RESEARCH ARTICLE

Symmetric core-cohesive blockmodel in preschool children's interaction networks

Marjan Cugmas[1]*, **Dawn DeLay**[2], **Aleš Žiberna**[1], **Anuška Ferligoj**[1,3]

1 Centre for Methodology and Informatics, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia, **2** Sanford School of Social and Family Dynamics, Arizona State University, Tempe, Arizona, United States of America, **3** International Laboratory for Applied Network Research, National Research University Higher School of Economics, Moscow, Russia

* marjan.cugmas@fdv.uni-lj.si

Abstract

Researchers have extensively studied the social mechanisms that drive the formation of networks observed among preschool children. However, less attention has been given to global network structures in terms of blockmodels. A blockmodel is a network where the nodes are groups of equivalent units (according to links to others) from a studied network. It is already shown that mutuality, popularity, assortativity, and different types of transitivity mechanisms can lead the global network structure to the proposed asymmetric core-cohesive blockmodel. Yet, they did not provide any evidence that such a global network structure actually appears in any empirical data. In this paper, the symmetric version of the core-cohesive blockmodel type is proposed. This blockmodel type consists of three or more groups of units. The units from each group are internally well linked to each other while those from different groups are not linked to each other. This is true for all groups, except one in which the units have mutual links to all other units in the network. In this study, it is shown that the proposed blockmodel type appears in empirical interactional networks collected among preschool children. Monte Carlo simulations confirm that the most often studied social network mechanisms can lead the global network structure to the proposed symmetric blockmodel type. The units' attributes are not considered in this study.

Introduction

One of the key attempts in sociology, and also in psychology, is to reveal the (social) mechanisms that are responsible for a given (social) output. When the relationships among individuals are studied, the social output is a social network. In social network analysis, there are different approaches to study the underlying social mechanisms of 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 objective of the current study is to identify fundamental social mechanisms that guide the formation of a global network structure.

In this study, the global network structure is narrowed to a structure with three or more groups. The units from the first group (called the core group) have symmetric links established with all units in the network, while the units from the other groups (called cohesive groups) are internally well linked. The units from different cohesive groups are not linked to each other. This global network structure (called symmetric core-cohesive blockmodel, described in more detail in subsection Global network structure) is proposed since it is a combination of cohesive and symmetric core-periphery global network structures and because these global network structures can arise from the well-known transitivity [1, 2] and popularity [2] mechanisms. These two mechanisms were found to be present in the formation of many liking and friendship networks collected among preschoolers (see subsection Local mechanisms).

The assumption made in this study is that the proposed global network structure appears among preschool children. Entrance to preschool brings a set of peers together who were previously unknown to one another. This is rare in the natural world and, thus, the shift into preschool peer groups offers a unique opportunity to assess and understand the mechanisms behind peer group formation. Preschool entry is also distinct from other social network settings in that it offers a closed network space in which peers interact. Preschool also provides a unique developmental context in which children are motivated, perhaps for the first time, to form new and enduring social relationships with similar-age peers [3, 4].

This assumption (of the emergence of the proposed global network structure) is tested by using the blockmodeling approach [5] on the symmetrized networks previously analyzed by Schaefer et al. [6]. Their study's main focus was the network dynamics rather than the global network structure. They showed that the selected local network mechanisms are important in such networks and that the importance of different local network mechanisms change throughout the school year.

Building on the assumption (tested in this paper) that the proposed global network structure emerges in interactional preschool networks, the following research question is posed: **can the proposed global network structure appear due to the selected local network mechanisms without considering the nodes' attributes?** Here, the same local network mechanisms are assumed as in the study of Schaefer et al. [6] and other previous studies on preschool network dynamics. The research question is addressed using Monte Carlo simulations, specifically, by applying the proposed model from the family of network evolution models.

The study is relevant since understanding of the local network mechanisms at play, in the context of global network structures, is important while studying real (empirically observed) networks. Namely, the proposed global network structure's emergence at preschools raises very important developmental questions, e.g., how children in the core group differ from children in cohesive groups and what are the implications (if any) for their further individual development? Should such a global network structure be encouraged or discouraged? Is this a period where scholars may be able to document the emergence of social cliques and associated social norms? Will some children be integrated into cohesive groups, while others are left with minimal peer affiliations in the global network [7]?

The paper is organized as follows: a new global network structure is formally defined and the local network mechanisms are proposed and described (section Global network structure). Next, the global network structures of the empirical interactional preschool networks are analyzed (section The empirical case). The main research question concerned with the proposed global network structure's emergence is addressed in section Simulation approach and some conclusions are outlined.

Global network structure and local network mechanisms through structural processes

In the following section, the proposed global network structure is defined in the blockmodel context. Different local network mechanisms that may drive the global network structure of preschool children's interactional networks towards the proposed one are discussed.

Global network structure

A blockmodel is a network in which the units are groups of equivalent units from a studied network [5]. The term reflects the fact that if a network is represented by a matrix, which is then split according to a partition (groups), blocks (submatrices) are formed in the matrix. The term "block" refers to a submatrix showing the links among units from the same or different group(s). Two selected units are structurally equivalent if they have the same pattern of links to the other units [8, 9]. The possible block types are identified through a selected definition of equivalence which is based on links among the units. Structural equivalence [8] and its generalization, regular equivalence [10], are the most common. When structural equivalence is used, only null and complete blocks are possible. In ideal complete blocks, all possible links are present while no link exists in ideal null blocks.

A demonstration of blockmodeling according to structural equivalence is given in Fig 1. The original network is visualized in matrix form in Fig 1A. Here, each row and column represents a unit. Gray colored cells in a matrix represent a link from the i-th unit (row) to the j-th unit (column). Cells on the diagonal represent loops (a given unit is linked to itself). The units are permuted (see Fig 1B) in such a way that those with the same pattern of links are placed together and form a cluster (group). Two groups are shown in Fig 1B.

In the blockmodeling context, the clusters of units are shrinked into nodes. The blockmodel that is obtained is visualized in Fig 1C. The obtained blockmodel has two nodes (shrinked groups). Here, two types of blocks appear (complete and null). Complete blocks are on the diagonal of the matrix because the units from both groups are internally linked to each other. Off-diagonal blocks refer to the relationships between different groups. Since the units from different groups 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 empirical networks. In such networks, there are usually some non-links in complete blocks and some links in null blocks (see Fig 1D). Such links are called errors or inconsistencies.
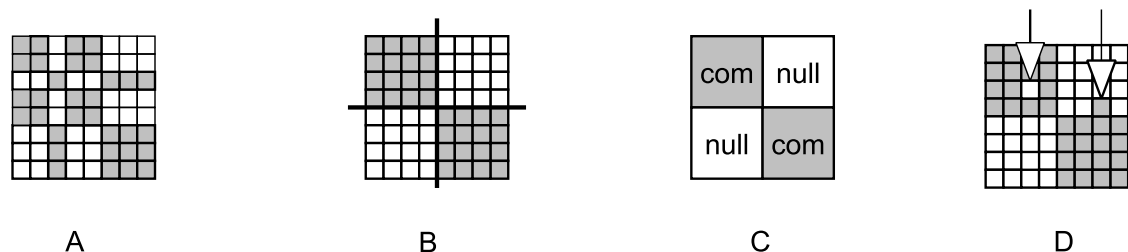

Fig 1. Example of an empirical network and its blockmodeling solution. (A) empirical network, (B) empirical network drawn in line with the blockmodeling solution, (C) blockmodel, (D) blockmodeling solution with two inconsistencies.

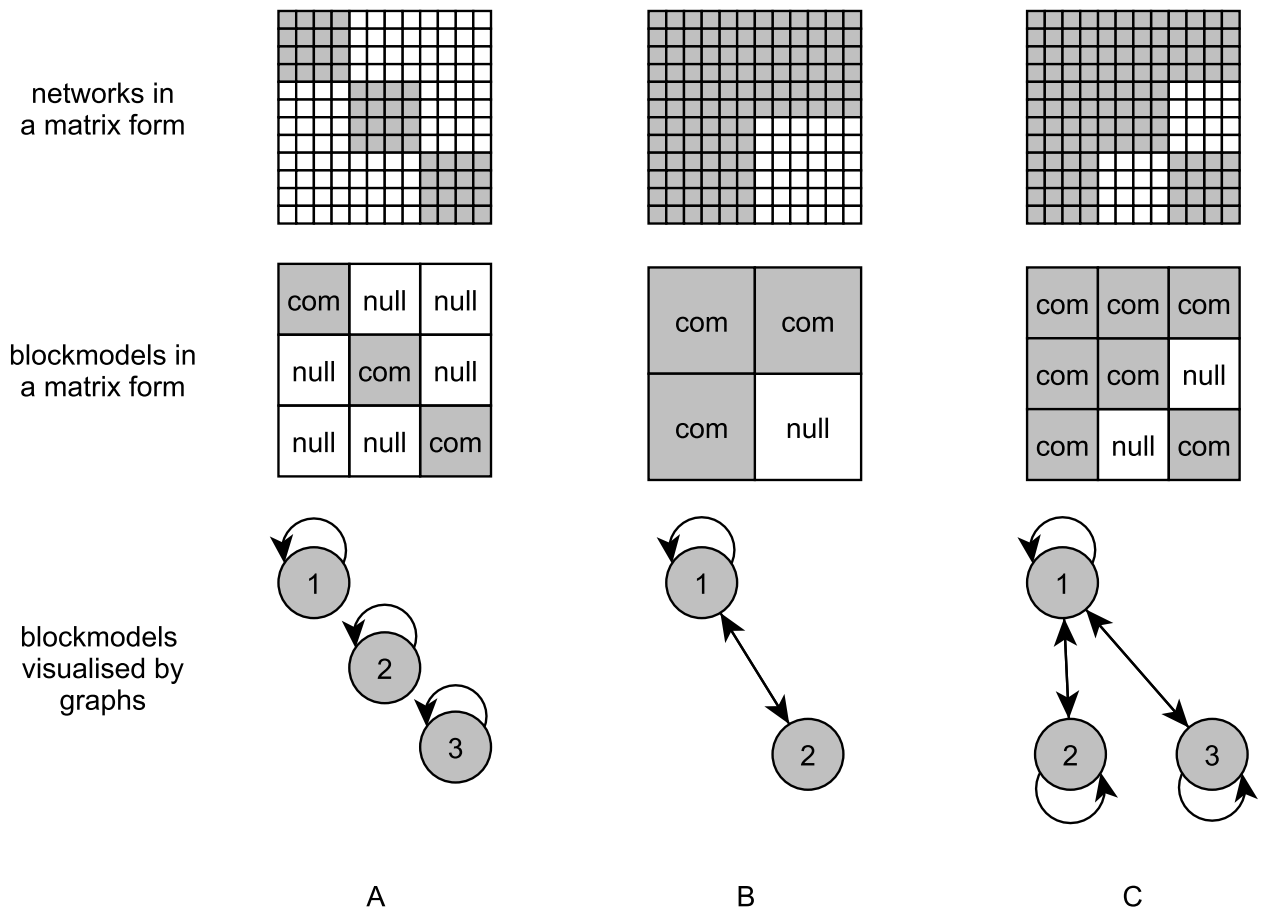https://doi.org/10.1371/journal.pone.0226801.g001

Fig 2. Different representations of networks with a cohesive blockmodel, symmetric core-periphery blockmodel, and symmetric core-cohesive blockmodel. (A) cohesive blockmodel, (B) symmetric core-periphery blockmodel, (C) symmetric core-cohesive blockmodel.

There are several well-known blockmodel types, with two being the cohesive and (symmetric or asymmetric) core-periphery blockmodel types. The cohesive blockmodel type (Fig 2A) contains at least two groups of units where units from different groups are not linked to each other, while all units inside each cohesive group are linked to each other. On the other hand, the symmetric core-periphery blockmodel (Fig 2B) is defined by two groups of units. The units from the core group are internally well linked to each other and units from the periphery are not linked to each other. The units from the core are also linked to the units from the periphery and vice versa (in the asymmetric case, the units from the periphery are linked to the core ones or vice versa).

The newly proposed symmetric core-cohesive blockmodel type (Fig 2C) is seen as a combination of a cohesive blockmodel and a symmetric core-periphery blockmodel. A symmetric core-cohesive blockmodel consists of one core group of units to which all units in the network are linked, and where units from the core group are linked to all other units in the network. The other units are classified into cohesive groups. Units from each cohesive group are internally linked to each other, while units from different cohesive groups are not linked to each other. The model can be extended in such a way that a group of units which are not linked to each other would also exist.

Local mechanisms

It has been hypothesized that the asymmetric core-cohesive blockmodel type might appear in networks observed among preschool children where the links are defined by "friend nominations" or "liking" [2]. Although research is lacking on the global network structure's evolution in the blockmodel context, many studies address the mechanisms that affect the creation and dissolution of ties. The social mechanisms of attraction most often discussed are mutuality (also known as reciprocity), popularity (also known as the Matthew effect or preferential attachment), transitivity, and assortativity (also known as assortative mixing or homophily) [11]. The last one may be considered through the assortativity of in-degree or other units' attributes, such as gender [12, 13]. Simulations confirm that an asymmetric core-cohesive blockmodel can appear as a result of the listed mechanisms.

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 analyzed in such settings are often observational. In such studies, a link is often operationalized as an interaction and therefore the observed links are undirected. If such interactions are considered as an indicator of friendship, popularity or liking, the same mechanisms must be considered 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 through the reciprocation of ties and is one of the most fundamental social network mechanisms (besides creating links) and a basic feature of social life [14]. Analyzing 49- to 62-month-old preschool children, Snyder et al. [15] not only found that children spend much time with selected friends and less with others, but also strong evidence of mutuality. Observed mutual links in the empirical global network structures can also emerge since children prefer to interact with peers who are similar to themselves. This tendency often fosters the emergence of mutual peer relationships during childhood [6, 16–18].
  The researcher cannot indirectly study this mutuality when analyzing non-directed interactional empirical networks. However, the mechanism can play a role in the process of creating the initiative for interactions (also see subsection The algorithm for generating networks).

- **Popularity** is defined through an in-degree in social network analysis and is usually an operationalization of likeability or social status [14]. As a social network mechanism, popularity expresses the tendency to create links to others with a relatively high (in)degree. This is especially the case for less popular ones who wish to increase their own popularity by creating links with those who are most popular [19]. The fact that some units become more popular than others can relate to their personal attributes (e.g., wealth, being good at something, etc.) or positive or negative behavior [20].

- **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 [6].

Many empirical studies highlight the importance of these mechanisms. For example, Snyder et al. [15] noticed that children spend considerable time with selected friends and less with others. They also observed a strong mutual affiliation of friendships, which is subjected to the level of *positive social consequences available from peers in the classroom.*

Daniel et al. [14] used ERGM [21] 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.

Schaefer et al. [6] 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 [22–24]. 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 over time, which is expected since very early on friendships are typically play-oriented dyads that primarily socialize children into group life [25]. When children gain more social contacts and greater confidence, they move into larger groups [26].

The empirical case

The hypothesis about the symmetric core-cohesive blockmodel being present in empirical interactional networks is tested in the subsections below. To this end, the empirical data collected among preschool children are analyzed using generalized blockmodeling. The research was approved by a University Internal Review Board at Arizona State University.

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 (the active consent to participate in the study was obtained from parents or guardians of children included in the study). The data were also analyzed in the study by Schaefer et al. [6]. 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 in different activities, e.g. free play, talking, aggressive behavior, 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 analyzed 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 are generated for each class. Each network's construction is based on a two-month period, as presented in Table 1. The networks are in matrix form in which each row and each column represents a child. The number of a given child's (ego, in a given row) observed interactions with other children (alter, in a given column) is shown in the corresponding cells of the matrix. The obtained networks were transformed from directed to undirected (data in S1 Dataset) and binarized (data in S2 Dataset): 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) 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%.

Methodology

Binarized networks are blockmodeled to evaluate the global network structure. Blockmodeling is a way of reducing a large, potentially incoherent network to a smaller, comprehensible, and

Table 1. Some basic descriptive statistics for the undirected networks.

CLASS ID	No. of units				No. of observations				Age span in the last period (in months)	% of males in the last period
	TP 1	TP 2	TP 3	TP 4	TP 1	TP 2	TP 3	TP 4		
	Sep-Oct 2004	Nov-Dec 2004	Feb-Mar 2005	Apr-May 2005	Sep-Oct 2004	Nov-Dec 2004	Feb-Mar 2005	Apr-May 2005		
1	21	20	20	19	814	510	484	321	42-58	63
2	17	17	15	14	57	95	236	374	48-59	50
3	16	17	14	14	75	200	190	184	50-58	50
4	17	18	18	16	104	410	525	548	49-55	69
5	17	17	14	14	280	406	862	413	37-57	50
6	15	15	14	14	202	343	1005	510	46-59	43
	Sep-Oct 2005	Nov-Dec 2005	Feb-Mar 2006	Apr-May 2006	Sep-Oct 2005	Nov-Dec 2005	Feb-Mar 2006	Apr-May 2006		
7	21	19	17	16	594	564	196	589	46-60	44
8	18	18	16	16	396	432	273	855	43-58	69
9	18	18	16	15	663	406	368	1237	37-59	40
10	16	15	15	14	931	496	309	1609	39-60	64
11	15	16	15	15	172	241	395	574	48-60	47

interpretable structure [5]. In a blockmodeling procedure, a list of allowed and forbidden block types is given. Since structural equivalence is used, these block types are null and complete. In order to not constrain the blockmodeling procedure, the relationships between groups (image matrix) is not pre-specified.

The blockmodeling was done using the "blockmodeling" package [27] for the R programming language. The number of iterations in the blockmodeling was 500 and 3 clusters were set for all networks.

Results: Empirical blockmodels

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

Some of the blockmodels obtained 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 group (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 towards the symmetric core-cohesive is addressed in the next section. Attributes of the units are not considered in this study.
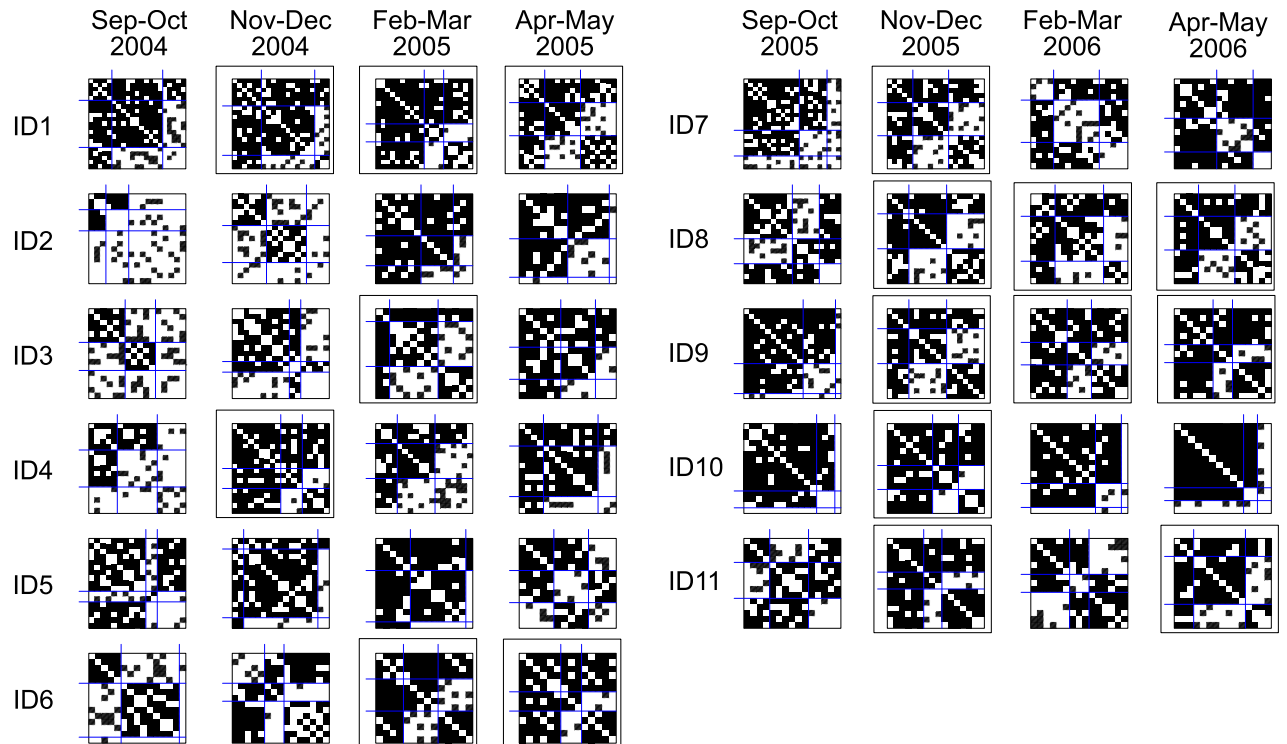
Fig 3. Obtained blockmodel structures for each class (ID1 to ID11) and each time period. Undirected and binarized empirical networks are considered. The obtained symmetric core-cohesive blockmodels are presented in the frame matrices.

https://doi.org/10.1371/journal.pone.0226801.g003

It should be noted that simulations can never prove that certain mechanisms cause global structure in empirical networks, only that they could cause it.

Simulation approach

Cugmas et al. [2] have already shown that the mutuality, popularity, assortativity, and outgoing-two-path mechanisms can lead towards the asymmetric core-cohesive blockmodel type and that different combinations of the local network mechanisms lead to this global network structure. The results were different when only one of the mechanisms was considered. For example, when only the popularity mechanism was considered, the resulting blockmodel was an asymmetric core-periphery blockmodel, while, on the other hand, the transitivity mechanism plays a role while forming the cohesive groups. Since the mechanisms are not independent, the role of the assortativity and mutuality mechanisms when considered together with popularity and transitivity in a blockmodel context is unclear and depends on the strengths of the other mechanisms.

In this paper, the *symmetric* interactional networks of preschool children are studied. Therefore, the simulation approach proposed by Cugmas et al. [2] for asymmetric networks is adapted to the symmetric case.

To evaluate whether the selected local network mechanisms can lead the global network structure towards the symmetric core-cohesive blockmodel, the adapted algorithm for generating networks is presented in the next subsection, and followed by definitions of the selected local network mechanisms. Many networks are generated using the proposed algorithm. In

this generating process, different strengths of the mechanisms are considered. The global network structures of the generated networks are then evaluated in the Results subsection by applying the concepts of inconsistent blocks and relative fit value, which are also described in subsection Simulation design.

The algorithm for generating networks

A symmetric core-cohesive blockmodel may be generated in several ways by considering different local mechanisms. Three approaches are identified with regard to whether symmetric or asymmetric links are generated and considered in the symmetrization procedure.

- **Immediate any-link symmetrization approach**: here symmetric links are formed or dissolved. A link between the ego and the alter is formed or dissolved based only on the ego's preferences. An alter can change this link at subsequent iterations (when he is in a position to form or dissolve a link, i.e. the alter is the ego). This means a symmetric link exists if at least one of the actors chooses that link, and does not exist if at least one actor does not want it. The reciprocity mechanism is not considered in this approach.

- Asymmetric links are generated in the case of the second and third approaches. Here, only asymmetric links are formed at one time. The generated networks are symmetrized before being analyzed since it is assumed that an observer does not know who initialized an interaction (interaction is seen as symmetric by an observer). In this approach to generating networks, it is also considered that the social process of interactions is an asymmetric process. To achieve symmetric networks, one of the following approaches can be used:

  - **reciprocal-link symmetrization approach**: the generated networks are symmetrized in such a way that a symmetric link exists only if both actors choose the link. This approach implies that an interaction exists only if both actors perceive each other as a desirable interaction partner. This is inconsistent with the assumption that an interaction can be established without the alter's consent.

  - **any-link symmetrization approach**: the generated networks are symmetrized in such a way that a symmetric link exists if at least one actor chooses the link, and does not exist if neither actor wants it. Considering asymmetric and symmetric links implies the assumption that an interaction is independent of the alter.

The approach where the links or non-links are reciprocated or dissolved independently of the alter (the immediate any-link symmetrization approach) is, to the best of the author's knowledge, rarely used in practice. However, it can be used when both the observer (e.g., researcher) and other members of the network (other units, except the ego and the alter) do not know who initiated an observed interaction. On the opposite, the logic behind the approach where a (symmetric) link exists between the ego and the alter if both choose it and it does not otherwise exist (reciprocal-link symmetrization approach) is used more often in applied studies; for example, to study different types of collaborations between organizations [28, 29]. Similar and also some other possible approaches to generating networks also exist within Stochastic Actor Oriented Models [30].

In this study, the observed interactional networks are symmetric by the definition of "interaction", although the process which 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 where an interaction is actively rejected by the alter, it can still be observed, although it is more likely to be recorded if it is either accepted or tolerated. Therefore, the approach where asymmetric ties are formed (by

considering the mutuality mechanism) and the network is symmetrized in such a way that all links are symmetrizied, before being further analyzed, is the closest representation of the emergence of empirical networks.

Networks are represented in the form of an adjacency matrix X of size $n * n$ where n is the number of units. The possible values are 1s and 0s where 1s represent links, while 0s represent non-links. Because loops are not present, the diagonal values are 0. The proposed algorithm (see Algorithm 1) comes from the family of network evolution models (NEM) [31] and can take initial networks with different blockmodels.

Algorithm 1 The algorithm for generating networks used in this study

```
Require: initial network X
Require: vector of strengths of the mechanisms θ
Require: probability of establishing a link q
Require: number of iterations k
1: for l in 1: k do
2:    randomly select unit i
3:    calculate network statistics according to the selected mechanisms
for unit i and all other units and save it in S
4:    calculate φ = Sθ^T
5:    if φ ≥ Q_3(φ), classify unit j into set C, where Q_3 is 3rd quartile
6:    if φ ≤ Q_1(φ), classify unit j into set F, where Q_1 is 1st quartile
7:    with probability q set i → j where j is randomly selected from
set C
8:    with probability 1 − q set i ↛ j where j is randomly selected from
set F
9: end for
10: return generated network X
```

The algorithm is iterational where the number of iterations k can be determined based on the desired number of changes in the global network structure. Further, parameter q must be set. It reflects the tendency towards the creation of a link and can be estimated based on the density of the network with the expected blockmodel. Yet, there is no guarantee the generated networks' density will equal q since it depends on several factors, including the selected local mechanisms.

In the iterational process, a unit i is randomly selected with probability $\frac{1}{n}$. Then, the network statistics S are calculated based on the operationalized selected mechanisms (see the next subsection). These network statistics are weighted by the vector of strengths of local mechanisms θ producing vector $\phi = S\theta^T$. These units, for which it holds that their corresponding weighted network statistic is higher than or equal to the third quartile of all weighted network statistics, are classified in the set C and are the candidates to accept the incoming tie from unit i. The other units, for which it holds that their corresponding weighted network statistic is lower than or equal to the first quartile of all weighted network statistics, are classified in the set F and are candidates for being dissolved of an incoming tie by unit i. With probability q, the link from i to randomly selected j from set C is set and with probability $1 − q$ a non-link from i to randomly selected j from set F is set. Since the unit can establish a link that already exists or dissolve a link that does not exist, there could be no visible change of a link upon a given iteration.

Formal definitions of the mechanisms

The mechanisms are operationalized by different network statistics defined on a binary network, and normalized so that the minimum corresponding values are 0 and the maximum values are 1.
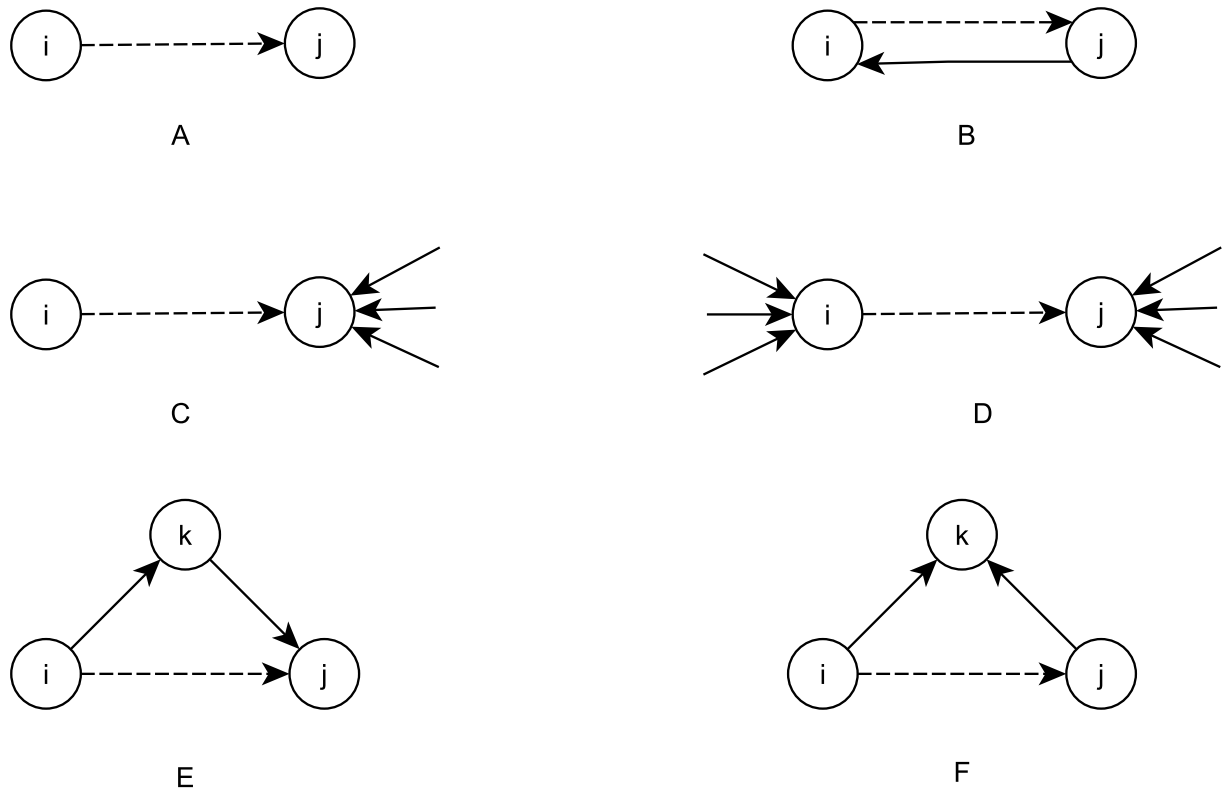
Fig 4. Illustrations of different mechanisms considered. (A) parameter q, (B) mutuality mechanism, (C) popularity mechanism, (D) assortativity mechanism, (E) transitivity mechanism, (F) outgoing-shared-partners mechanism.

These network statistics (S) are weighted (by considering θ) and summed to produce vector ϕ as described in the previous subsection. The local network mechanisms of which the network statistics are weighted with higher weights (in an absolute value) are more important in the network's evolution. The interpretation of a given mechanism depends on the sign of a corresponding weight. For example, positively weighted popularity statistics refers to the tendency to create a link to those with a relatively high in-degree. On the contrary, a negative sign reflects the tendency to avoid establishing links to those with a relatively high in-degree.

The mechanisms are defined in the same way as in a study of the asymmetric core-cohesive blockmodel type [2]. Therefore, only a brief description of the proposed mechanisms is given here (the mechanisms are schematically shown in Fig 4, where dashed lines illustrate the links under evaluation appear, are confirmed, or disappear):

0. **Parameter** q (Fig 4A) reflects the tendency to have a link. Since this is not a focal mechanism, it is implemented in the NEM algorithm as parameter q and is therefore technically not considered as a mechanism in this study.

1. The **mutuality mechanism** (M) (Fig 4B) reflects the tendency to reciprocate links.

2. The **alter popularity mechanism** (P) (Fig 4C) reflects the tendency to create links to the most popular ones.

3. The **assortativity mechanism** (A) (Fig 4D) reflects the tendency to create links to those units with the same level of popularity (in-degree).

4. The **transitivity mechanism** (*T*) ([Fig 4E](#)) is a tendency for a unit to directly connect to units, to which it is indirectly connected with (one or more) paths of length two (with more paths increasing the tendency).

5. The **outgoing-shared-partner mechanism** (*OSP*) ([Fig 4F](#)) represents a "structural homophily" effect which is traditionally based on similarity according to the units' attributes. In the case of the OSP, it is defined by similar choices of partners [32].

Simulation design

The described NEM algorithm is used to generate the networks by considering the selected social mechanisms. Since different mechanism strengths are to be considered, 300 randomly selected θ are generated. The random values are generated by first sampling five values from the standard normal distribution Φ and then multiplying them by a scalar [33, 34] (after such normalization, the sum of the squared elements of θ equals 1).

$$\theta = \frac{\Phi}{\sqrt{\sum \Phi_i^2}} \tag{1}$$

Within the NEM algorithm, parameter q is set to 5/9 and a total of 116,490 iterations are applied. Parameter q, which indicates the tendency to create a link, is set arbitrarily but with reference to generating asymmetric core-cohesive blockmodels. Initial networks are empty with 24 units.

The 30 networks are generated for each θ. Generalized blockmodeling for binary networks (on symmetrized generated networks, structural equivalence is used) is done after the selected number of iterations of the algorithm. More precisely, the intermediate number of iterations m, at which the global network structure is analyzed, is determined as $m_i = m_{i-1} * 1.9$, where $m_1 = 100$ [2]. This approach is used since most changes in the structure of the links happen at a lower number of iterations.

Based on the generalized blockmodeling solution, the number of inconsistent blocks is calculated on the network generated on the 32,969th iteration and used as the fit function. It is defined as the number of different blocks between the symmetric core-cohesive blockmodel with three groups and the empirically obtained blockmodel with three groups.

Some θs that generate networks with the lowest number of inconsistent blocks are selected. Based on the selected θs, the networks are again generated but with a higher number of iterations (116,490 iterations). For each network generated by the selected θs, the relative fit (RF) [2] is calculated (for networks with a different number of iterations up to 116,490) as

$$RF = 1 - \frac{P^m}{\frac{1}{k} \sum_{i=1}^{k} P_i^r} \tag{2}$$

where P^m is the value of the criterion function [35, 36] obtained on the empirical network and P_i^r is the value of the criterion function obtained on the *i*-th randomized network. There are k randomized networks. The mean value of the criterion function in the case of random networks is estimated by simulations. RF is a more detailed measure of the fit of a given blockmodel to the empirical data and its use is most valid when the presence of a given blockmodel type is confirmed by non-specified blockmodeling. Higher values indicate a better fit (the value of 1 indicates a perfect fit) and the expected value of the RF measure in the case of a random network is zero.

Table 2. Mean number of inconsistent blocks and mean RF values with the corresponding parameter values. For those θs which generated networks with the mean number of inconsistent blocks at 32,969 iterations equal to zero. Initial is an empty network.

θ ID	θ					MRF
	M	P	A	T	OSP	
136	-.18	.74	.37	-.35	.42	0.96
25	-.43	.27	.66	.25	-.50	0.80
279	.17	-.11	.43	.60	.65	0.50
248	.11	-.58	.49	.78	-.38	0.48
72	-.57	.68	.04	-.46	.10	0.35
22	-.24	-.51	.21	-.21	-.78	0.26

θ ID	NO. OF ITERATIONS											
	100	190	361	686	1,303	2,478	4,705	8,939	16,948	32,969	61,311	116,490
136	4.93	4.97	4.40	3.30	0.33	0.03	0.03	0.07	0.03	0.00	0.00	0.00
25	4.90	4.77	3.57	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
279	4.73	4.93	3.43	0.20	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
248	4.73	4.97	4.00	1.37	0.33	0.00	0.00	0.03	0.00	0.00	0.00	0.00
72	4.90	4.97	4.13	3.77	0.70	0.20	0.03	0.00	0.10	0.00	0.00	0.00
22	5.00	5.10	4.45	2.03	0.17	0.03	0.00	0.00	0.00	0.00	0.00	0.00

M = mutuality mechanism, P = popularity mechanism, A = assortativity mechanism, T = transitivity mechanism, OSP = outgoing-shared-partners mechanism,

MRF = the mean RF value

Results: Generated networks

There are six 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.

The θs that generate each network with a symmetric core-cohesive blockmodel are shown in Table 2 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 RF.

A more detailed insight into RF for a selected θ is given in Fig 5. 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 low 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 on the generated networks Fig 6. 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. At this step, the links among different groups are established yet, in some cases, links within the core units are not present. 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 is only increasing until 61,311 iterations.

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 Fig 6). The mean RF for the cohesive blockmodel is lower while the mean RF for the symmetric core-periphery blockmodel
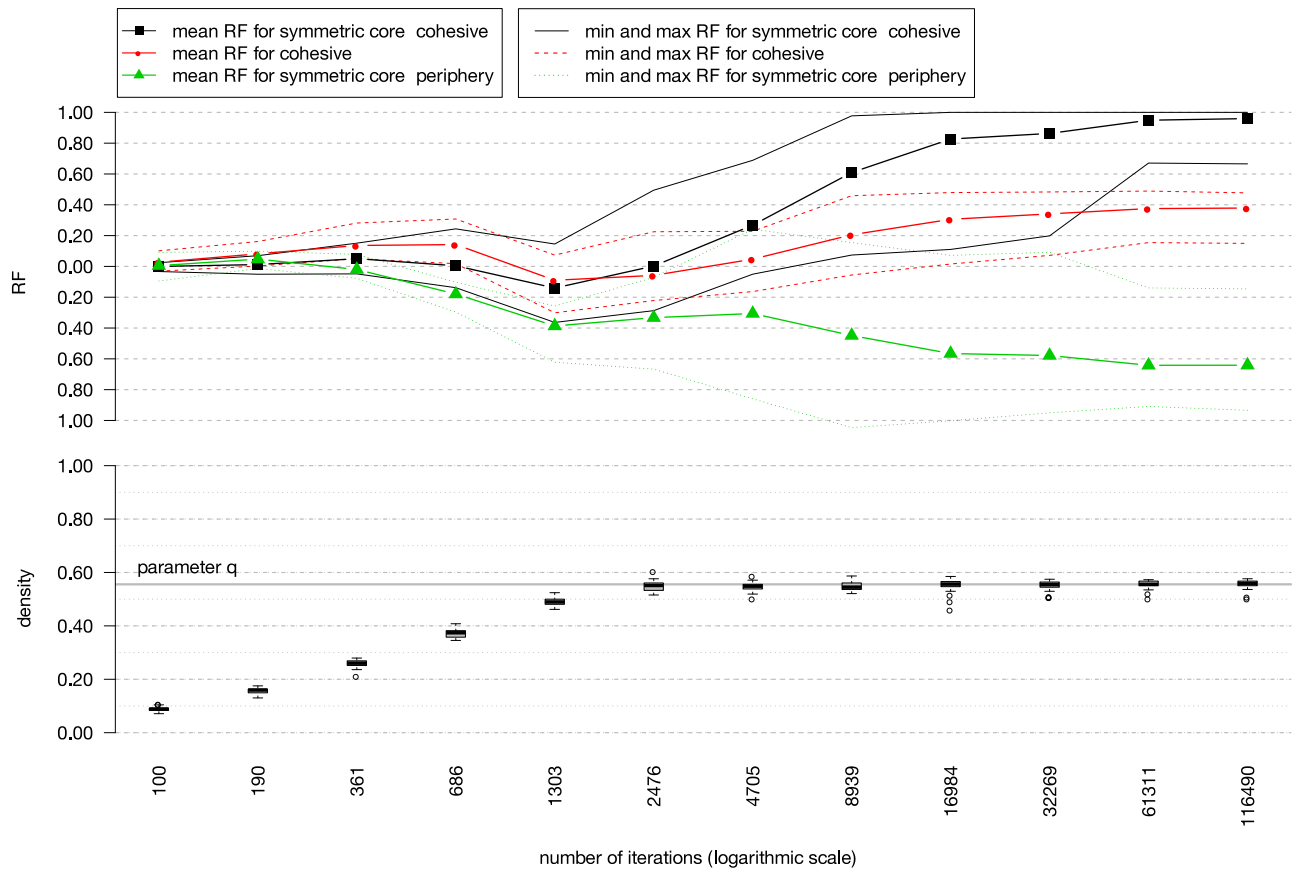
Fig 5. Mean RF for each blockmodel type visualized by lines and the distribution of the density visualized by boxplots. The networks are generated by considering $\theta = \{M = -0.18, P = 0.74, A = 0.37, T = -0.35, OSP = 0.42\}$, $q = 5/9$, $d_0 = 0$.
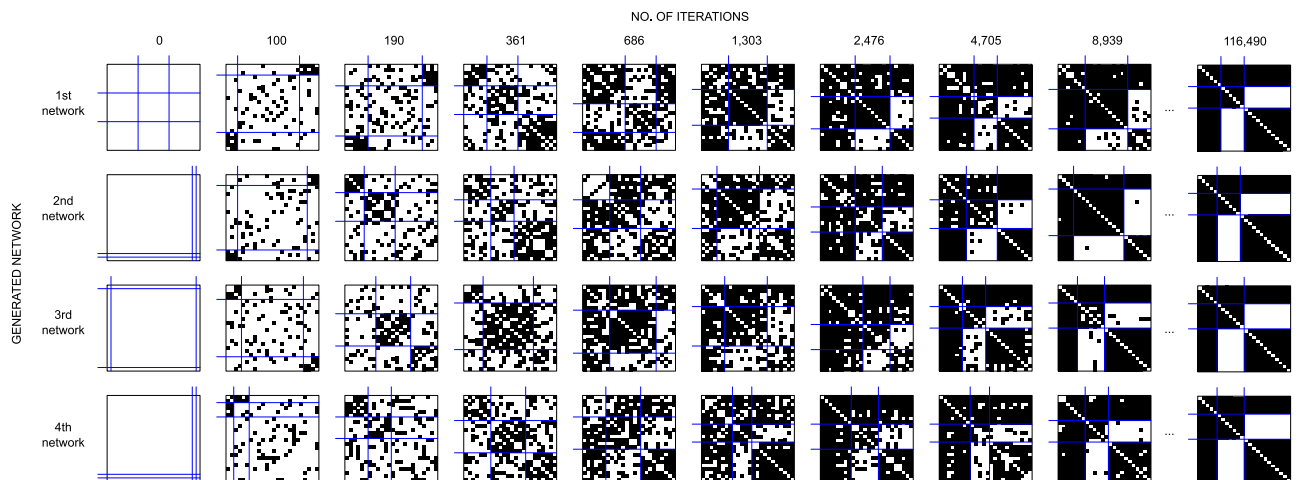
Fig 6. Some networks generated. The networks are generated by considering $\theta = \{M = -0.18, P = 0.74, A = 0.37, T = -0.35, OSP = 0.42\}$, $q = 5/9$, $d_0 = 0$. The networks are drawn in line with the blockmodels obtained by generalized blockmodeling (non-specified model). Networks for different repetitions of the algorithm for generating networks are drawn in lines for different numbers of iterations.

type is highly negative, indicating that the randomized networks fit this blockmodel type much more than the networks generated by using the proposed algorithm.

Generating networks with other approaches

Because different approaches to generating networks rely on different assumptions (not only on the underlying social process under study, but also on the data collecting technique), they must be carefully chosen with respect to these assumptions to draw valid conclusions. As stated above, the chosen approach (any-link symmetrization approach) to generate and symmetrize links is seen as the most appropriate for simulating the social process of interactions and corresponding data collecting technique. While the other approaches listed are regarded as less appropriate, they are not a priori discarded. Therefore, the following results aim to evaluate the extent to which different approaches affect the ability to generate the symmetric core-cohesive global network structure. The results are obtained using the same methodology as in the previous sections, but with different approaches to generating networks.

The set of 300 θs is the same in the case of the reciprocal-link symmetrization approach and any-link symmetrization approach, and similar for the immediate any-link symmetrization approach (values corresponding to the mutuality mechanism are removed from the set of θs while values corresponding to other local network mechanisms are normalized such that the sums of squared elements of θs equal to 1). The parameter q is set to $q = 7/5$ in the case of the immediate any-link symmetrization approach and in the case of the reciprocal-link symmetrization approach.

The number of inconsistent blocks is evaluated on the 32,969 iteration. It can be seen in Table 3 that there are 6 θs that generated all 30 networks without any inconsistent block in the case of the any-link symmetrization approach. On the contrary, there is no such θ when using the other approaches. However, there are several θs that generated networks with less than or equal to 0.5 mean inconsistent blocks in the case of all three approaches to generating networks. The highest number of such θs is in the case of the any-link symmetrization approach, followed by the reciprocal-link symmetrization approach and any-link symmetrization approach. There are 18 commonly chosen θs (out of 30 randomly generated θs) when using the immediate any-link symmetrization approach and the any-link symmetrization approach and 3 θs when using the reciprocal-link symmetrization approach and the any-link symmetrization approach.

Table 3. The number of selected θs (out of 300 randomly generated θs) according to different selection criteria, the minimum mean IB and the maximum RF for different approaches to generating networks.

		immediate any-link symmetrization approach	reciprocal-link symmetrization approach	any-link symmetrization approach
number of θs that generated all networks without any inconsistent block at 32,969 iteration		0	0	6
number of θs that generated networks with $\leq$0.5 mean number of inconsistent blocks at 32,969 iteration		30	7	76
overall minimum mean number of inconsistent blocks	at 32,969 iteration	0.03	0.17	0.00
	at 116,490 iteration	0.13	0.10	0.00
the overall maximum RF value	at 32,969 iteration	0.90	0.39	0.86
	at 116,490 iteration	0.96	0.37	0.96

The networks were generated again by considering the θs that generated networks with equal to or less than 0.5 inconsistent blocks, but where the number of iterations is increased to 116,490. The lowest overall mean number of inconsistent blocks can be observed with the any-link symmetrization approach, which is expected since some θs generated all networks without any inconsistent block.

The overall maximum mean RF value is highest in the case of the immediate any-link symmetrization approach (0.90 at 32,969 iterations and 0.96 at 116,490 iterations), followed by the any-link symmetrization approach with similar overall maximum RF values. The global network structures are much more stable in the case of the reciprocal-link symmetrization approach, despite the high mean RF values. Besides the fact that fewer generated networks with the asymmetric core-cohesive blockmodel emerge in the case of the immediate any-link symmetrization approach, it is also very common for the asymmetric core-cohesive blockmodel to appear at lower iterations but, later, the share of nodes in cohesive groups increases and the global network structure becomes closer to cohesive with two clusters. When the true blockmodel type is cohesive, instead of symmetric core-cohesive, the RF values can be overestimated when calculated on the assumption that the true blockmodel is symmetric core-cohesive.

The mean RF values for the networks generated using the reciprocal-link symmetrization approach are very low, indicating that the symmetric core-cohesive blockmodel cannot be generated with this approach. The density of the networks generated with the selected θs is extremely high.

Comparing the results of the different approaches to generating networks revealed that use of the any-link symmetrization approach generates networks with a very small number of inconsistent blocks, very high mean RF values, and very stable global network structures. According to the analysis, which is based on 300 randomly generated θs, this cannot be claimed for the other approaches to generating networks.

Conclusion

Interactional networks collected in preschool classrooms are studied in this paper. In such an environment, children start to form groups. Children within individual groups spend more time with each other than they do with children from other groups. At the same time, a group of children is formed which spends a considerable amount of time with all the others from any group. This leads to the newly proposed blockmodel type, i.e. symmetric core-cohesive. It consists of one group of units which are called core units and two or several other groups of units called cohesive groups. The units from all groups are internally linked to each other. The units from all cohesive groups are linked to the units from the core group and vice versa. The units from different cohesive groups are not linked to each other.

The existence of this blockmodel type is evaluated on empirical data. The data were collected within a larger longitudinal study among preschool children in the United States 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 analyzed classes in at least one time period. This proves that the proposed global structure (blockmodel type) is relevant for such data.

The most common local network mechanisms (popularity, assortativity, transitivity, and outgoing-shared-partners) are considered. The attributes of the units are not taken into account and the initial networks are empty. Based on an adapted version of the algorithm proposed by Cugmas et al. [2], three approaches to simulating the emergence of symmetric interactional networks are proposed. The any-link symmetrization approach is seen as the most appropriate for

studying the emergence of the symmetric core-cohesive blockmodel. Using this approach, asymmetric links are formed throughout the simulation process, but the generated asymmetric links are symmetrized prior to the analysis. The results of the Monte Carlo simulations confirmed that the selected mechanisms can generate networks with a 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 a 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 for describing and analyzing empirical interactional network global structures. The fact that this global network structure can emerge in empirical interactional networks may form the basis for future research on the developmental questions related to this global network structure that were raised in the introduction section.

Second, understanding the link between the global network structure and the local network mechanisms in a given context is necessary for studying (e.g., modelling) empirically obtained networks. It has been shown that the selected local network mechanisms are important in the formation of a symmetric core-cohesive blockmodel even without considering any other attributes of the units.

Supporting information

S1 Dataset. Non-binarized symmetrizied networks for the considered school classes and time periods. One network is for one school class and one time period. The networks are in matrix form. Each matrix is in its own file (.csv format).
(ZIP)

S2 Dataset. Binarized symmetrizied networks for the considered school classes and time periods. One network is for one school class and one time period. The networks are in matrix form. Each matrix is in its own file (.csv format).
(ZIP)

Author Contributions

Conceptualization: Marjan Cugmas, Dawn DeLay, Anuška Ferligoj.

Data curation: Dawn DeLay.

Formal analysis: Marjan Cugmas, Aleš Žiberna.

Funding acquisition: Aleš Žiberna, Anuška Ferligoj.

Investigation: Marjan Cugmas, Dawn DeLay, Anuška Ferligoj.

Methodology: Marjan Cugmas, Aleš Žiberna, Anuška Ferligoj.

Project administration: Marjan Cugmas.

Software: Marjan Cugmas, Aleš Žiberna.

Supervision: Aleš Žiberna, Anuška Ferligoj.

Validation: Anuška Ferligoj.

Visualization: Marjan Cugmas.

Writing – original draft: Marjan Cugmas, Dawn DeLay.

Writing – review & editing: Dawn DeLay, Aleš Žiberna, Anuška Ferligoj.

References

1. Bianconi G, Darst RK, Iacovacci J, Fortunato S. Triadic closure as a basic generating mechanism of communities in complex networks. Physical Review E. 2014; 90(4):042806. https://doi.org/10.1103/PhysRevE.90.042806

2. Cugmas M, Žiberna A, Ferligoj A. Mechanisms generating asymmetric core-cohesive blockmodels. Metodološki zvezki. 2019; 16(1):17–41.

3. Fabes RA, Martin CL, Hanish LD. The next 50 years: Considering gender as a context for understanding young children's peer relationships. Merrill-Palmer Quarterly. 2004; 50(3):260–273. https://doi.org/10.1353/mpq.2004.0017

4. Martin CL, Fabes RA, Hanish LD, Hollenstein T. Social dynamics in the preschool. Developmental Review. 2005; 25(3-4):299–327. https://doi.org/10.1016/j.dr.2005.10.001

5. Doreian P, Batagelj V, Ferligoj A. Generalized blockmodeling. Cambridge: Cambridge University Press; 2005.

6. Schaefer DR, Light JM, Fabes RA, Hanish LD, Martin CL. Fundamental principles of network formation among preschool children. Social networks. 2010; 32(1):61–71. https://doi.org/10.1016/j.socnet.2009.04.003 PMID: 20161606

7. Brown BB, Klute C. Friendships, cliques, and crowds. In: Adams G, Berzonsky M, editors. Blackwell Handbook of Adolescence. Wiley-Blackwell; 2003. p. 266–290.

8. Lorrain F, White H. Structural equivalence of individuals in social networks. The Journal of Mathematical Sociology. 1971; 1(1):49–80. https://doi.org/10.1080/0022250X.1971.9989788

9. Batagelj V, Ferligoj A, Doreian P. Direct and indirect methods for structural equivalence. Social Networks. 1992; 14(1-2):63–90. https://doi.org/10.1016/0378-8733(92)90014-X

10. White DR, Reitz KP. Graph and semigroup homomorphisms on networks of relations. Social Networks. 1983; 5(2):193–234. https://doi.org/10.1016/0378-8733(83)90025-4

11. Stadtfeld C, Takács K, Vörös A. The Emergence and Stability of Groups in Social Networks. SSRN. 2018; p. 1–60.

12. Maccoby EE, Jacklin CN. Gender segregation in childhood. In: Advances in child development and behavior. Elsevier; 1987. p. 239–287.

13. Vaughn BE, Colvin TN, Azria MR, Caya L, Krzysik L. Dyadic analyses of friendship in a sample of preschool-age children attending Head Start: Correspondence between measures and implications for social competence. Child development. 2001; 72(3):862–878. https://doi.org/10.1111/1467-8624.00320 PMID: 11405587

14. Daniel JR, Santos AJ, Peceguina I, Vaughn BE. Exponential random graph models of preschool affiliative networks. Social Networks. 2013; 35(1):25–30. https://doi.org/10.1016/j.socnet.2012.11.002

15. Snyder J, West L, Stockemer V, Gibbons S, Almquist-Parks L. A social learning model of peer choice in the natural environment. Journal of Applied Developmental Psychology. 1996; 17(2):215–237. https://doi.org/10.1016/S0193-3973(96)90026-X

16. Block P. Reciprocity, transitivity, and the mysterious three-cycle. Social Networks. 2015; 40:163–173. https://doi.org/10.1016/j.socnet.2014.10.005

17. Kandel DB. Homophily, selection, and socialization in adolescent friendships. American journal of Sociology. 1978; 84(2):427–436. https://doi.org/10.1086/226792

18. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annual review of sociology. 2001; 27:415–444. https://doi.org/10.1146/annurev.soc.27.1.415

19. Dijkstra JK, Cillessen AH, Borch C. Popularity and adolescent friendship networks: Selection and influence dynamics. Developmental Psychology. 2013; 49(7):1242–1252. https://doi.org/10.1037/a0030098 PMID: 22985296

20. Cillessen AH, Rose AJ. Understanding popularity in the peer system. Current Directions in Psychological Science. 2005; 14(2):102–105. https://doi.org/10.1111/j.0963-7214.2005.00343.x

21. Robins G, Pattison P, Kalish Y, Lusher D. An introduction to exponential random graph (p*) models for social networks. Social networks. 2007; 29(2):173–191.

22. Block P, Stadtfeld C, Snijders TA. Forms of Dependence Comparing SAOMs and ERGMs From Basic Principles. Sociological Methods & Research. 2019; 48(1):202–239. https://doi.org/10.1177/0049124116672680

23. Handcock MS, Robins G, Snijders T, Moody J, Besag J. Assessing Degeneracy in Statistical Models of Social Networks. Journal of the American Statistical Association. 2003; 76:33–50.

24. Snijders TA, Van de Bunt GG, Steglich CE. Introduction to stochastic actor-based models for network dynamics. Social networks. 2010; 32(1):44–60. https://doi.org/10.1016/j.socnet.2009.02.004

25. Hartup WW, Stevens N. Friendships and adaptation in the life course. Psychological bulletin. 1997; 121(3):355–370. https://doi.org/10.1037/0033-2909.121.3.355

26. Hartup WW. Adolescents and their friends. New Directions for Child and Adolescent Development. 1993; 1993(60):3–22. https://doi.org/10.1002/cd.23219936003

27. Žiberna A. Generalized and Classical Blockmodeling of Valued Networks; 2018. Available from: https://cran.r-project.org/web/packages/blockmodeling/blockmodeling.pdf.

28. Van de Bunt GG, Groenewegen P. An Actor-Oriented Dynamic Network Approach: The Case of Interorganizational Network Evolution Organizational Research Methods. 2007; 10(3):463–482. https://doi.org/10.1177/1094428107300203

29. Bauer F, Hansen T, Hellsmark H. Innovation in the bioeconomy—dynamics of biorefinery innovation networks Technology Analysis & Strategic Managment. 2018; 30(8):935–947. https://doi.org/10.1080/09537325.2018.1425386

30. Snijders T. Stochastic Actor-Oriented Models for Network Dynamics Annual Review of Statistics and Its Application. 2017; 4(1):434–363. https://doi.org/10.1146/annurev-statistics-060116-054035

31. Toivonen R, Kovanen L, Kivelä M, Onnela JP, Saramäki J, Kaski K. A comparative study of social network models: Network evolution models and nodal attribute models. Social Networks. 2009; 31(4):240–254. https://doi.org/10.1016/j.socnet.2009.06.004

32. Robins G, Pattison P, Wang P. Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. Social Networks. 2009; 31(2):105–117. https://doi.org/10.1016/j.socnet.2008.10.006

33. Marsaglia G. Choosing a point from the surface of a sphere. The Annals of Mathematical Statistics. 1972; 43(2):645–646. https://doi.org/10.1214/aoms/1177692644

34. Muller ME. A note on a method for generating points uniformly on n-dimensional spheres. Communications of the ACM. 1959; 2(4):19–20. https://doi.org/10.1145/377939.377946

35. Batagelj V. Notes on blockmodeling. Social Networks. 1997; 19(2):143–155. https://doi.org/10.1016/S0378-8733(96)00297-3

36. Doreian P, Batagelj V, Ferligoj A. Partitioning networks based on generalized concepts of equivalence. Journal of Mathematical Sociology. 1994; 19(1):1–27. https://doi.org/10.1080/0022250X.1994.9990133