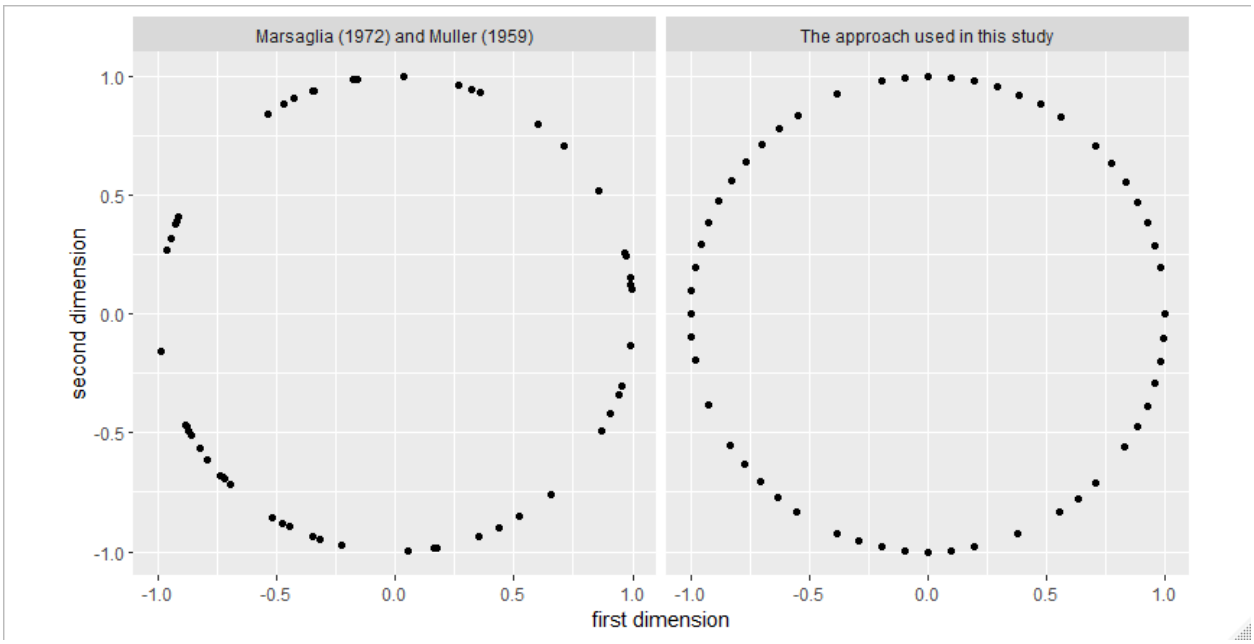
Appendix C: Generating equally distributed $\boldsymbol{\theta}$s

θ is a vector of length corresponding to the number of local network mechanisms that are being considered. To limit the space of all possible values, the restriction $\sum {\theta_i}^2 = 1$ is applied. Individual θs can be seen as points in the space. When each θ consists of three values, the points are distributed across the sphere. There are several approaches to generating random points on the sphere that can be generalized on a higher dimension. One approach is by generating the values from the standard normal distribution Φ and multiplying them by a scalar, $\theta = \Phi\left(1/\sqrt{\sum \Phi_i}\right)$ (Marsaglia, 1972; Muller, 1959). The values (points) are randomly distributed when this approach is applied, although some are much closer to each other than others. To solve this issue, several approaches have been proposed but they are all limited to the three-dimensional space.

Therefore, the following approach to generating random points in n-dimensional space is used. When this approach is used, the points are relatively equally distributed. The algorithm is as follows. A researcher must set the desired number of θs (points) to be generated as well as the initial set of θs, I. Thee iterative process is then initiated. In this process, the approach put forward by Marsaglia (1972) and Muller (1959) is used to generate a set of θs, called C, which are the candidates for being added to the initial set I. The number of elements of set C is arbitrarily, but the highest number results in more equally distributed points. Then, the θ from set C with maximal minimal distance to any θ from set I is identified and added to set S.

As shown in Figure 1: An example of some $\boldsymbol{\theta}$s generated by two different approaches, the points are more equally distributed with the described approach than with the approach proposed by Marsaglia (1972) and Muller (1959) (54 θs are generated by each approach, each set C consists of 1,000 θs).

Figure 1: An example of some θs generated by two different approaches



Appendix D: Number of inconsistent blocks and mean RF values for the generated networks with a hierarchical-cohesive blockmodel with last non-cohesive group or with a hierarchical-cohesive blockmodel

Table 1: The selected θs for generating networks with a hierarchical-cohesive blockmodel with last non-cohesive group

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter		
34	Fixed to 0.	0.943	Fixed to 0.		-0.334	Fixed to 0.		0.31	0.004
295		0.944			-0.331			0.31	0.003
298		0.941			-0.337			0.31	0.004
535		0.942			-0.335			0.31	0.004
623		0.941			-0.339			0.31	0.004
739		0.943			-0.332			0.31	0.003
1069		0.942			-0.335			0.31	0.004
1242		0.941			-0.338			0.31	0.003
1265		0.943			-0.333			0.31	0.004
1348		0.942			-0.336			0.31	0.005
1642		0.944			-0.330			0.31	0.004
1658		0.943			-0.332			0.31	0.004

Note: Only newcomers are possible. The signs of the mechanisms' weights are fixed.

Table 2: The selected θs for generating networks with a hierarchical-cohesive blockmodel with last non-cohesive group

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter		
1868	0.455	Fixed to 0.	0.030	0.020	-0.740	Fixed to 0.	-0.494	0.29	0.019
532	0.419		0.057	0.050	-0.660		-0.619	0.28	0.020
368	0.640		0.011	0.134	-0.592		-0.471	0.27	0.016
962	0.307		0.170	0.094	-0.616		-0.698	0.26	0.027
374	0.351		0.008	0.011	-0.526		-0.775	0.25	0.016

Note: Only newcomers are possible. The signs of the mechanisms' weights are fixed.

Table 3: The selected θs for generating networks with a hierarchical-cohesive blockmodel with last non-cohesive group

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter		
798	0.308	0.703	0.104	-0.023	-0.621	-0.003	-0.121	0.50	0.029
1861	0.281	0.380	0.123	-0.016	-0.737	-0.005	-0.468	0.50	0.024
1222	0.396	0.557	0.075	0.095	-0.636	-0.015	-0.337	0.46	0.030
1656	0.339	0.609	0.001	-0.157	-0.600	-0.117	-0.340	0.43	0.029
1814	0.041	0.732	0.082	-0.011	-0.612	-0.060	-0.280	0.43	0.044
147	0.428	0.377	0.002	-0.121	-0.756	-0.026	-0.297	0.42	0.022
1301	0.039	0.590	0.004	-0.096	-0.483	-0.103	-0.630	0.42	0.023
1757	0.412	0.301	0.197	-0.038	-0.836	-0.008	-0.035	0.42	0.017
446	0.137	0.546	0.178	-0.118	-0.675	-0.027	-0.425	0.40	0.032
777	0.136	0.542	0.236	0.164	-0.518	-0.041	-0.579	0.40	0.022
216	0.009	0.907	0.019	-0.027	-0.410	-0.061	-0.073	0.39	0.028
621	0.376	0.360	0.179	0.147	-0.597	-0.173	-0.537	0.39	0.040
802	0.295	0.688	0.135	-0.092	-0.547	-0.337	-0.030	0.39	0.041
821	0.051	0.705	0.306	-0.029	-0.613	-0.149	-0.092	0.39	0.029
483	0.001	0.794	0.049	0.085	-0.413	-0.153	-0.406	0.38	0.029
1835	0.360	0.521	0.124	-0.145	-0.564	-0.340	-0.359	0.38	0.024
1910	0.376	0.182	0.047	0.032	-0.827	-0.002	-0.371	0.38	0.032
1377	0.270	0.073	0.006	-0.048	-0.761	-0.024	-0.583	0.37	0.032
1982	0.246	0.526	0.401	0.217	-0.631	-0.135	-0.197	0.33	0.032
1956	0.189	0.384	0.178	-0.095	-0.852	-0.032	-0.222	0.31	0.035

Note: Only newcomers are possible. The signs of the mechanisms' weights are fixed.

Table 4: The selected θs for generating networks with selected versions of a hierarchical blockmodel (only newcomers are possible, the signs of the mechanism' weights are not fixed)

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF	Number of generated networks with the
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter			
Hierarchical-cohesive blockmodel with last non-cohesive group										
501	0.367	0.331	0.202	0.158	-0.622	-0.195	-0.515	0.33	0.038	29
Hierarchical-cohesive blockmodel										
1632	0.410	0.668	-0.227	-0.163	-0.490	-0.185	0.180	0.25	0.023	29

Table 5: The selected θs for generating networks with selected versions of a hierarchical blockmodel (newcomers and outgoers are possible, the signs of the mechanism' weights are not fixed)

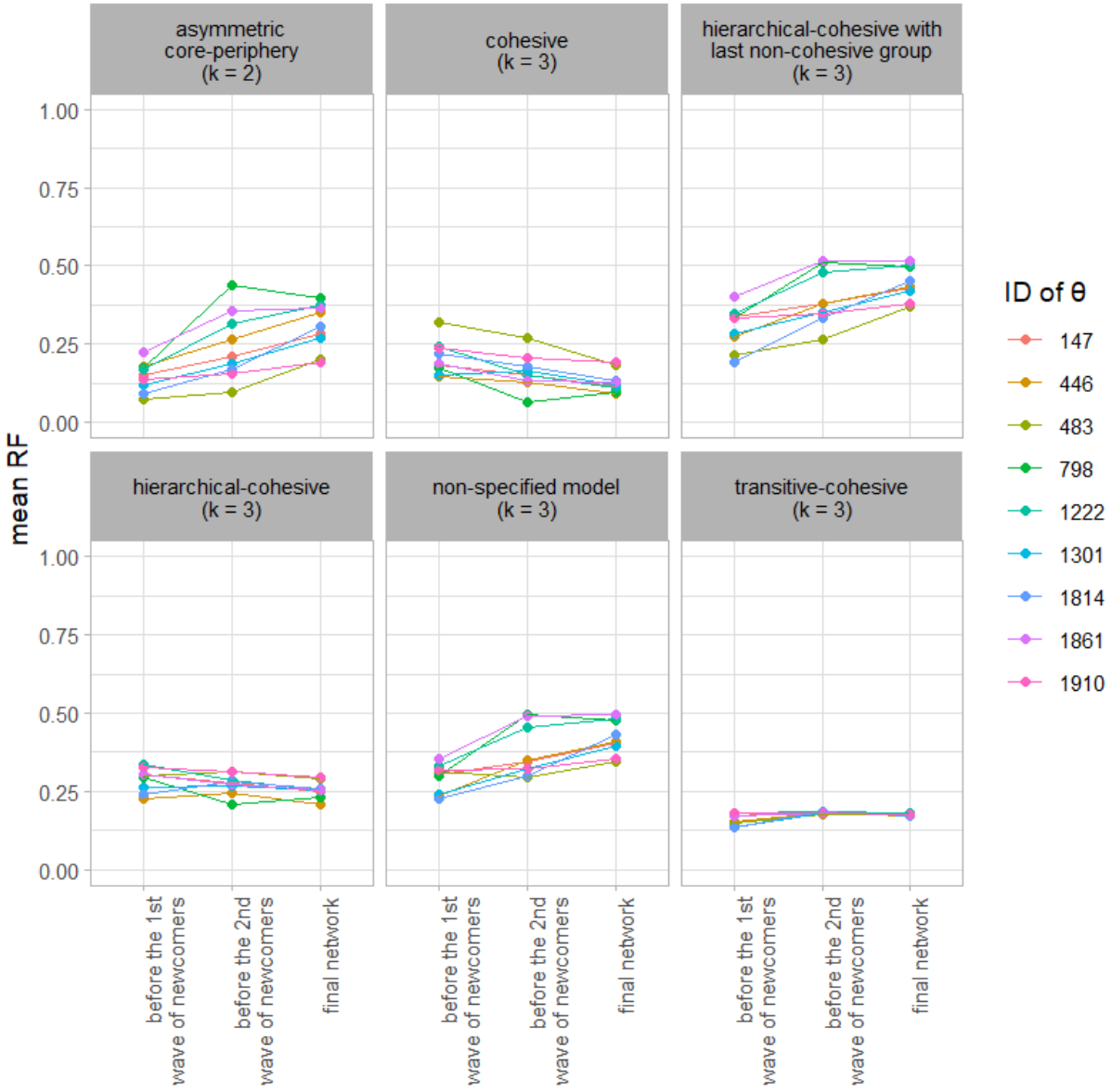
ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF	Number of generated networks with the
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter			
Hierarchical-cohesive blockmodel with last non-cohesive group										
880	0.564	0.130	0.008	-0.283	-0.743	-0.044	-0.179	0.25	0.040	29
Hierarchical-cohesive blockmodel										
1456	-0.033	0.848	-0.378	0.064	0.121	-0.338	0.067	0.17	0.041	26

Table 6: The selected θs for generating networks with a hierarchical-cohesive blockmodel

ID	the θs that generated 30 (out of 30) networks without any inconsistent blocks							mean RF	sd of RF	Number of generated networks with the
	Hierarchical position of the alter	Tenure of the alter	Popularity level of the alter	Outgoing shared partners	Difference in hierarchical position between the ego and the alter	Difference in tenure between the ego and the alter	Distance between the ego and the alter			
35	0.035	0.635	0.025	-0.038	-0.740	-0.186	-0.100	0.21	0.023	28
562	0.022	0.729	0.075	-0.282	-0.537	-0.290	-0.097	0.21	0.027	28

Note: Only newcomers are possible. The signs of the mechanisms' weights are fixed.

Figure 1: The mean RF values for different blockmodel types for selected θs



Appendix E: Interpreting the RF for sparse networks

The amount of errors (inconsistencies) between different networks with the same blockmodel is evaluated by using the relative fit measure (RF), defined as

$$RF = 1 - \frac{P^m}{\frac{1}{k}\sum_{i=1}^{k} P_i^r}$$

where k is the number of randomized networks, P^m is the value of a criterion function of the network of the interest (e.g. empirical) and P_i^r is the value of a criterion function of the i-th random network. The criterion function for structural equivalence (Doreian et al., 2005) is defined for nondiagonal blocks as

$$\delta(R, B) = \sum_{x \in C_u,\ y \in C_v} \left| r_{xy} - b_{xy} \right|$$

where R corresponds to the observed nondiagonal block and B corresponds to the ideal block. Next, r_{xy} is the observed tie and b_{xy} is the corresponding value in the ideal block. In line with the generalized blockmodeling approach for sparse networks (Žiberna, 2013), the errors in null and complete blocks can be given different weights. In this study, complete blocks are weighted by $d/(1-d)$ and null blocks are weighted by 1, where d is the density of the whole network. When there are no inconsistencies (i.e. no links in null blocks and all links in complete blocks), the value of one criterion function or another would be equal to 1. Yet the number of outgoing links is constrained by the algorithm for generating networks (which is an operationalization of the restricted ability to maintain an infinite number of relationships) to a number which is lower to what is implied by the chosen blockmodel type with three clusters. This means it is theoretically impossible to obtain the value of a criterion function equal to 1.

The simulation study was conducted to estimate the expected and maximum values of the criterion function used in the blockmodeling approach for sparse networks. For this purpose, 10,000 random networks were generated in line with the chosen blockmodel such that there were no links in null blocks and the out-degree of each node was 5. This is very close to the best theoretical fit of the global network structure to the chosen blockmodel that could be achieved by considering the

constraints of the out-degree. The blockmodeling approach for sparse networks was applied to each generated network and the value of the criterion function was calculated.

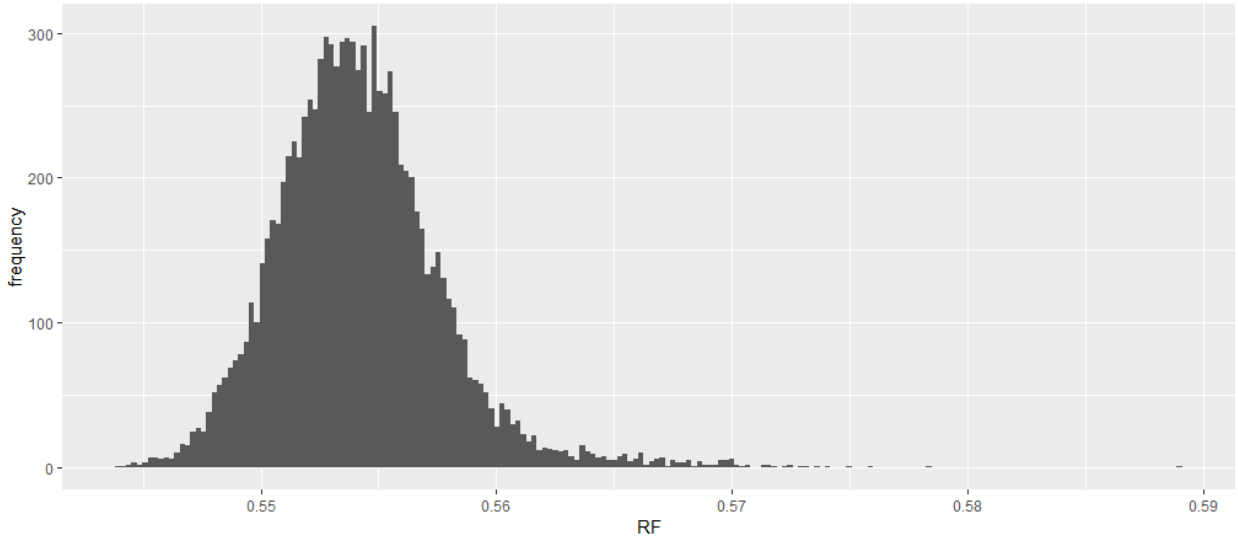
In addition, RF values were calculated for all networks that were generated. The mean value of the criterion function for the case of random networks was estimated using simulations in which 30 random networks (with the same density as in ideal networks) were generated for each of 10,000 generated random networks with the chosen blockmodel.

Table 1: Some summary statistics of the distribution of the relative fit for the case of randomized networks for each blockmodel type

	min	mean	max
Hierarchical-cohesive blockmodel with last non-cohesive group	0.53	0.55	0.59
Hierarchical-cohesive blockmodel	0.32	0.34	0.36

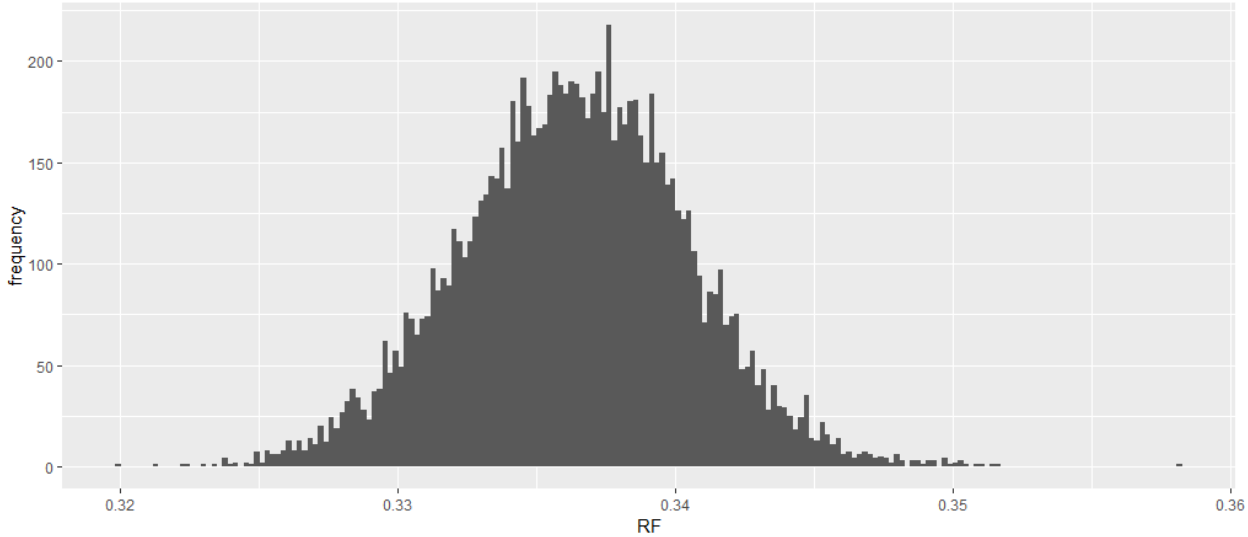
The maximum observed RF value for the case of hierarchical-cohesive blockmodel with last non-cohesive group is 0.59 (see Table 1 and Figure 1) whereas the maximum RF value for the hierarchical-cohesive blockmodel is 0.36 (see Table 1 and Figure 2), which is significantly lower. This is because five links (corresponding to the nodes on the lowest hierarchical level) in total are distributed between the nodes from two, instead of one, cluster in the case of the second blockmodel. This increases the number of inconsistencies in complete blocks. RF values close to the maximum observed RF values in random networks for the corresponding blockmodel can be interpreted as very high.

Figure 1: Distribution of the relative fit values for a hierarchical-cohesive blockmodel with last non-cohesive group



Note: The RF values are calculated based on 10,000 randomly generated networks in line with the chosen blockmodel.

Figure 2: Distribution of the relative fit values for a hierarchical-cohesive blockmodel



Note: The RF values are calculated based on 10,000 randomly generated networks in line with the chosen blockmodel.

Appendix F: Characteristics of the RF

The characteristics of the RF are studied by conducting simulations. More precisely, the relationship between the RF obtained by pre-specified blockmodeling ($\mathrm{RF_P}$) and the RF obtained by non-specified blockmodeling ($\mathrm{RF_N}$) is studied for different levels of errors and different numbers of clusters.

Networks with different blockmodel types (cohesive, hierarchical, hierarchical-cohesive, transitivity and transitive-cohesive), with different levels of errors, $LE = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, and with different numbers of clusters, $k = \{3, 4, 6, 8\}$, are generated. For each combination of blockmodel type, level of errors and number of clusters, 30 networks (each consisting of 24 nodes) are generated. In total, $5 * 11 * 4 * 30 = 6600$ networks are generated.

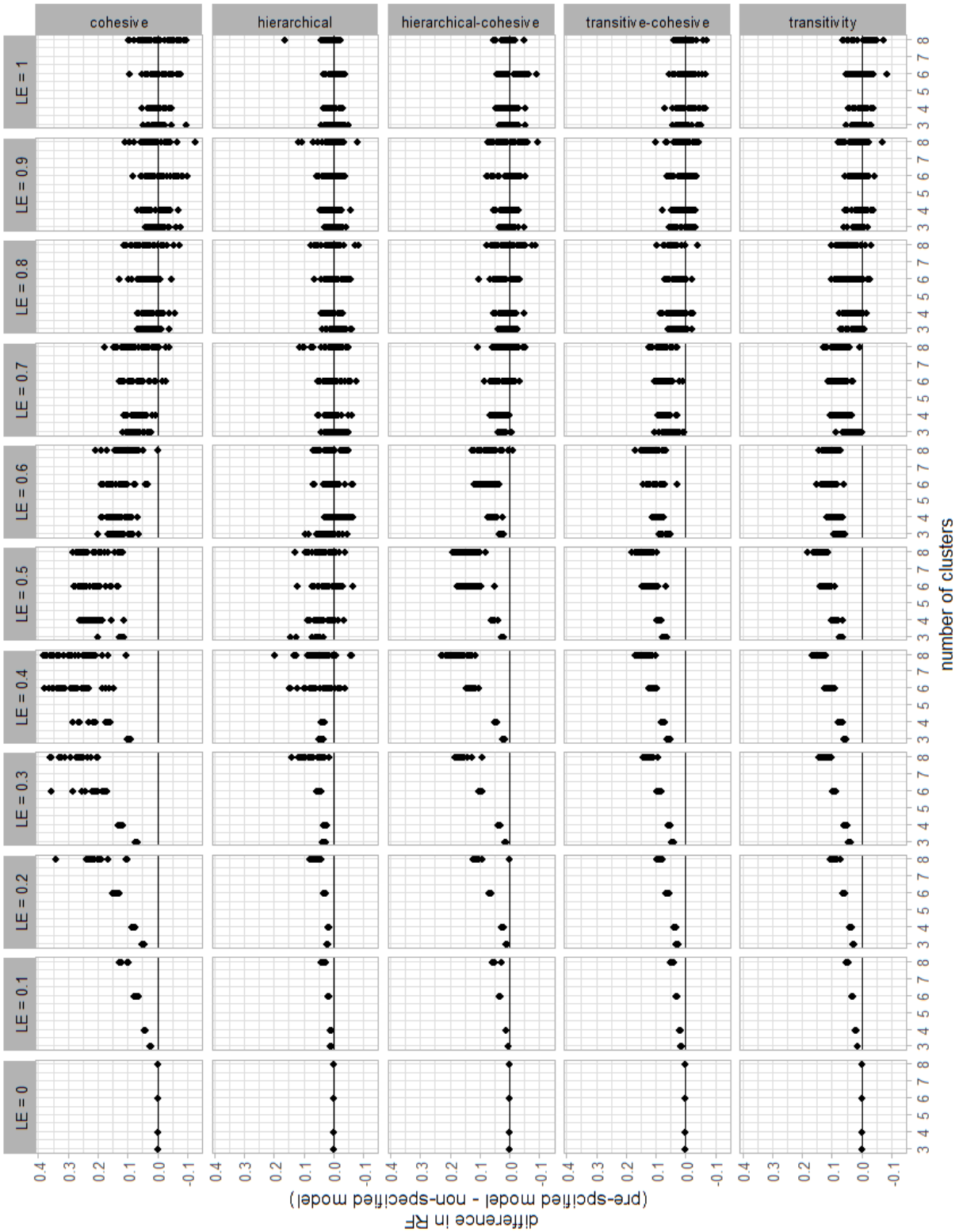
Direct blockmodeling with a non-specified and with a pre-specified model is applied to each generated network. The true number of clusters is used and, for the pre-specified model, the true blockmodel is considered. Each RF is calculated based on 30 random networks.

The results are shown in Figure . When the blockmodels without inconsistencies ($LE = 0$) are considered, the difference between the RF corresponding to the pre-specified model and the RF corresponding to the non-specified model (simply called difference) equals 0 because both blockmodeling approaches are able to find a blockmodel without any inconsistency.

When LE is increasing, the difference is also increasing. This means that RF values corresponding to pre-specified blockmodeling are generally higher than the RF values corresponding to the non-specified model. This is due to the different CF values obtained by the non-specified and pre-specified models, where CF is higher in the case of randomized networks compared to when original (non-randomized) networks are being analysed. The values of CF obtained with non-specified blockmodeling are usually lower than those obtained by pre-specified blockmodeling.

The difference starts to shrink at higher levels of errors, namely $LE > 0.5$. At such a level of errors, it is hard to talk about the presence of a given blockmodel type in the network (see subsection 2.5.4). The mean difference in RF when it comes to totally random networks is very close to 0.

Figure 1: The difference between the relative fit values obtained by pre-specified blockmodeling and the relative fit values obtained by non-specified blockmodeling, controlling for different numbers of clusters and different levels of errors



Razširjen povzetek

Uvod

Eno izmed osrednjih prizadevanj v sociologiji je razumeti družbene mehanizme, ki vplivajo na določen družben pojav. V primeru proučevanja odnosov med posamezniki je mogoče izbran družbeni pojav predstaviti v obliki omrežja, kjer vozlišča predstavljajo posameznike, povezave med njimi pa izbrano vrsto odnosa. V takem primeru je mogoče družbene mehanizme opisati kot silnice, ki vplivajo na vzpostavljanje in prekinjanje povezav v omrežju in prek tega na globalno zgradbo omrežja oziroma pojav določenega družbenega pojava.

V okviru analize omrežij je razvitih mnogo pristopov za proučevanje dinamike v omrežjih, a noben neposredno ne obravnava odnosa med lokalnimi omrežnimi mehanizmi in globalnimi omrežnimi zgradbami, opisanimi z bločnimi modeli, kar je namen pričujoče disertacije. V preliminarnem delu raziskave je naslovljeno tudi vprašanje, ali je mogoče generirati omrežja z izbranimi globalnimi omrežnimi zgradbami z upoštevanjem različnih vrst triad.

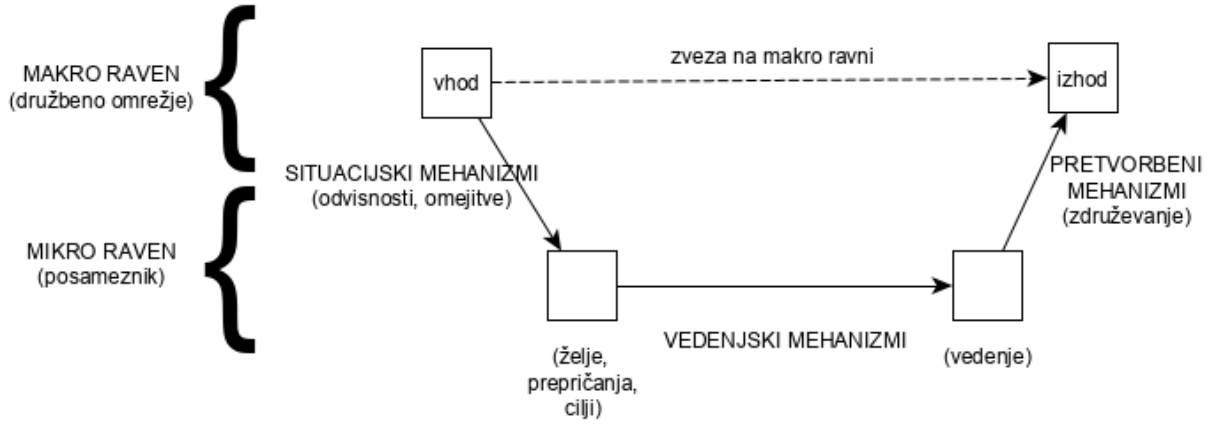
Lokalni omrežni mehanizmi

Eden izmed prvih sociologov, ki je predstavil idejo o družbenih mehanizmih, je bil Robert Merton. Čeprav ni nikoli razvil zelo jasne opredelitve koncepta (Hedström in Wennberg, 2017), pa je znano, da je o družbenih mehanizmih razmišljal kot o »družbenih procesih, ki imajo določen vpliv na določene dele družbene zgradbe« (Merton, 1949, str. 451). Gambetta (1998, str. 102) je družbene mehanizme opredelil kot »hipotetične vzročne modele, ki pojasnjujejo posameznikovo vedenje«. Imeli naj bi obliko »v določenih pogojih bo enota z verjetnostjo p storila x zaradi [mehanizma] M« (prav tam). Tako je ena izmed najpomembnejših vlog družbenih mehanizmov vloga, ki jo imajo pri pojasnjevanju družbenih pojavov. Na primer, raziskovalec, ki razmišlja v kontekstu družbenih mehanizmov, ni zadovoljen zgolj z morebitno opaženo korelacijo med I (vhod; angl. *imput*) in O (izhod*;* angl. *output),* temveč želi pojasniti vzrok opažene korelacije.

Pri proučevanju družbenih omrežij sta I in O pogosto dve različni globalni zgradbi omrežja. Pri iskanju družbenih mehanizmov, ki bi lahko vplivali na spremembo globalne zgradbe omrežja, je treba dovolj podrobno opisati spreminjanje povezav na ravni vozlišč, kar je prvi pogoj za opis spremembe globalne zgradbe omrežja. Tako je, v skladu s Colemanovim makro-mikro-makro

modelom (Coleman, 1986), družbene mehanizme mogoče razvrstiti v tri skupine (Slika 1): (i) situacijski mehanizmi (angl. *situational mechanisms*), ki so povezani z vplivom globalne zgradbe omrežja na posameznikove želje, prepričanja, cilje, in podobnega; (ii) vedenjski mehanizmi (angl. *action-formation mechanisms*), ki so povezani z vplivom posameznikovih želja, prepričanj, ciljev in podobnega, ki vplivajo na posameznikovo vedenje; ter (iii) pretvorbeni mehanizmi (angl. *transformational mechanisms*), ki so povezani z vplivom posameznikovega vedenja na globalno zgradbo omrežja (Hedström in Swedberg, 1998; Stadtfeld, 2018).

Slika 1 Colemanov makro-mikro-makro model z vidika analize omrežij



Pri proučevanju družbenih mehanizmov se pogosto uporabljajo modeli na osnovi agentov (angl. *agent-based models*), ki niso nujno namenjeni pojasnitvi točno določenega empiričnega pojava, temveč omogočajo okvir za razmišljanje o pojavu na bolj splošni ravni (Hedström in Ylikoski, 2010). Z uporabo tovrstnih modelov se predpostavlja, da je mogoče osnovne značilnosti zelo zapletenih družbenih pojavov razložiti z zelo preprostimi matematičnimi modeli (Hedström in Ylikoski, 2011, str. 396). Pogosto se izkaže, da je mogoče izid na makroravni pojasniti že zgolj na osnovi dogodkov na mikroravni. V primerjavi s klasično teorijo iger in analitičnim modeliranjem je mogoče z modeli na osnovi agentov upoštevati heterogenost družbenega vedenja posameznikov ali pa proučevati neuravnoteženo družbeno dinamiko (angl. *out-of-equilibrium social dynamic*). V primerjavi s statističnimi modeli, ki temeljijo na spremenljivkah, je mogoče z modeli na osnovi agentov upoštevati različne mikrogenerativne družbene procese (angl. *micro-generative processes*) (Bianchi in Squazzoni, 2015). Pogosto uporabljeni verjetnostni modeli na ravni

posameznika (angl. *Stochastic Actor-Oriented Models;* SAOM) (Block in drugi, 2016; Snijders, 2001) so poseben primer modelov na osnovi agentov (Snijders in Steglich, 2015).

V okviru slednjih se najpogosteje, kot lokalni omrežni mehanizmi, omenjajo mehanizem vzajemnosti (težnja k vzpostavljanju vzajemnih povezav), mehanizem popularnosti (težnja k vzpostavljanju povezav k bolj popularnim vozliščem v omrežju, kjer je popularnost običajno opredeljena z upoštevanjem vhodne stopnje vozlišča) ter mehanizmi, ki so povezani s tranzitivnostjo (na primer, težnja k vzpostavljanju povezav k vozliščem, ki imajo vzpostavljene povezave k veliko skupnim vozliščem, kot vozlišče, ki vzpostavlja povezavo) ali podobnostjo oziroma homofilijo (težnja k vzpostavljanju povezav k podobnim vozliščem, glede na neko lastnost). Obstajajo še drugi omrežni mehanizmi.

Globalne omrežne zgradbe

Globalne zgradbe omrežij so pogosto opisane s tako imenovanimi bločnimi modeli. Bločni model je omrežje, kjer vozlišča predstavljajo skupine enakovrednih vozlišč iz proučevanega omrežja. Obstaja več vrst enakovrednosti vozlišč, od katerih je najbolj znana in najpogosteje uporabljena strukturna enakovrednost (angl. *structural equivalence*) (Lorrain in White, 1971). Vozlišči sta strukturno enakovredni, če sta z drugimi vozlišči v omrežju enako povezani. Izraz blok se nanaša na podmatriko matrike sosednosti in prikazuje povezave med vozlišči iz dveh različnih skupin ali med vozlišči znotraj ene skupine (Doreian in drugi, 2005). V primeru strukturne enakovrednosti sta mogoči dve vrsti blokov, in sicer poln blok (angl. *complete block*) in prazen blok (angl. *null block*). V idealnem primeru znotraj praznega bloka ni nobene povezave, znotraj polnega bloka pa so vse možne povezave. V primerih empiričnih omrežij se praviloma pojavi nekaj povezav tudi v praznih blokih in nekaj nepovezav v polnih blokih. Takšne povezave se imenujejo napake (angl. *errors*) ali neskladnosti (angl. *inconsistencies*). Najbolj značilne so naslednje vrste bločnih modelov (Doreian in drugi, 2005):

- **Kohezivni bločni model** (Slika 2a): V primeru kohezivnega bločnega modela so vozlišča znotraj različnih kohezivnih skupin dobro povezana, medtem ko vozlišča iz različnih kohezivnih skupin med seboj niso povezana.
- **Simetrični središčno-periferni bločni model** (Slika 2b): Navadno je sestavljen iz dveh skupin. Vozlišča iz ene skupine se imenujejo središčna vozlišča (to so vozlišča, ki so med

seboj dobro povezana), vozlišča iz druge skupine pa se imenujejo periferna vozlišča. Slednja med seboj niso povezana, a so povezana z drugimi, središčnimi vozlišči, ki pa so povezana med seboj.

- **Asimetrični središčno-periferni bločni model**: Podoben je simetričnemu središčno-perifernemu bločnemu modelu, le da v primeru asimetričnega obstajajo povezave le iz perifernih k centralnim vozliščem (Slika 2c).
- **Hierarhični bločni model** (Slika 2d): V primeru hierarhične vrste bločnega modela je mogoče skupine vozlišč urediti hierarhično. Kadar hierarhični bločni model sestoji iz treh skupin, takrat obstaja skupina med seboj nepovezanih vozlišč, ki pa so močno povezane z naslednjo skupino nepovezanih vozlišč. Slednja je močno povezana z vozlišči iz tretje skupine, katere so prav tako nepovezane med seboj. Vozlišča iz prve skupine niso povezana z vozlišči iz tretje skupine, kot to velja za tranzitivni in tranzitivno-kohezivni bločni model.
- **Hierarhično-kohezivni bločni model** (Slika 2e): Tudi v primeru hierarhično-kohezivnega bločnega modela je mogoče skupine urediti hierarhično. V nasprotju s hierarhičnim bločnim modelom so enote v primeru hierarhično-kohezivnega bločnega modela znotraj skupin med sabo močno povezane.
- **Tranzitivni bločni model** (Slika 2f): Tranzitivni model je podoben hierarhičnemu modelu, le da v primeru tranzitivnega obstajajo povezave iz skupine iz višjega nivoja z vozlišči, ki pripadajo skupinam na vseh nižjih nivojih.
- **Tranzitivno-kohezivni bločni model** (Slika 2g): Ta bločni model je podoben tranzitivnemu bločnemu modelu, le da so vozlišča v primeru tranzitivno-kohezivnega bločnega modela znotraj skupin močno povezana.

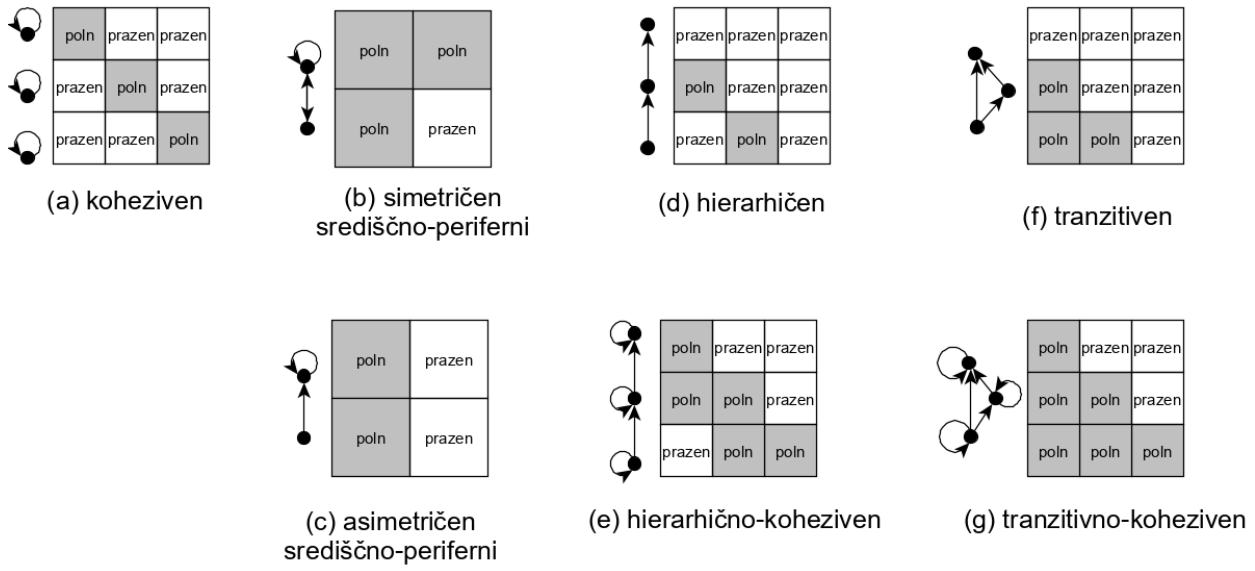
Zgoraj so naštete le najbolj značilne vrste bločnih modelov ter lokalnih omrežnih mehanizmov, a obstaja še zelo veliko drugih.[34] Zaradi tega je v disertaciji analiziranih nekaj vrst bločnih modelov

[34] Že zgoraj navedene asimetrične vrste bločnih modelov (mednje spadajo vse naštete vrste bločnih modelov, razen kohezivnega in asimetričnega središčno-perifernega) je mogoče opredeliti drugače, in sicer na način, ki upošteva povezave od enot iz skupin na višjih hierarhičnih ravneh k enotam iz skupin na nižjih hierarhičnih ravneh (v primeru asimetričnega središčno-perifernega bločnega modela pa povezave od središčnih k perifernim enotam). Za nastanek takšnih vrst bločnih modelov bi bilo nemara treba upoštevati drugačne lokalne omrežne mehanizme, kot v primeru izbranih vrst bločnih modelov.

in lokalnih omrežnih mehanizmov, ki so izbrani na osnovi upoštevanja različnih družbenih kontekstov. Upoštevanje družbenega konteksta namreč bistveno pripomore k izbiri najprimernejših lokalnih omrežnih mehanizmov in pripadajočih bločnih modelov.

Obravnavana sta dva družbena konteksta. Prvi se nanaša na razred predšolskih otrok (omrežja prijateljstev, omrežja naklonjenosti in omrežja interakcij so analizirana v okviru tega družbenega konteksta), drugi pa na delovno okolje srednje velikega, na znanju temelječega podjetja (omrežja pretoka znanja so analizirana v okviru tega družbenega konteksta). Obravnavanje dveh družbenih kontekstov naslavlja enega izmed osrednjih raziskovalnih vprašanj pričujoče disertacije, in sicer »Kateri mehanizmi (ali kombinacija več mehanizmov) vplivajo na vzpostavitev določene vrste bločnega modela?«.

Slika 2 Različne vrste bločnih modelov



Disertacija se začne z analizo zmožnosti generiranja najpogostejših vrst bločnih modelov zgolj z upoštevanjem različnih vrst triad. Zmožnost generiranja omrežij na tak način lahko nakazuje obstoj lokalnih omrežnih mehanizmov, ki bi lahko povzročili nastanek izbrane vrste bločnega modela, brez upoštevanja lastnosti vozlišč (spremenljivk) (angl. *attributes*), kot sta spol in starost. Neupoštevanje atributov vozlišč je sicer glavna omejitev raziskave. Namen te omejitve je poiskati bolj splošne povezave med lokalnimi omrežnimi mehanizmi in globalnimi omrežnimi zgradbami. Družben kontekst je torej upoštevan zgolj pri izbiri lokalnih omrežnih mehanizmov in globalnih omrežnih zgradb, ne pa tudi pri proučevanju povezave med njima.

Generiranje omrežij z izbranimi bločnimi modeli z upoštevanjem različnih vrst triad

Odnos med različnimi vrstami triad in različnimi globalnimi zgradbami omrežja so proučevali že zelo zgodnji raziskovalci s področja analize omrežij (Davis in Leinhardt, 1967; Holland in Leinhardt, 1970; Johnsen, 1985). Davis in Leinhardt (1967), avtorja klasifikacije MAN,[35] ki povzema vse možne vrste grafov velikosti tri (triad), sta dokazala, da je potreben in zadosten pogoj za obstoj določene globalne zgradbe omrežja, ki sta jo proučevala, prisotnost sedmih točno določenih vrst triad. Podobne klasifikacije dovoljenih in prepovedanih vrst triad so bile narejene še za nekatere druge vrste globalnih zgradb omrežij. Takšna klasifikacija je omogočala preverjanje prisotnosti izbranih globalnih zgradb v empiričnih omrežjih z metodo ugotavljanja prisotnosti dovoljenih in prepovedanih vrst triad v omrežju. Omenjena metodologija pa ima omejen obseg uporabe pri analizi empiričnih omrežij z določeno ravnjo napak. Holland in Leinhardt (1970) sta zato poudarjala pomembnost verjetnostnega pristopa k preverjanju globalnih zgradb omrežij na osnovi prisotnosti ali odsotnosti določenih vrst triad. Tako sta izpeljala porazdelitev različnih vrst triad v slučajnih omrežjih in definirala testno statistiko

$$\tau = \frac{T - \mu_T}{\sigma_T} \tag{0.1}$$

kjer je T število določene vrste triad v empiričnem omrežju, μ_T je pričakovano število triad v slučajnem omrežju in σ_T je varianca števila istih vrst triad v slučajnem omrežju. Holland in Leinhardt (1970) sta predpostavila, da se testna statistika porazdeljuje asimptotično normalno, kar sta upoštevala pri preverjanju števila triad v empiričnih omrežjih.

Danes se pogosto uporablja tudi pojem »motif«, ki je definiran kot vzorec povezav, ki se v omrežju pojavlja pogosteje, kot bi bilo pričakovati po slučaju (Milo in drugi, 2002), kar poleg triad vključuje tudi druge vrste (običajno večjih) podomrežij. Podomrežja velikosti treh vozlišč so pogosto upoštevana, ker bi bilo upoštevanje podomrežij velikosti dveh vozlišč nezadostno (na

[35] Vsaki vrsti triade je pripisana oznaka iz treh števk: prva števka označuje število vzajemnih povezav (angl. *mutual*), druga števka označuje število asimetričnih povezav (angl. *asymmetric*), tretja števka pa označuje število nepovezav oziroma število negativnih povezav (angl. *non-link* ali *negative link*). Od tod izhaja poimenovanje »MAN«. Nekatere vrste triad so nadalje označene s črko (C označuje cikel, T označuje tranzitivnost, U označuje »navzgor« in D označuje »navzdol«).

osnovi tega bi bilo mogoče proučevati zgolj nepovezave, usmerjene povezave ter vzajemne povezave), po drugi strani pa upoštevanje podomrežij z večjim številom vozlišč lahko zelo poveča računsko zahtevnost ocene pričakovanega števila izbranih vzorcev povezav v slučajnih omrežjih. Davis in Leinhardt (1967) sta prepoznala triade kot najmanjšo enoto, ki omogoča opazovanje oziroma proučevanje odnosov med več vozlišči. Kljub nekaterim opozorilom, da je treba biti previden pri definiranju slučajnih omrežij (ta predstavljajo osnovo pri ugotavljanju, ali se določeno podomrežje pojavi pogosteje, kot bi bilo pričakovati po slučaju) (Artzy-Randrup in drugi, 2004), pa so raziskovalci motife uspešno uporabili v mnogo raziskavah, na primer za opisovanje hierarhične zgradbe omrežja v možganih (Yu in Gerstein, 2006) in za razvrščanje empiričnih omrežij v skupine (Milo in drugi, 2004).

Kljub pogosti uporabi različnih vrst triad pri opisovanju globalnih omrežnih zgradb pa do zdaj še niso bile sistematično proučevane v okviru najpogostejših vrst bločnih modelov. Čeprav je znano, da se v določenih globalnih zgradbah omrežij določene vrste triad pojavljajo pogosteje kakor v drugih, pa ni znano, ali je mogoče različne vrste triad povezati z nastankom takih globalnih omrežnih zgradb.

Za odgovor na to vprašanje so vse vrste triad, za vsako vrsto izbranega bločnega modela, razvrščene v skupino prepovedanih ali v skupino dovoljenih vrst triad. Razvrstitev je opravljena na osnovi idealnih omrežij brez napak z izbranim bločnim modelom. Izbrana vrsta triade je razvrščena v množico dovoljenih vrst triad, če se pojavi v idealnem omrežju z izbranim bločnim modelom. V nasprotnem primeru pa je razvrščena v množico prepovedanih vrst triad. Izkaže se, da je na osnovi takšne razvrstitve mogoče razločevati med vsemi že omenjenimi vrstami bločnih modelov.

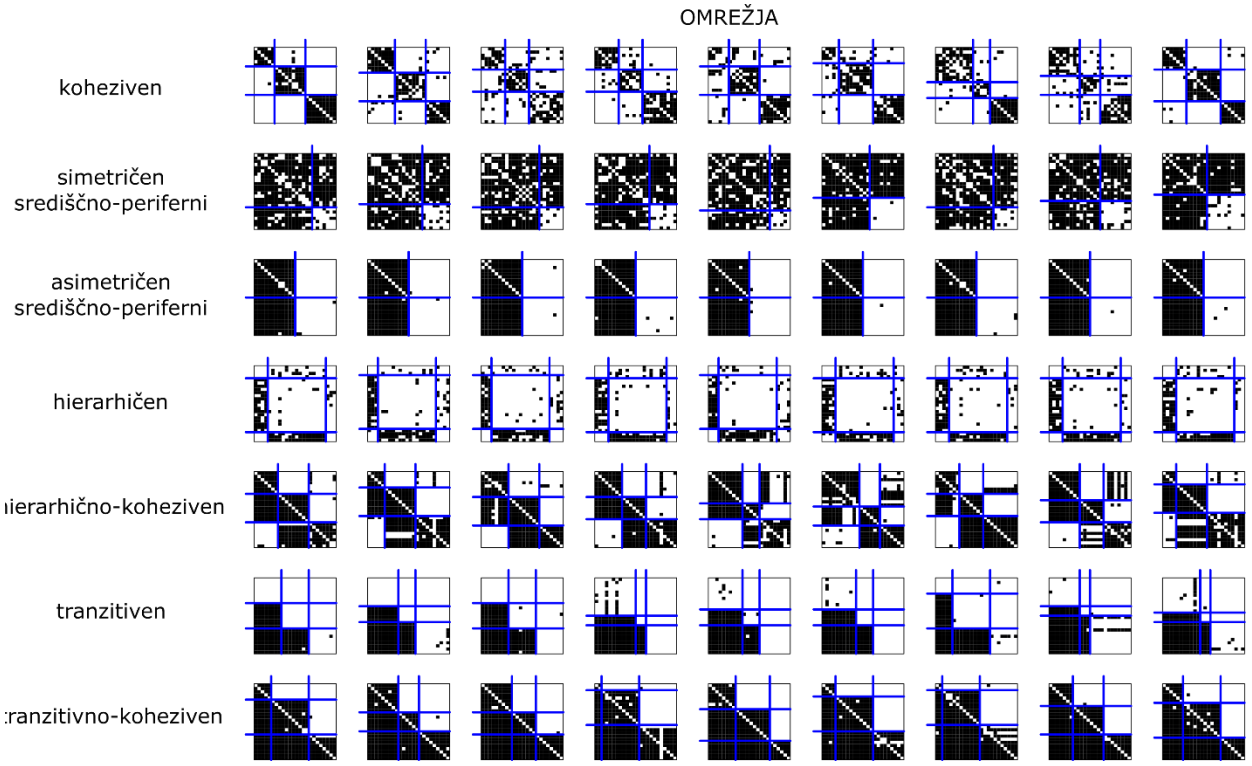
Množica dovoljenih in množica prepovedanih vrst triad sta uporabljeni za generiranje omrežij z izbrano vrsto bločnega modela. Za generiranje omrežij sta uporabljena dva različna algoritma. Prvi je algoritem prestavljanja povezav (algoritem RL), drugi pa algoritem na osnovi Monte Carlo Markovskih verig (angl. *Markov Chain Monte Carlo;* MCMC), ki je implementiran v paketku »ergm« (Hunter in drugi, 2013) za programski jezik R (Team, 2000). Dva algoritma sta uporabljena z namenom zmanjšanja verjetnosti nezmožnosti generiranja omrežij kot posledice lastnosti posameznega algoritma.

Rezultati kažejo, da je vse proučevane vrste bločnih modelov mogoče generirati z upoštevanjem izbranih vrst triad (Slika 3). V splošnem obseg napak v generiranih omrežjih ni bistveno višji, ko je pri generiranju upoštevana zgolj množica prepovedanih vrst triad, v primerjavi z generiranimi omrežji na osnovi obeh množic vrst triad – dovoljenih in prepovedanih. Ima pa upoštevanje zgolj množice prepovedanih vrst triad pomembno praktično prednost pri generiranju omrežij z izbrano vrsto bločnega modela. Namreč, raziskovalec mora (če želi generirati omrežja z uporabo algoritma RL na osnovi množice dovoljenih vrst triad) določiti število vsake izmed vrst triad v idealnem omrežju z izbrano vrsto bločnega modela. To lahko vsebuje informacijo o številu in velikosti skupin v omrežju. Po drugi strani pa mora raziskovalec, če želi generirati omrežja z uporabo algoritma RL, na osnovi množice prepovedanih vrst triad določiti samo množico prepovedanih vrst triad, ne pa tudi pojavnosti posameznih vrst triad v idealnem omrežju z dano vrsto bločnega modela (saj je število vsake prepovedane vrste triade vedno 0). Seznam prepovedanih vrst triad lahko še vedno vsebuje podatek o številu skupin v omrežju, a je količina informacij, ki jih mora algoritmu posredovati raziskovalec, bistveno manjša.

Kakorkoli, generirana omrežja s ciljnim hierarhičnim bločnim modelom navadno vsebujejo nekoliko višjo stopnjo napak v primerjavi z generiranimi omrežji z drugimi vrstami bločnih modelov. To je še posebej izrazito, ko so omrežja generirana z algoritmom MCMC. V primeru, ko je uporabljen algoritem RL in so upoštevane vse vrste triad, imajo generirana omrežja želen hierarhičen bločni model, a so nekatere skupine izrazito majhne. Upoštevanje dodatnih vrst podomrežij, kot so poti dolžine tri, rezultira v generiranih omrežjih skoraj brez napak in s približno enako velikimi skupinami.

Glavna ugotovitev tega dela disertacije je, da lahko izbrane vrste bločnih modelov nastanejo kot posledica lokalnih omrežnih zgradb, brez upoštevanja dodatnih lastnosti vozlišč. To je pokazatelj obstoja zapletenejših lokalnih omrežnih mehanizmov, ki lahko povzročijo nastanek omrežij z izbranimi vrstami bločnih modelov. Take vrste lokalnih omrežnih mehanizmov so obravnavane v nadaljevanju.

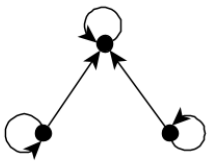
Slika 3 Nekaj primerov omrežij z različnimi bločnimi modeli, generiranimi z uporabo algoritma RL z upoštevanjem izbranih vrst triad



Nastanek simetričnega in asimetričnega središčno-kohezivnega bločnega modela

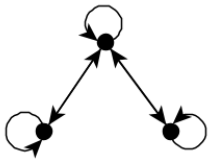
V disertaciji sta predstavljena simetrični (Slika 4a) in asimetrični (Slika 4b) središčno-kohezivni bločni model. Oba sta mešanica kohezivnega ter središčno-perifernega bločnega modela. Sestavljena sta iz vsaj treh skupin. V primeru asimetričnega bločnega modela so vsa vozlišča v omrežju neposredno povezana z vozlišči iz središčne skupine. Vsa vozlišča znotraj posameznih kohezivnih skupin so med seboj povezana, vozlišča iz različnih kohezivnih skupin pa med seboj niso povezana.

Slika 4 Simetričen in asimetričen središčno-kohezivni bločni model s tremi skupinami



poln	prazen	prazen
poln	poln	prazen
poln	prazen	poln

(a) asimetričen središčno-kohezivni



poln	poln	poln
poln	poln	prazen
poln	prazen	poln

(b) simetričen središčno-kohezivni

Na osnovi ugotovitev predhodnih raziskav je mogoče domnevati, da lahko asimetričen središčno-koheziven bločni model nastane v omrežjih (nominacij) prijateljstev ali v omrežjih naklonjenosti med predšolskimi otroki, simetričen središčno-kohezivni bločni model pa bi lahko nastal v omrežjih interakcij med predšolskimi otroki. Obsežni pregled literature kaže, da se z nastankom in razvojem tovrstnih omrežij najpogosteje povezujejo naslednji lokalni omrežni mehanizmi: mehanizem vzajemnosti, mehanizem popularnosti, mehanizem podobnosti (glede na vhodne stopnje vozlišč) ter različni mehanizmi, povezani s tranzitivnostjo. V povezavi z mehanizmom podobnosti se v literaturi pogosto pojavljajo lastnosti vozlišč, kot sta spol in starost, pa tudi nekatere psihosocialne lastnosti (DeLay in drugi, 2016; Kandel, 1978; McPherson in drugi, 2001). Zaradi omejitev disertacije na neupoštevanje lastnosti vozlišč te niso upoštevane v analizah.

Za odgovor na vprašanje o možnosti nastanka izbranih vrst bločnih modelov, kot posledice vpliva izbranih lokalnih omrežnih mehanizmov, je predlagan model iz družine modelov razvoja omrežij (angl. *Network Evolution Models*; NEM). Model je definiran z iterativnim algoritmom. V vsaki iteraciji je slučajno izbrano eno vozlišče. Izbrano vozlišče, z določeno verjetnostjo q, določi povezavo do slučajno izbranega drugega vozlišča, ki ima visoko vrednost utežene omrežne statistike. Nato, z verjetnostjo $1 - q$, določi nepovezavo do slučajno izbranega vozlišča z nizko vrednostjo utežene omrežne statistike. Omrežna statistika je določena kot linearna kombinacija lokalnih omrežnih mehanizmov.[36] Algoritem se zaključi po izbranem številu iteracij.

Uteži lokalnih omrežnih mehanizmov (običajno zapisane v obliki vektorja θ dolžine k, kjer je k število upoštevanih lokalnih omrežnih mehanizmov) odražajo stopnjo njihove pomembnosti oziroma moči (poljubno jih lahko določi raziskovalec, lahko tudi na osnovi slučajno generiranih vrednosti uteži). Kljub naporom za dosego primerljivosti uteži, ki pripadajo različnim lokalnim omrežnim mehanizmom, pa te žal niso splošno primerljive, saj različni lokalni omrežni mehanizmi niso neodvisni. Podoben problem primerljivosti je skupen tudi SAOM in modelom eksponentnih slučajnih grafov (angl. *Exponential Random Graph Models*; ERGM) (Goodreau, 2007; Hunter in drugi, 2008; Robins in drugi, 2007). Čeprav je bila problematika primerljivosti uteži že večkrat naslovljena (Indlekofer in Brandes, 2013; Snijders, 2004; Snijders, Van de Bunt, in drugi, 2010),

[36] V okviru algoritmov NEM se izraz "mehanizem" nanaša na operacionalizacije lokalnih omrežnih mehanizmov (kot so opredeljeni v poglavju Lokalni omrežni mehanizmi) z različnimi statistikami (na primer stopnja vozlišča), ki odražajo izbrane lokalne omrežne mehanizme.

pa trenutno ne obstaja nobena splošno sprejeta rešitev. V vsakem primeru pa je mogoče v grobem primerjati vsaj predznake uteži ter uteži z ekstremno različnimi vrednostmi (na primer uteži, katerih vrednost je blizu 0, z utežmi, katerih vrednost je blizu 1).[37]

To je pomembno, saj so uteži mehanizmov v tej disertaciji določene slučajno. Na osnovi slednjih je generiranih mnogo omrežij, katerih globalne zgradbe so preverjane z uporabo različnih obstoječih in predlaganih mer prileganja bločnih modelov. Že obstoječa mera je število neskladnih blokov (angl. *inconsistent blocks*) (Žnidaršič in drugi, 2012), ki je definirana skozi število različnih vrst blokov med empiričnim bločnim modelom ter med izbranim (idealnim) bločnim modelom in se uporablja kot kriterij za ugotavljanje prisotnosti izbranega bločnega modela v podatkih. Predlagana pa je tudi normalizirana vrednost kriterijske funkcije, poimenovana funkcija relativnega prileganja (angl. *Relative Fit function*; funkcija RF), ki se uporablja za oceno ravni napak v empiričnem omrežju v primerjavi z danim bločnim modelom. Definirana je kot razmerje med vrednostjo kriterijske funkcije, dobljene na empiričnem omrežju, ter med pričakovano vrednostjo kriterijske funkcije, ocenjene na slučajnih omrežjih. Višja vrednost funkcije RF pomeni boljše prileganje (nižjo raven napak) empiričnega omrežja v primerjavi z izbranim bločnim modelom.

Asimetrični središčno-kohezivni bločni model

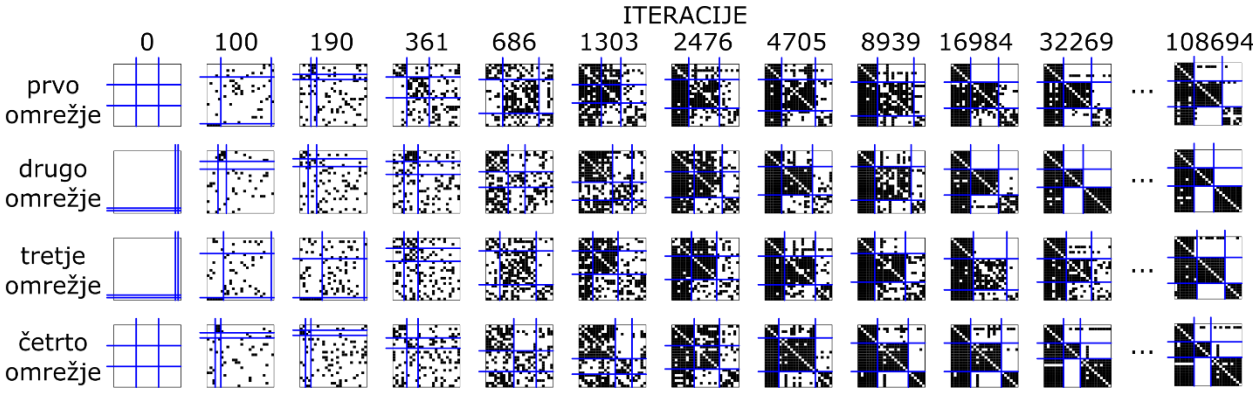
Rezultati Monte Carlo simulacij potrjujejo, da lahko asimetričen središčno-koheziven bločni model nastane kot posledica mehanizma vzajemnosti, mehanizma popularnosti, mehanizma podobnosti stopenj ter dveh vrst mehanizmov, povezanih s tranzitivnostjo (primer generiranih omrežij z upoštevanjem naštetih lokalnih omrežnih mehanizmov je na Sliki 5). Slednja sta mehanizem izhodnih poti dolžine dva (angl. *outgoing two-paths mechanism*; OTP) in mehanizem izhodnih skupnih vozlišč (angl. *outgoing shared-partners mechanism*; OSP). To velja za vse upoštevane začetne zgradbe omrežij: prazno omrežje, omrežje z asimetričnim središčno-perifernim bločnim modelom ter omrežje s kohezivnim bločnim modelom.

Za vsako vrsto bločnega modela je izbranih deset vektorjev uteži θ, ki generirajo omrežja s povprečno najnižjim številom neskladnih blokov. Izkaže se, da so izbrane θ podobne v primeru,

[37] V tej disertaciji so uteži definirane tako, da lahko zavzamejo vrednosti med -1 in 1, vsota kvadriranih vrednosti uteži vseh upoštevanih mehanizmov pa je enaka 1.

ko je začetno omrežje prazno omrežje, ter v primeru, ko imajo začetna omrežja kohezivni bločni model. Množici izbranih θ se nekoliko bolj razlikujeta od primera, ko imajo začetna omrežja asimetričen središčno-periferen bločni model. V tem primeru so uteži mehanizma izhodnih poti dolžine dva nekoliko višje. To je pričakovano, saj morajo v primeru generiranja omrežij z začetnim asimetričnim središčno-perifernim bločnim modelom nastati kohezivne skupine, katerih nastanek se pogosto povezuje z mehanizmom izhodnih poti dolžine dva. To pa ne pomeni, da izbrane θ ne morejo povzročiti nastanka asimetričnega središčno-kohezivnega bločnega modela iz katerekoli (analizirane) začetne zgradbe omrežja – običajno je treba le povečati število iteracij v algoritmu za generiranje omrežij.

Slika 5 Nekaj primerov generiranih omrežij z asimetričnim središčno-kohezivnim bločnim modelom



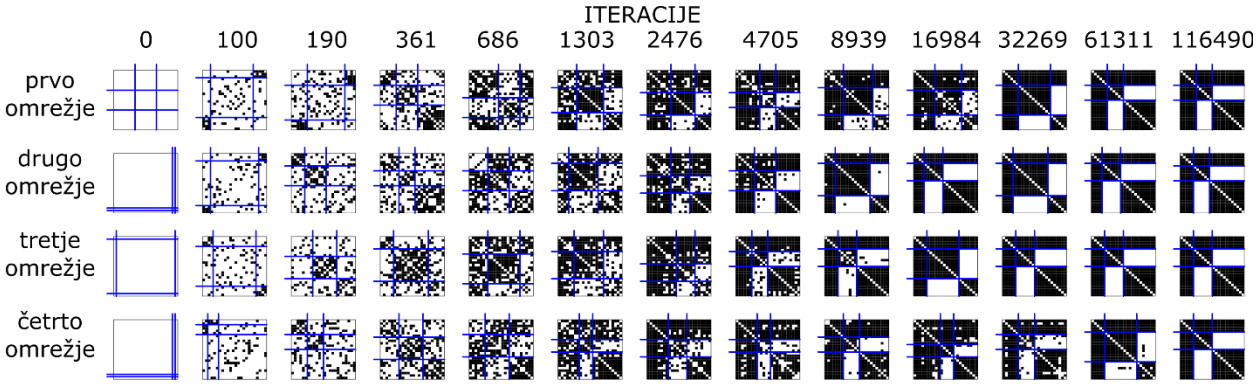
Opomba: Omrežja so narisana v skladu z dobljenim bločnim modelom. Začetno omrežje je prazno omrežje.

Simetrični središčno-kohezivni bločni model

V nasprotju z asimetričnim središčno-kohezivnim bločnim modelom so za primer simetričnega središčno-kohezivnega bločnega modela analizirana tudi empirična omrežja, ki so bila zbrana med predšolskimi otroki v ZDA (analizirani so sekundarni podatki) (Schaefer in drugi, 2010) v več šolskih razredih in skozi celotno šolsko leto. Namen tovrstne analize je preveriti, ali se izbrana globalna zgradba omrežja pojavljajo v empiričnih omrežjih in tako, po eni strani, ugotoviti smiselnost proučevanja izbrane globalne zgradbe, po drugi strani pa preveriti smiselnost upoštevanja izbiralnega družbenega konteksta za to vrsto globalne zgradbe. Rezultati bločnega modeliranja kažejo, da se izbrana globalna zgradba omrežja pojavijo v skoraj vsakem proučevanem razredu v vsaj eni opazovani časovni točki.

Analize nastanka simetričnega središčno-kohezivnega bločnega modela, ki so opravljene s pomočjo Monte Carlo simulacij, kažejo podobne rezultate, kot analize za asimetričen primer: izbrana vrsta bločnega modela lahko nastane kot posledica mehanizma vzajemnosti, mehanizma popularnosti, mehanizma podobnosti stopenj ter mehanizmov, povezanih s tranzitivnostjo (mehanizem izhodnih poti dolžine dva in mehanizem izhodnih skupnih vozlišč). Primer generiranih omrežij prikazuje Slika 6.

Slika 6 Nekaj primerov generiranih omrežij s simetričnim središčno-kohezivnim bločnim modelom



Opomba: Omrežja so narisana v skladu z dobljenimi bločnimi modeli. Začetno omrežje je prazno omrežje.

Rezultati so pričakovani, saj sta lahko prijateljstvo oziroma naklonjenost (ki sta analizirana v primeru asimetričnega bločnega modela) vzrok za začetek interakcije, ki je analizirana v primeru simetričnega bločnega modela. To je tudi razlog za izbiro enakih lokalnih omrežnih mehanizmov v primeru obeh vrst bločnih modelov. Tudi uporabljeni algoritem za generiranje simetričnih in asimetričnih omrežij je enak. Kljub različnim možnim načinom generiranja simetričnih omrežij je izbrani način generiranja omrežij najbližji način posnemanja nastanka analiziranih omrežij interakcij.

Pri generiranju omrežij z Monte Carlo simulacijami se stremi k čim bolj natančni poustvaritvi procesa nastajanja empiričnih omrežij, ki ga v tem primeru predstavlja postopek zbiranja omrežij interakcij med otroki v vrtcu. Interakcije so opredeljene kot asimetričen družben proces, pri katerem je vedno vozlišče i tisto, ki začne interakcijo z drugim vozliščem j. Glede na stopnjo pomembnosti lokalnega omrežnega mehanizma vzajemnosti vozlišču j (in glede na nekatere druge lokalne omrežne mehanizme) lahko vozlišče j nadaljuje interakcijo z vozliščem i ali pa se umakne.

V tem oziru je, pri generiranju omrežij, treba upoštevati tudi lokalni omrežni mehanizem vzajemnosti. V primeru analiziranih empiričnih omrežij raziskovalci niso zabeležili podatka o tem, kdo je začel interakcijo, zaradi česar so omrežja analizirana kot simetrična. Tako so tudi omrežja, generirana z algoritmom NEM, pred analizo simetrizirana.

Druge vrste bločnih modelov

Obravnavano je tudi generiranje preostalih vrst bločnih modelov z upoštevanjem enakih lokalnih omrežnih mehanizmov kot v primeru generiranja simetričnega in asimetričnega središčno-kohezivnega bločnega modela.

Eno izmed pomembnejših spoznanj je, da mehanizem popularnosti povzroči nastanek omrežja z asimetričnim središčno-kohezivnim bločnim modelom, medtem ko kombinacija lokalnih omrežnih mehanizmov popularnosti in vzajemnosti povzroči nastanek omrežij s simetričnim središčno-perifernim bločnim modelom.

V primeru simetričnih omrežij obstaja zgolj ena vrsta lokalnega omrežnega mehanizma tranzitivnosti (imenovan tranzitivnost ali pa učinek zapiranja triad) (angl. *triad closure*), ki se pogosto povezuje z nastankom kohezivnih skupin. V primeru asimetričnih omrežij pa obstaja več vrst lokalnih omrežnih mehanizmov, povezanih s tranzitivnostjo. V tej disertaciji sta proučevana predvsem mehanizem izhodnih poti dolžine dva in mehanizem izhodnih skupnih vozlišč. Čeprav lokalni omrežni mehanizem izhodnih poti dolžine dva po zgradbi spominja na tranzitivnost, pa njegov vpliv povzroči nastanek omrežij z asimetričnim središčno-perifernim bločnim modelom, in ne omrežij s kohezivnim bločnim modelom. Slednjega lahko povzroči lokalni omrežni mehanizem izhodnih skupnih vozlišč.

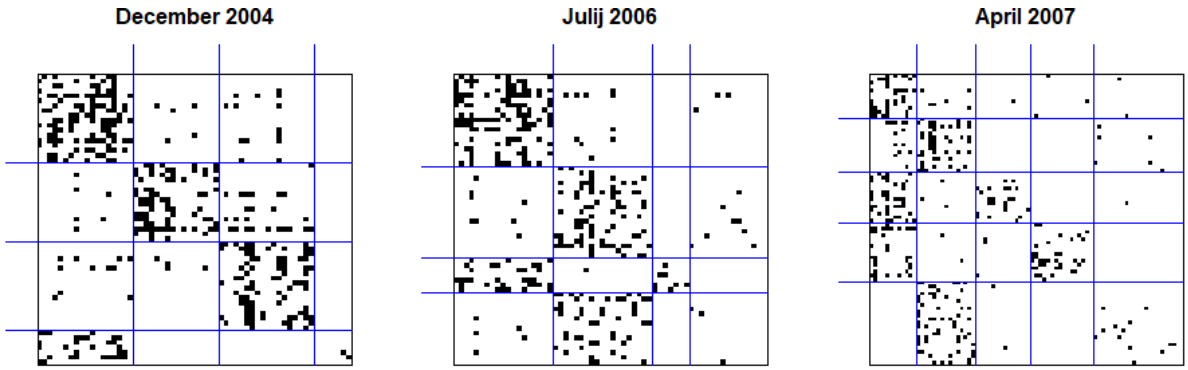
Preostalih vrst bločnih modelov ni mogoče generirati z upoštevanjem posameznih izbranih lokalnih omrežnih mehanizmov. Prav tako, z upoštevanjem vseh izbranih lokalnih omrežnih mehanizmov, ni mogoče generirati omrežij s hierarhičnim ali tranzitivnim bločnim modelom, ki bi vsebovala relativno nizko število napak.

Nastanek hierarhičnega bločnega modela v omrežjih pretoka znanja

Tako kot v prejšnjih primerih je tudi v primeru omrežij pretoka znanja v organizaciji oziroma v podjetju družbeni kontekst upoštevan z namenom izbire najustreznejših vrst bločnih modelov in

pripadajočih lokalnih omrežnih mehanizmov. Povezava v takem omrežju operacionalizira prenos znanja, posamezna vozlišča v omrežju pa pomenijo zaposlene posameznike.

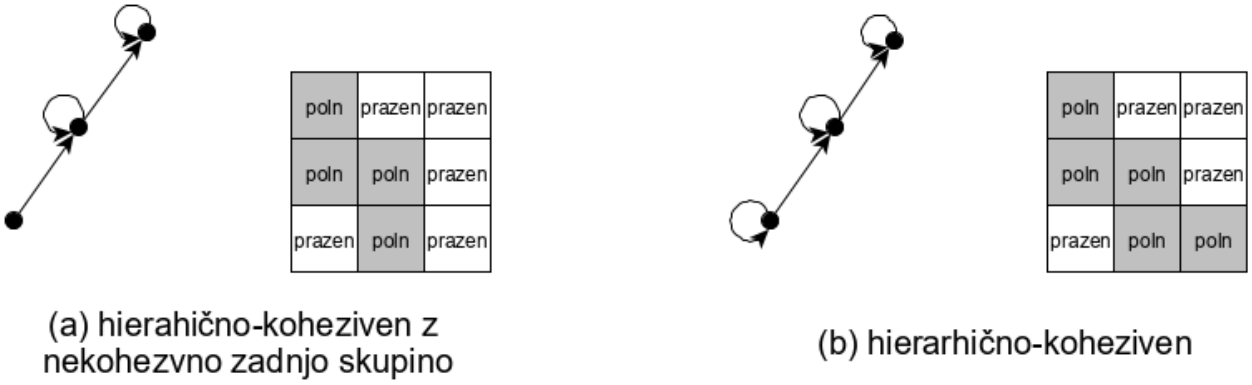
Slika 6 Empirični bločni modeli omrežij prenosa znanja za različna časovna obdobja



Za izbiro vrste bločnega modela in lokalnih omrežnih mehanizmov so bila analizirana omrežja pretoka znanja, ki so bila zbrana v treh časovnih točkah med letoma 2004 in 2007 med zaposlenimi v rastočem podjetju, ki temelji na znanju (Škerlavaj, 2007; Škerlavaj, Dimovski in Desouza, 2010; Škerlavaj, Dimovski, Mrvar in drugi, 2010). Rezultati bločnega modeliranja za redka omrežja (Žiberna, 2013) kažejo na relativno nizko raven napak v dobljenih bločnih modelih empiričnih omrežij. Globalna zgradba omrežja iz zgodnejših časovnih obdobij je kombinacija kohezivnega in središčno-perifernega bločnega modela. Dobljene (kohezivne) skupine v veliki meri določa pripadnost vozlišča k poslovni enoti (angl. *business unit*), središče in periferija dela bločnega modela, ki spominja na središčno-periferni bločni model, pa sta določeni s stažem. Pozneje globalna omrežna zgradba postane bolj podobna kombinaciji dveh središčno-perifernih bločnih modelov in v tretjem časovnem obdobju hierarhičnemu bločnemu modelu. Tudi v tem primeru se izkaže, da so skupine v veliki meri določene s pripadnostjo vozlišč k poslovni enoti, znanje pa se praviloma pretaka od vozlišč z višjim stažem k vozliščem z nižjim stažem. Vozlišča z višjim stažem zavzamejo višjo hierarhično stopnjo kakor vozlišča z nižjim stažem.

Na osnovi rezultatov analize empiričnih omrežij (pa tudi na osnovi proučitve predhodne literature s področja pretoka znanja) sta izbrani dve vrsti bločnih modelov. Prvi je omenjeni hierarhično-koheziven bločni model (Slika 7a), drugi pa je hierarhično-koheziven bločni model z nekohezivno zadnjo skupino (Slika 7b). Slednje pomeni, da so vozlišča iste hierarhične ravni neposredno povezana, razen vozlišča na najnižji hierarhični ravni.

Slika 7 Hierarhično-koheziven bločni model z nekohezivno zadnjo skupino in hierarhično-koheziven bločni model



Lokalni omrežni mehanizmi, ki bi lahko privedli globalno zgradbo omrežja k izbranim vrstam bločnih modelov, so določeni na osnovi pregleda že opravljenih raziskav s področja omrežij nasvetov (angl. *advice networks*), omrežij učenja (angl. *knowledge networks*) ter drugih vrst omrežij, povezanih s prenosom znanja v podjetju ali v organizaciji. Končna množica izbranih lokalnih omrežnih mehanizmov je izbrana na osnovi teorije, ki jo je predstavil Nebus (2006) in v kateri naslavlja proces izmenjave nasvetov med egom (tistim, ki išče nasvet) in alterjem (možni dajalec nasveta) v podjetju oziroma v organizaciji. Nebus (2006) predpostavlja, da ego tehta med stroški pridobivanja nasveta ter med vrednostjo nasveta, ki bi ga dobil od določenega alterja. Na osnovi tega je mogoče lokalne omrežne mehanizme razvrstiti v skupino mehanizmov, povezanih z zaznavo vrednosti nasveta (angl. *value-related mechanisms*) s strani ega, ter med mehanizmov, ki so povezani z zaznavo stroškov pridobivanja nasveta (angl. *cost-related mechanisms*) s strani ega.

V skupino lokalnih omrežnih mehanizmov, povezanih z zaznavo vrednosti nasveta, spadajo naslednji lokalni omrežni mehanizmi, ki so upoštevani v tej raziskavi: popularnost alterja (v tej raziskavi je opredeljena z vhodno stopnjo alterja), hierarhičen položaj alterja (v tej raziskavi je opredeljen s številom vozlišč, ki lahko dosežejo izbranega alterja prek usmerjenih povezav), staž alterja ter število skupnih vozlišč, ki si jih delita ego in alter (mehanizem izhodnih skupnih vozlišč). Med lokalne omrežne mehanizme, ki so povezani z zaznavo stroškov pridobivanja nasveta, pa spadajo: razlika v stažu med egom in alterjem, razdalja med egom in alterjem (v tej disertaciji je opredeljena z geodezično razdaljo), razlika hierarhičnega položaja med egom in alterjem ter mehanizem izhodnih skupnih vozlišč, ki spada v obe skupini mehanizmov. Vsi našteti

lokalni omrežni mehanizmi so operacionalizacije različnih socioloških in psiholoških konstruktov. Na primer, razdalja med egom in alterjem je operacionalizacija egove zaznave psihične razdalje (angl. *psychic distance*), kognitivne razdalje in geografske razdalje, medtem ko je (višja) stopnja popularnosti alterja lahko indikator egove zaznavne pripravljenosti deljenja znanja in ravni kognitivnega zaupanja alterju s strani ega.

Lokalni omrežni mehanizmi so opredeljeni tako, da je pozitiven predznak uteži mehanizmov, ki so povezani z zaznavo vrednosti nasveta, ter negativen predznak uteži mehanizmov, ki so povezani z zaznavo stroškov pridobivanja nasveta, v skladu s teorijo po Nebusu (2006).

Za proučevanje raziskovalnega vprašanja, ki se nanaša na nastanek izbranih hierarhičnih vrst bločnih modelov, z upoštevanjem izbranih lokalnih omrežnih mehanizmov, je uporabljena podobna metodologija kot pri proučevanju nastanka (simetričnega in asimetričnega) središčno-kohezivnega bločnega modela. Poglavitna razlika je v uporabljenem algoritmu za generiranje omrežij, ki upošteva številne dodatne značilnosti rastočih omrežij pretoka znanja v organizaciji oziroma v podjetju. Tako, na primer, upošteva možnost prihoda novih zaposlenih (angl. *newcomers*) in odhoda že zaposlenih (angl. *outgoers*), prekinjanje povezav v omrežju pa ni odvisno od izbranih lokalnih omrežnih mehanizmov, ampak je trajanje povezave časovno omejeno.

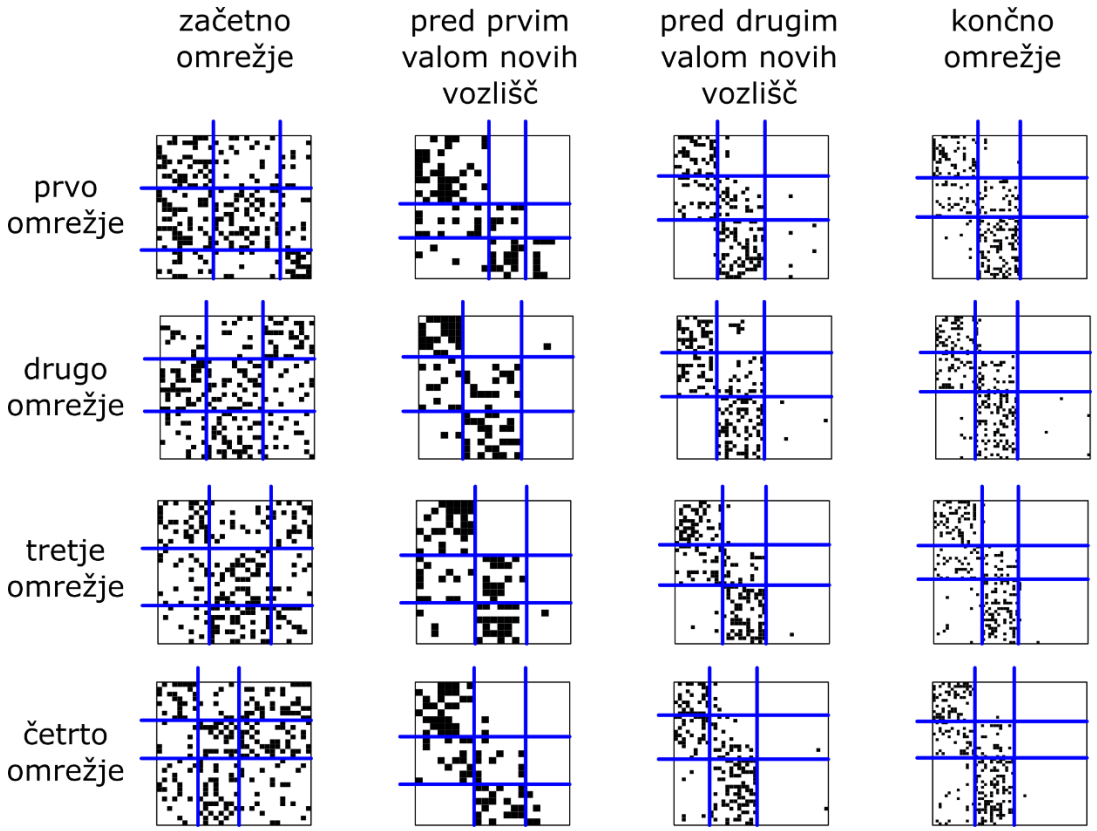
Razlika je tudi v oceni globalnih zgradb generiranih omrežij. Poleg pogoja, da generirano omrežje vsebuje izbrano vrsto hierarhičnega bločnega modela, mora povprečen staž vozlišč po skupinah naraščati z naraščanjem hierarhične ravni (vozlišča z višjim stažem so praviloma višje na hierarhiji). Samo v primeru, ko sta izpolnjena oba pogoja, je mogoče govoriti o prisotnosti izbrane vrste hierarhičnega bločnega modela v generiranem omrežju.

Omrežja so, z uporabo omenjenega algoritma, generirana na različne načine. V nekaterih primerih so upoštevani tako novi zaposleni kakor tudi osip zaposlenih, spet v drugih primerih so upoštevani samo novi zaposleni. Nadalje, predznaki uteži lokalnih omrežnih mehanizmov so v nekaterih primerih prosti, v drugih primerih pa so preddoločeni glede na način interpretacije pripadajočega lokalnega mehanizma. Na primer, pozitiven predznak uteži lokalnega omrežnega mehanizma popularnosti je mogoče interpretirati kot težnjo ega k izbiri alterja z višjo stopnjo popularnosti, medtem ko je lokalni omrežni mehanizem z negativnim predznakom pripadajoče uteži mogoče

interpretirati kot izogibanje ega k izbiri alterja z višjo stopnjo popularnosti. Upoštevanje prostih predznakov uteži povečuje računsko zahtevnost simulacij, a lahko omogoči nekatera nepričakovana spoznanja o razvoju izbranih bločnih modelov.

Rezultati analiz kažejo, da je generiranje omrežij z obema izbranima vrstama bločnih modelov mogoče v primeru, ko so upoštevani vsi našteti lokalni omrežni mehanizmi (Slika 8). Nižja stopnja napak je običajno v omrežjih s ciljnim hierarhično-kohezivnim bločnim modelom z nekohezivno zadnjo skupino.

Slika 8 Nekaj primerov generiranih omrežij s hierarhično-kohezivnim bločnim modelom z nekohezivno zadnjo skupino



Opomba: Upoštevana so tako nova vozlišča kot osip. Predznaki uteži lokalnih omrežnih mehanizmov so preddoločeni. Vsi proučevani lokalni omrežni mehanizmi so upoštevani pri generiranju omrežij. Omrežja so narisana v skladu s pripadajočim bločnim modelom.

Nadaljnji rezultati analiz kažejo, da je mogoče generirati omrežja s hierarhično-kohezivnim bločnim modelom z zadnjo nekohezivno skupino v primerih, ko so upoštevani vsi našteti lokalni omrežni mehanizmi, pa tudi v primeru, ko so upoštevani zgolj tisti lokalni omrežni mehanizmi, ki ne vključujejo staža. V primeru, ko so upoštevani samo novi zaposleni, ne pa tudi osip zaposlenih,

je mogoče generirati omrežja z omenjeno vrsto bločnega modela že zgolj z upoštevanjem mehanizmov, ko so povezani s stažem. V večini primerov generiranih omrežij je zaznati relativno nizko stopnjo napak. Mogoče je sklepati, da so najpomembnejši lokalni omrežni mehanizmi za generiranje omrežij s takim bločnim modelom tisti, ki so povezani s stažem, hierarhičnim položajem vozlišč ter mehanizmom razdalje med egom in alterjem. To je pričakovano, saj so našteti lokalni omrežni mehanizmi operacionalizacije mnogo pomembnih družbenih konstruktov, vključno z zaznavo stroškov pridobivanja nasveta (družbeni stroški, psihološki stroški, institucionalni stroški in drugi) ter zaznavo različnih vrst razdalj med egom in alterjem (psihiatrična razdaja, kognitivna razdalja, geografska razdalja in druge).

Izbrani vrsti bločnih modelov se pojavita tudi v primeru, ko predznaki uteži lokalnih omrežnih mehanizmov niso preddoločeni. Izkaže se, da so predznaki tistih uteži lokalnih omrežnih mehanizmov, ki generirajo omrežja z izbranimi vrstami bločnih modelov popolnoma (v primeru hierarhično-kohezivnega bločnega modela z nekohezivno zadnjo skupino) ali v glavnem (v primeru hierarhično-kohezivnega bločnega modela) v skladu s predpostavljenimi.

Pomembnost raziskave

Gre za prvo raziskavo, ki naslavlja odnos med lokalnimi omrežnimi mehanizmi in nastankom izbranih vrst bločnih modelov brez upoštevanja lastnosti vozlišč. Rezultati kažejo, da lahko določeni lokalni omrežni mehanizmi povzročijo nastanek nekaterih vrst bločnih modelov.

Raziskava je še posebej pomembna, saj prispeva k razumevanju nastanka izbranih vrst bločnih modelov, hkrati pa prinaša pomembna praktična spoznanja. Zelo splošna ugotovitev je, da je z ustreznimi politikami mogoče spodbujati nastanek izbrane globalne zgradbe v realnih omrežjih. Bolj konkretno spoznanje pa je, da na osnovi enega ali zgolj nekaj opazovanj omrežja ni mogoče veljavno sklepati o lokalnih omrežnih mehanizmih, ki so vplivali na nastanek opažene globalne zgradbe omrežja. Med razvojem globalne zgradbe omrežja se namreč lahko pojavijo vmesne globalne zgradbe, raziskovalec pa običajno ne ve, na kateri točki razvojnega procesa globalne zgradbe omrežja je izmeril globalno zgradbo opazovanega omrežja. Zato je v takih primerih globalno zgradbo omrežja nujno opazovati v več časovnih točkah v daljšem obdobju. Tudi v tem primeru ni garancije, da se globalna zgradba omrežja pozneje ne bo spremenila, četudi bodo prisotni isti lokalni omrežni mehanizmi z istimi pripadajočimi utežmi. Uporaba Monte Carlo

simulacij je lahko v veliko pomoč pri razumevanju in napovedovanju razvojnega procesa izbranega bločnega modela ob upoštevanju vpliva izbranih lokalnih omrežnih mehanizmov.

Nadalje, predlagane algoritme NEM je mogoče uporabiti za generiranje omrežij z izbrano globalno zgradbo z določeno ravnjo napak. V nasprotju z drugimi pristopi, ki se pogosto uporabljajo v ta namen, lahko z algoritmi NEM generiramo omrežja z upoštevanjem lokalnih omrežnih mehanizmov. To pomeni, da napake v generiranih omrežjih niso slučajne, temveč so v skladu z izbranimi lokalnimi omrežnimi mehanizmi. Tako generirana omrežja je mogoče uporabiti za različna nadaljnja proučevanja globalnih omrežnih zgradb, lokalnih omrežnih mehanizmov ali modelov za ocenjevanje globalne omrežne zgradbe.

V tej disertaciji je predlagana metodologija za proučevanje odnosa med lokalnimi omrežnimi mehanizmi in različnimi vrstami globalnih omrežnih zgradb. Metodologija vključuje generiranje omrežij z upoštevanjem lokalnih omrežnih mehanizmov, pa tudi pristop za evalvacijo dobljenih globalnih omrežnih zgradb. Prav tako je narejenih nekaj pomembnih korakov pri razvoju normalizirane kriterijske funkcije, poimenovane funkcija relativnega prileganja, katere vrednost je primerljiva med dvema bločnima modeloma in ki se lahko uporabi tudi za določitev števila skupin ter za izbiro najustreznejše vrste bločnega modela pri preddoločenem bločnem modeliranju.

Poglavitne pomanjkljivosti raziskave

Čeprav so lokalni omrežni mehanizmi pogosto proučevani v empiričnih študijah družbenih omrežij, do zdaj še ni bilo opravljene raziskave, ki bi sistematično obravnavala odnos med lokalnimi omrežnimi mehanizmi in globalnimi omrežnimi zgradbami, opredeljenimi z bločnimi modeli. Eden izmed razlogov je dejstvo, da je izvedba takih študij lahko zelo zahtevna brez upoštevanja izbranega družbenega konteksta, saj je število vseh možnih lokalnih omrežnih mehanizmov in globalnih omrežnih zgradb zelo veliko.

Tako pričujoča disertacija je majhen korak k boljšemu razumevanju odnosa med lokalnimi omrežnimi mehanizmi in globalnimi omrežnimi zgradbami, ki so opredeljene skozi izbrane vrste bločnih modelov. Zaradi izjemno širokega raziskovalnega vprašanja so vpeljane določene predpostavke in omejitve.

Omejitve, povezane z lokalnimi omrežnimi mehanizmi

Ena izmed pomembnejših predpostavk je ta, da lokalni omrežni mehanizmi vplivajo na vedenje vozlišč (vzpostavljanje povezav, prekinjanje povezav ali ohranjanje statusa *quo*) in prek tega na pojav določene globalne zgradbe omrežja, pa tudi predpostavka, da globalna zgradba omrežja neposredno ne vpliva na pomembnost (moč) posameznih lokalnih omrežnih mehanizmov. S tem je povezana tudi predpostavka, da je pomembnost lokalnih omrežnih mehanizmov enaka za vsa vozlišča v omrežju, ne glede na to, v katerem delu omrežja so. Obe predpostavki sta redko neposredno obravnavani pri analizi empiričnih omrežij z uporabo SAOM ali ERGM. V primeru slednjih je mogoče oceniti pomembnost lokalnih mehanizmov posebej za različne skupine vozlišč, kjer so skupine vozlišč določene na osnovi pripadajočih lastnosti vozlišč. Ti lahko predstavljajo različne osebne značilnosti vozlišč ali njihov položaj v omrežju (na primer upoštevanje dejstva ali vozlišče v omrežju pripada centru ali periferiji). Block (2015) je na osnovi metaanaliz omrežij prijateljstev pokazal, da vozlišča, ki pripadajo tranzitivnim triadam, bolj verjetno ohranjajo recipročne povezave, saj tovrstne triade omogočajo družbene interakcije, ki se sicer ne bi pojavile. Za primer omrežij interakcij med predšolskimi otroki pa Schaefer in drugi (2010) ugotavljajo, da se manj kompleksni lokalni omrežni mehanizmi pogosteje pojavljajo med mlajšimi otroki, medtem ko se bolj kompleksni lokalni omrežni mehanizmi pojavljajo med starejšimi otroki, kar je skladno z njihovim psihološkim razvojem.

Poleg upoštevanja različnih uteži lokalnih omrežnih mehanizmov za različna vozlišča bi bilo smiselno upoštevati tudi spreminjanje pomembnosti lokalnih omrežnih mehanizmov v času. Na primer, Schaefer in drugi (2010) ugotavljajo, da pomembnost lokalnega omrežnega mehanizma popularnost in mehanizma zapiranja triad (ang. *triadic closure*) v omrežjih interakcij med predšolskimi otroki narašča skozi šolsko leto. Za upoštevanje spreminjanja moči lokalnih omrežnih mehanizmov v tej disertaciji bi bilo treba razviti pristop za normalizacijo uteži lokalnih omrežnih mehanizmov, tako da bi bile te primerljive v času. To je zelo zahteven problem. Tako SAOM kakor tudi (časovni) ERGM, ki sta verjetno najpogosteje uporabljena modela za proučevanje dinamike v omrežjih, predpostavljata konstantne ravni pomembnosti mehanizmov v času.

Posamezna vozlišča imajo pri uporabi predstavljenih algoritmov NEM enake verjetnosti za to, da bodo dobila priložnost za spremembo statusa povezave. Kljub temu pa so algoritmi NEM

definirani tako, da lahko vozlišče v določeni iteraciji ne spremeni nobene povezave. Kljub temu bi bilo koristno upoštevati primer, ko bi imela določena vozlišča več priložnosti za spremembo povezav kakor druga. Na primer, v primeru omrežij dajanja nasvetov v organizaciji je upravičeno domnevati, da zaposleni z nižjim stažem pogosteje sprašujejo za nasvet kakor zaposleni z višjim stažem. V primeru omrežij interakcij med predšolskimi otroki pa tisti z višjim številom vzajemnih povezav manj pogosto vzpostavljajo nove povezave kakor oni z nižjim številom vzajemnih povezav (Daniel in drugi, 2019). Takšne omejitve je mogoče upoštevati znotraj modela (kot del definicije algoritma NEM) ali pa prek vključitve ustreznih lokalnih omrežnih mehanizmov.

Omejitve, povezane z bločnimi modeli

Ena izmed osnovnih omejitev, ki je povezana z izbranimi bločnimi modeli, je omejitev na bločne modele s tremi skupinami oziroma z dvema skupinama. Vprašanje, ki ni v celoti odgovorjeno, se nanaša na to, v kolikšni meri je mogoče dobljene rezultate posplošiti na bločne modele z več kot tremi skupinami. Eno izmed pomembnejših vprašanj v tem pogledu je, ali bi bilo smiselno (v primeru več skupin) prilagoditi uteži izbranih lokalnih omrežnih mehanizmov ali pa dodati lokalne omrežne mehanizme.

Medtem ko je mnogo zgodnejših raziskovalcev s področja analize omrežij (Cartwright in Harary, 1956; Moreno, 1934) obravnavalo tako pozitivne kakor tudi negativne povezave, so se raziskovalci iz poznejših obdobij v glavnem osredotočili na analizo binarnih omrežij, natančneje omrežij z opazovanimi zgolj pozitivnimi povezavami. To je morda povezano z načinom zbiranja podatkov, ki je v primeru binarnih omrežij za respondente manj kognitivno zahtevno. Kljub temu pa nekateri raziskovalci današnjega časa opozarjajo, da je upoštevanje negativnih povezav ključno pri proučevanju razvoja globalnih omrežnih zgradb. Stadtfeld in drugi (2019) poudarjajo, da zgolj prisotnost pozitivnih povezav v omrežju ni zadostno za pojasnitev nastanka skupin v omrežju, Doreian in Mrvar (2014) sta pokazala, da upoštevanje samo pozitivnih povezav pri proučevanju empiričnih omrežij lahko pripelje do lažnih rezultatov. Razširitev pričujoče študije na raven omrežij s pozitivnimi in negativnimi povezavami (označenih omrežij; angl. *signed networks*) bi lahko dodatno osvetlila tudi nekatere praktične vidike zbiranja tovrstnih omrežij. V primeru omrežij, zbranimi med predšolskimi otroki, je upoštevanje negativnih povezav zelo pomembno, saj se lahko uporablja kot operacionalizacija dodatnih družbenih pojavov, kakršen je medvrstniško nasilje (angl. *bullying*). Medvrstniško nasilje je opredeljeno kot ponavljajoče se in namerno

negativno vedenje storilca do žrtve, ki se težko brani (Olweus, 1994). Eden izmed namenov medvrstniškega nasilja je povečati ali vzdrževati družbeni položaj (Caravita in drugi, 2009; de Bruyn in drugi, 2010; Salmivalli in drugi, 1996). Kljub temu pa obstajajo razlike v načinih odzivov do storilcev, glede na različne starostne skupine otrok. Mlajši otroci navadno sankcionirajo medvrstniško nasilje, medtem ko ga starejši nagradijo s prepoznavanjem višjega družbenega položaja (Van der Ploeg, Steglich in Veenstra, v tisku).

Priporočila za nadaljnja raziskovanja

Nekaj pomembnih priporočil za nadaljnje raziskovanje je že predstavljenih v prejšnjih poglavjih, skupaj z razpravo o omejitvah raziskave. Spodaj sledijo dodatna priporočila za prihodnje raziskave, ki niso neposredno povezana z omejitvami te študije.

Eden izmed pomembnih metodoloških vprašanj, ki so bila naslovljena v disertaciji, je način ocene prileganja danega omrežja k izbrani vrsti bločnega modela. Zelo splošno oceno prileganja omogoča izračun števila neskladnih blokov, za oceno ravni napak v empiričnem omrežju pa je mogoče uporabiti že omenjeno funkcijo RF, ki lahko zavzame vrednosti do 1 (v primeru omrežja brez napak), pričakovana vrednost v primeru slučajnega omrežja pa je 0. Vrednost funkcije RF je dobljena s pomočjo mnogo randomiziranih omrežij. Njena uporaba je smiselna, kadar ni omejitev pri vzpostavljanju in prekinjanju povezav, ki bi onemogočale nastanek omrežja z izbranim bločnim modelom brez napak. Primer take omejitve je omejitev navedbe števila zaposlenih, do katerih bi se obrnili po nasvet, ali pa naravna omejitev na število možnih vzdrževanih prijateljskih vezi (Dunbar, 1992). V takšnih primerih so opazovana omrežja navadno redka in tako je pogosto smiselna uporaba bločnega modeliranja za redka omrežja (Žiberna, 2013), kjer so napake v polnih in napake v praznih blokih različno utežene, polni bloki pa so navadno redkejši, kakor v primeru bločnega modeliranja, ko so napake v praznih in polnih blokih enako utežene. V takih primerih vrednost funkcije RF pogosto ne more doseči vrednosti 1, kar pomeni, da je treba dodatno oceniti zgornjo mejo, ki jo lahko zavzame. Kljub navedenemu so potrebne dodatne analize lastnosti funkcije RF, upoštevajoč število in velikosti skupin, dejstvo, ali je mera dobljena na osnovi preddoločenega bločnega modeliranja ali ne, ter z upoštevanjem različnih vrst enakovrednosti.

V tej disertaciji je predstavljenih več algoritmov za generiranje omrežij z upoštevanjem izbranih lokalnih omrežnih mehanizmov. Vsi spadajo v družino NEM. Pokazano je, da je pogosto

nemogoče oceniti parametre SAOM in ERGM v primeru, ko ima omrežje določen bločni model z nizko stopnjo napak. To velja tudi za primere, ko so omrežja z izbranimi bločnimi modeli generirana na osnovi lokalnih omrežnih mehanizmov, ki so vključeni v SAOM oziroma ERGM. Razlog je lahko v izrazito nenaravni porazdelitvi stopenj v omrežju. V realnih omrežjih se take nenaravne porazdelitve stopenj pojavijo kot posledica omejitev pri vzpostavljanju in prekinjanju povezav v omrežju. To nakazuje na potrebo po nadaljnjem razvoju obstoječih modelov (na primer SAOM in ERGM) na način, ki bi pri oceni parametrov modela upošteval bločni model omrežja, skupaj z obstoječimi lokalnimi omrežnimi mehanizmi. Druga možnost je razvoj novih statističnih modelov, ki bi temeljili na predlaganem algoritmu NEM in bi omogočali generiranje omrežij z izbrano vrsto bločnega modela ter oceno parametrov modela, ki bi temeljila na empiričnih omrežjih z izbranim bločnim modelom.

Prilagojeno metodologijo, razvito v tej disertaciji, je mogoče uporabiti za zasnovo tipologije izbranih lokalnih omrežnih mehanizmov, ki povzročijo nastanek izbranih vrst bločnih modelov oziroma spremembe iz ene v drugo vrsto bločnega modela. Takšna klasifikacija bi bila okvir za proučevanje empiričnih omrežij in za generiranje omrežij z dano vrsto bločnega modela, kar je pomembno pri proučevanju in preverjanju algoritmov za bločno modeliranje dinamičnih omrežij (angl. *dynamic blockmodeling algorithms*) (Matias in Miele, 2015; Xing in drugi, 2010; Xu in Hero, 2014).