Cluster analysis of multiplex networks

Snijders, Thomas; Vörös, András

Published in:
Social Networks

DOI:
10.1016/j.socnet.2017.01.002

Social Networks 49 (2017) 93–112

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Cluster analysis of multiplex networks: Defining composite network measures

András Vörös [a,b,d,*], Tom A.B. Snijders [b,c]

[a] ETH Zürich, Chair of Social Networks, Clausiusstrasse 50, 8092 Zürich, Switzerland
[b] University of Oxford, Nuffield College, New Road, Oxford OX1 1NF, United Kingdom
[c] University of Groningen, Department of Sociology, Grote Rozentstraat 31, Groningen 9712 TG, The Netherlands
[d] MTA-TK "Lendület" Research Center for Educational and Network Studies, [Hungarian Academy of Sciences], Országház utca 30, Budapest 1014, Hungary

ARTICLE INFO

Article history:
Available online 16 January 2017

Keywords:
Multiplex networks
Multilayer networks
Dimension reduction
Cluster analysis
Peer perceptions
Adolescents

ABSTRACT

Social relations are multiplex by nature: actors in a group are tied together by various types of relationships. To understand and explain group processes it is, therefore, important to study multiple social networks simultaneously in a given group. However, with multiplexity the complexity of data also increases. Although some multivariate network methods (e.g. Exponential Random Graph Models, Stochastic Actor-oriented Models) allow to jointly analyze multiple networks, modeling becomes complicated when it focuses on more than a few (2–4) network dimensions. In such cases, dimension reduction methods are called for to obtain a manageable set of variables. Drawing on existing statistical methods and measures, we propose a procedure to reduce the dimensions of multiplex network data measured in multiple groups. We achieve this by clustering the networks using their pairwise similarities, and constructing composite network measures as combinations of the networks in each resulting cluster. The procedure is demonstrated on a dataset of 21 interpersonal network dimensions in 18 Hungarian high-school classrooms. The results indicate that the network items organize into three well-interpretable clusters: positive, negative, and social role attributions. We show that the composite networks defined on these three relationship groups overlap but do not fully coincide with the network measures most often used in adolescent research, such as friendship and dislike.

1. Introduction

In a classic study of U.S. high schools in the late 1950s, Coleman and his colleagues asked students to name both their peers from school whom they considered as friends and those whom they most wished to be like (Coleman, 1961). The distinction between friend and role model relations led the researchers to discover that every school had a distinct group of "local leaders" (students popular as friends) and "system leaders" (students popular as role models), and that it was the latter group which had a larger impact on the school community as a whole (Coleman, 1961, Chapter 4). This phenomenon, important as it is in understanding the value system of adolescents, would not have been revealed had the study focused solely on friendships between students. The work of Coleman and colleagues demonstrates that without taking into account the multiplexity of relations, the social forces shaping communities cannot be fully understood.

The importance of multiplexity has long been recognized in the research tradition of Social Network Analysis (Kadushin, 2012; Prell, 2011; Borgatti et al., 2009), but its potential for aiding explanations is far from being fully exploited. The collection of rich multiplex network datasets has begun already several decades ago (e.g. Sampson, 1969; Coleman, 1961; Homans, 1950).[1] Additionally, one can also witness the early development of methods appropriate for analyzing this kind of data (e.g. Davis, 1968; Fienberg et al., 1985). However, most multiplex studies and methods focus on only two or three types of relations at most. Common sense tells us that social ties, may they connect people, firms, or countries, are more heterogeneous than that. Yet, even nowadays we rarely see network studies that measure a large number, say

* Corresponding author. Present address: Clausiusstrasse 50, 8092 Zürich, Switzerland.
E-mail addresses: andras.voros@nuffield.ox.ac.uk, andras.voros@gess.ethz.ch (A. Vörös).

[1] The popular software packages for network analysis contain a variety of classic multiplex datasets. For example, networks among Sampson's monks and the bank wiring room data (Roethlisberger and Dickson, 1939) studied by Homans are available in UCINET (Borgatti et al., 2002).

http://dx.doi.org/10.1016/j.socnet.2017.01.002
0378-8733/

dozens, of relational dimensions in a given social context. This seems surprising especially in times when network data, even on a large scale, are becoming easier to collect and computational capacity is sufficient for analyzing complex data structures.

Studying a small set of relations can be a good practice as long as we have well-grounded knowledge about, and valid measures of, the most important types of social ties in a given context. However, this seems usually not to be the case, and simplicity of measurement tends to outweigh validity and thoroughness. For example, affective relations are quite often studied by friendship ties. Friendship is a complex relationship, yet it is mostly measured by the single question "Whom do you consider a friend?". As a price, additional work is needed to find out what people actually mean when they answer this simple question. And we do know, for instance, that they mean a lot of different things (Fischer, 1982), that what they mean changes with age (Berndt, 1982:1449), and that friendship is multidimensional (Marsden and Campbell, 1984). Consequently, it seems to be a better practice to complement our analyses of friendship networks with several related network dimensions.

Measuring a large number of relational dimensions, maybe dozens, makes it possible to define each of them very specifically, and thereby to obtain valid measures of various concrete social relations between actors (rhyming with the old wisdom of sociometricians). Even more importantly, however, by exploring the interrelations between these measures, one may be able to construct aggregated networks that represent some latent dimensions of social relations – just as in the case of psychological scale construction, but now tailored to networks. This approach to multiplexity can help us define more valid measures of complex relationships, such as friendship, and potentially new measures of so far unexplored relational concepts. The present paper attempts to contribute to a step in this direction by outlining a suitable analytical procedure and demonstrating its value on a real-world example.

1.1. Multiplexity as a challenge

The complexity of multiplex network data with many relational dimensions presents two general challenges compared to studying only one or a few networks at once. First, from a substantive point of view, with more network dimensions studied, the number of potential mechanisms involved in a social phenomenon increases rapidly. For example, in the context of Coleman's high-school study we may be interested in why certain students become popular in their community. The potential answers to this question become numerous as soon as we recognize that there are two distinct pathways to popularity: e.g., making friends and becoming a role model. How can students become role models outside of their local group? How can role models become popular in their local friendship groups? Do students tend to see their friends as role models? Or do they want to befriend those they look up to? These examples highlight the need for a series of theoretically and empirically grounded ideas about how multiple types of relationships are interlinked in a given context. Taking into account the multitude of possible interdependencies between several network dimensions, hypotheses about numerous social mechanisms may be necessary. These are bound to be ad hoc unless a strong theoretical framework is available.

The second, and related, challenge posed by the complexity of multiplex data is of a technical nature. It becomes necessary to use statistical tools which are able to efficiently and comprehensibly model the complex interdependencies of network ties within and between the types of social relations in question. Some of the recent multivariate statistical methods in the field allow the analysis of multiplex networks. For example, the structure of a dependent network can be explained by other networks using MRQAP (Krackhardt, 1988; Dekker et al., 2007), ERGMs (Lusher et al., 2012), or SAOMs (Snijders et al., 2010). Furthermore, the mutually dependent modeling of multiple network dimensions is also an option in case of the latter two methods (Lusher et al., 2012, Chapter 10; Snijders et al., 2013). These two techniques are flexible enough to model various specific forms of interdependencies between multiple networks, including actor-level, dyadic, and triadic interrelations. However, the same flexibility makes working with these techniques a daunting task already in case of three or four networks, let alone more.

1.2. Dimension reduction to the rescue

Dimension reduction is one of the successful methodological strategies to tackle both the substantive and the technical challenges that come with "too much" data. The general approach represented by Cluster Analysis, Factor Analysis, or other methods for dimension reduction of non-network, monadic data has been frequently applied to different problems in the analysis of social networks. It is enough to think about blockmodeling (Doreian et al., 2005; White et al., 1976) which aims to identify subsets of actors based on their position in the network. In some cases, the classification of actors may be based on their position in multiplex relational structures (e.g. Dabkowski et al., 2015; Lazega, 2001).

Of the classic techniques of scaling, we can find applications of Correspondence Analysis for the simultaneous clustering of tie senders and receivers in one-mode and two-mode networks (Faust, 2005; Wasserman et al., 1990; Wasserman and Faust, 1989), or for the comparison of structural features of networks measured in multiple groups (Faust and Skvoretz, 2002; Skvoretz and Faust, 2002). Cluster Analysis also appears as a method to classify groups based on their network structure (Brandes et al., 2011). Further, Multi-dimensional Scaling is used to visualize similarities between network actors (e.g., Hanneman and Riddle, 2005), or to assess the presence of a relational hierarchy in certain types of networks (Borgatti, 1994).[2]

Many more applications of dimension reduction methods to social networks could be mentioned. What seems to be missing from the literature, however, are studies that aim to analyze the multivariate similarities between multiple relational measures. Analogous to the approach of scale construction in the case of psychological tests, we could consider defining aggregated networks based on a large set of relational variables found to be similar in some respect.

1.3. A precursor in psychology: peer ratings of behavior in school

The first steps in constructing scale-type, aggregated network measures were actually taken in the discipline of psychology several decades ago. Starting from the middle of the 20th century, psychologists concerned with social adjustment problems of children and adolescents have studied the relationships between sociometric status and behavior in peer groups (for a review of this research stream, see Asher and Coie, 1990). To find the behavioral correlates of sociometric status in school, researchers developed tests in which students had to evaluate their peers on several (30–50) items which described various social behaviors.

As examples for the rich collection of tests that have been employed, we can mention the Peer Nomination Inventory (Wiggins and Winder, 1961), the Class Play Method (Bower, 1960; Masten et al., 1985), the Pupil Evaluation Inventory (Pekarik et al.,

[2] Although a linear model, we can also mention Canonical Correlation Analysis which some authors have applied to study and summarize the dyadic interrelations between a few types of relations in multiplex datasets (Carroll, 2006; Wasserman and Faust, 1989).

1976), and the Adjustment Scales for Sociometric Evaluation of Secondary-School Students or ASSESS (Prinz et al., 1978).

In all of these cases, peer nominations are aggregated within the classroom or group to individual scores that represent the number of nominations students received in each item. The individual scores are then usually studied by Factor Analysis to identify the latent dimensions underlying the pool of behavior items. Finally, the relationships between the factor variables and sociometric status, which is based on nominations received in independent like-dislike items, are explored.

Although the mentioned peer assessment methods vary in the selection procedure and final list of behavioral items, their findings point in the same general direction. In all age groups, sociometric status of children appears to be associated with the sociable/cooperative and the aggressive/disruptive dimensions of behavior. The peer evaluation approach outlined here has contributed greatly to the study of social processes related to rejection (Newcomb et al., 1993; Asher and Coie, 1990; Coie et al., 1982) and popularity (Cillessen et al., 2011) among children and adolescents.

An important strength of the above described studies is that they follow a multiplex network approach to the identification of the main behavioral dimensions in peer groups. The researchers start with a large set of network measures and arrive at a small number of aggregated variables which are easier to handle and which retain a significant part of the information in the original data. The networks of interest are interpersonal perceptions about various forms of social behavior among children. The classification of perception items depend on the pairwise similarities (inter-item correlations) between students' indegree centrality in each network: two perceived behavior networks will be more likely to end up on the same factor if students who receive many (few) nominations on one item, also receive many (few) on the other. Such a conceptualization suits the aim of the psychologists of the area as they investigate the differences in the behavioral profile of children in distinct sociometric status positions (e.g. rejected, popular).

In these studies, dyadic perceptions are aggregated to individual-level indegrees. This step discards network structure entirely and creates a risk of ecological fallacy: tie-level and individual-level, or even group-level, processes will be confounded by this aggregation. The case of sociometric status and aggression can serve as an example. Different studies found that aggression is related to popular (Cillessen and Rose, 2005:104), controversial (Coie et al., 1982:565), and rejected status (Coie et al., 1990:20). This variation in results may be due to differences between the studied groups with respect to dyadic and group-level social processes: the tendency of children to like or dislike those who they think are aggressive, and the tendency of the group to agree on which peers are aggressive or likeable. It is hard to tell the two apart. Reducing perceptional data to indegree scores obscures heterogeneities in the network, such as the existence of subgroups with different perceptions.

1.4. Multiplexity as an opportunity: the cluster analysis of multiplex networks

The argument discussed so far does not hold only for peer perceptions and sociometric status among children but points to a more general approach of dimension reduction in multiplex network data. In such data, the several network dimensions will be referred to as *network items*, because we regard them similarly as items in a psychological test: questions that are intended to measure an aspect of social reality – in this case a relational, not an individual aspect.

When the focus is on relations, similarity on the level of dyads is of more crucial importance than similarity between individuals. If actors tend to nominate the same peers in one network dimension as in another, then the two network items may capture different sides of the same relationship. For example, sharing secrets and willingness to borrow money may be thought of as constituents of trust. The classification of multiplex networks based on their dyadic similarities allows us to aggregate information on the level of network ties. That is, we could say that if two actors share their secrets and are willing to borrow money to each other, then they can be considered to have a trust relationship. This way we may be able to reduce a large multiplex dataset to a few composite networks, representing the main underlying relational dimensions emergent from the dyadic patterns of ties across the several measured items. Naturally, the same logic can be followed in the analysis of weighted networks (e.g. how much time people spend together, how much money they lend to each other). However, this paper focuses on the case of binary networks, as these are used most often in the study of children and adolescent communities.

Classifying network items in a single group, like a classroom or a workplace, can be meaningful for describing a particular social environment. However, the findings from one group are unlikely to be generalizable to a larger population, which has to do with the social interactions that occur in different groups. For example, dominant individuals might be found to be agreeable in one group, while disliked in another, merely because of the different personalities of the dominant individuals, or because of the occurrence of a conflict in one group and the role taken by the dominant informal group leaders. A greater degree of generalizability may be obtained when associations between relations are studied in multiple groups at the same time.

Similar to the approach represented by Factor and Cluster Analysis of monadic variables, here we present an analytical procedure for dimension reduction in multiplex network data measured in multiple groups. A network dimension, or network item, we define operationally by the question posed or the measurement instrument used; sometimes we refer to this just as the 'network' or 'relation'. The data structure is a crossed structure of network items by groups. By exploring the multivariate similarities between the measured network dimensions, the procedure aims at finding clusters of relations that are similar for each of the groups. For this purpose we apply Cluster Analysis to the similarity structure in every group. In doing so, we assume throughout that the studied groups are homogeneous in the sense that the measured network dimensions are related in more or less the same way in each of them, while allowing for some differences between the groups – related perhaps to their peculiar composition.

Based on a cluster solution that is reasonable for all of the studied groups, we define composite network measures. In some cases there may be straightforward ways for doing this, such as taking the union or the intersection of all networks that belong to a given cluster (see some examples in De Domenico et al., 2013). Instead of choosing either of these two extremes, we propose to explore several possibilities, with a tie in at least t (where t may range from 1 to the number of networks in a cluster) of the networks required to be present between two actors to qualify as a tie in the composed network. Exploring the properties of the resulting composite networks at different thresholds t, and also their relations with each other and with other basic network measures (such as friendship), can inform the specification of composite networks. We demonstrate that this procedure is capable of reducing the number of studied networks to a level that can be analyzed with existing multiplex network models, and it can also result in composite networks that meaningfully complement analyses based on more conventional network measures in the studied context.[3]

[3] Although the approach we take here produces binary networks (with dichotomous tie values) in the first place, we show how the resulting measures can be easily

The proposed procedure is explorative. It would perhaps be interesting to derive a procedure from an explicit model such as a model with latent networks that are measured with error or with item-specific contributions, like in Factor Analysis or other latent variable models. However, we fear that such a procedure would be quite cumbersome and subject to a variety of assumptions that are difficult to test, and therefore we restrict ourselves to what we think is a sensible explorative step-by-step approach.

In the following sections, we first describe the procedure in general terms, and then provide an example about the dimensions of peer perception in Hungarian high schools. We arrive at a clustering that represents the similarity between the networks in all groups to a reasonable extent. We highlight a number of decisions that need to be made by the researcher during the analysis, taking into account the context of the study. Some practical considerations and issues regarding future developments are discussed in the last section. We hope that the proposed analytical strategy may be useful for researchers in finding the main dimensions of relationships in a given social context, as a preliminary step for analyzing highly multiplex network data with existing statistical models, and that it may also provide clues for the construction of valid and rich multiplex network measures.

2. Defining composite network measures: an empirical procedure

When analyzing social processes within a group, we may be interested in a multitude of networks – relations, perceptions, common activities, etc. – among its members. Having measured a variety of network dimensions, we face data of high complexity that is difficult to analyze, even with the most advanced network methods. This is further complicated when we would like to investigate and compare processes across several groups. Below we propose and demonstrate a procedure, utilizing a set of existing statistical methods, that can aid researchers in reducing the number of studied networks by the construction of composite network measures in multiple groups.

2.1. The problem

We assume that we have conducted a network study in G disjoint groups (communities, firms, schools, etc.) where the g^{th} group has n_g actors.[4] In each of the groups, we have measured the same set of K networks representing various types of relations between the actors. The observed groups are likely to be heterogeneous in terms of network structure and the interrelations between network dimensions. However, we assume that there are inherent similarities between them, so that it is meaningful to look for consistent patterns across groups. We further assume that the number of measured network dimensions, K, is large enough so that the data cannot be analyzed with the multivariate methods mentioned above (practically, much larger than four).

We propose a procedure to reduce the complexity of such network datasets by identifying clusters of networks such that networks in the same cluster are similar and networks in different clusters are dissimilar; therefore it may be called a *cluster analysis of networks in a multiplex network data structure with multiple groups*. The procedure serves a similar purpose as Factor or Cluster Analysis of monadic variables: starting with a relatively large number of network items in a sample of groups, it results in a smaller number of composite networks which may capture the latent relational dimensions underlying the data better than the original items. (While here we focus on the case of binary networks, the procedure can be applied equally well to weighted networks.)

The five (plus one) steps of the procedure represent a strategy that is natural for Cluster Analysis in general. Here we briefly go through the steps and then elaborate on them in the following sections. Throughout the paper the terms “network”, “network item”, “item”, or “network dimension” are used interchangeably to refer to a specific instance of the K measured networks.

- Step 0. Before starting the analysis, it is worthwhile to think and state explicitly the aim of the application, initial hypotheses about the clustering of the K network items, etc. This step can help in making decisions later in the procedure. A few basic questions are collected below.
- Step 1. The set of K networks in G groups in our case have the same role as objects in a Cluster Analysis or items in a Scale Analysis. Accordingly, we choose a measure of pairwise similarity between the observed networks in each group – some suitable measures are presented below. For each of the G groups, this step results in a $K \times K$ similarity matrix.
- Step 2. The consistency of similarities across the groups are assessed, and network dimensions with highly inconsistent relations to the rest of the networks are dropped from further analysis. This is in line with the assumption that the observed groups are similar with respect to the network dimensions that are used for the clustering, and with the aim of finding similarities that are generalizable across groups.
- Step 3. Network dimensions are clustered using their pairwise similarities, with the aim of finding a classification that is appropriate for all G groups. The procedure proposed here is ordinal: it is based on only the rank order of the similarities.
- Step 4. The fit of the global solution is tested in each group.
- Step 5. Assuming that a well-fitting clustering solution has been found, it is used to combine the network dimensions belonging to the same clusters to create composite network measures.

2.2. Step 0: Thinking

Prior to the analysis, it is advisable to consider questions such as the following, as they may assist in making some of the decisions in later steps of the procedure. The answers should be based on theoretical considerations and prior empirical evidence. First of all, what is the aim of the study? What do we want to use the composite networks for? What should their ties represent? Given the aims, what are the best candidate measures of the similarities between the network items? Then, does it seem meaningful to compare all of our items? Are there some which are substantively not comparable with the rest? If the comparison seems appropriate, can we come up with an *a priori* intuitively acceptable grouping of the items? Are the groups themselves comparable, or are there systematic differences between them? Do they represent similar social settings? Finally, have others found something of relevance about the association between these or similar network items in similar groups?

2.3. Step 1: Measuring network similarity

The first step is to decide on how to measure similarity. This choice primarily depends on the level of analysis that is the focus of study: dyads or actors. The Jaccard index (Jaccard, 1908; Batagelj and Bren, 1995) and the simple matching coefficient (Sokal and Michener, 1958, Wasserman and Faust, 1994:370) can be used to

and meaningfully combined to form weighted (ordered) networks. Other, perhaps more elaborate ways of weighting are not discussed in this paper but may be possible and useful.

4 The paper focuses on studying multiple groups, where $G > 1$. A single case study, $G = 1$, is also meaningful, but it renders most of the non-trivial steps (Steps 2–4) explained in this paper unnecessary. We also present substantive arguments for studying multiple groups at the same time whenever possible in Section 5.

quantify similarities between two networks on the level of dyads. In- or outdegree correlations are simple measures of actor-level similarity between two networks, as we have seen in the examples from adolescent research (for a more recent application, see Szell et al., 2010).

Depending on the researcher's interest, further measures may be defined and used, e.g., measures of similarity between block structures. However, note that not all measures are equally appropriate for every research question, and the chosen measure has to fit to the aim of the application. This issue is discussed further in Sections 1.3, 2.2, and 5.

As our present aim is to combine information from several network items on the level of ties, we use a dyadic similarity measure, the Jaccard index. The Jaccard index between networks k and m ($k, m \in \{1, \ldots, K\}$) in a given group can be formulated as

$$J_{km} = \frac{n_{km,11}}{n_{km,11} + n_{km,10} + n_{km,01}}, \tag{1}$$

where

$$n_{km,ab} = \sum_{i,j} I\{x_{kij} = a, x_{mij} = b\} \quad (a, b \in \{0, 1\}; \quad k, m \in \{1, \ldots, K\}), \tag{2}$$

where $I\{A\}$ is the indicator function of set A and x_k and x_m are the adjacency matrices of the two networks. E.g., $n_{km,11}$ is the number of ties present in network k as well as m. The value of the Jaccard index ranges between 0 and 1, with 0 showing that there are no overlapping ties and 1 that exactly the same pairs of actors are tied in the two networks. By calculating the Jaccard index for each pair of the K networks, we get a $K \times K$ matrix of pairwise similarities for each of the G groups in our dataset.[5]

2.3.1. Projecting and visualizing the similarity structure

Before starting any further analysis, it is advisable to visualize the similarity structure of the networks in each group. A good representation may give a preliminary idea about the main dimensions of relations between the network dimensions. In addition, comparing images from each group can inform us about between-group differences that should be taken into account in the later steps. Since the starting point was that we have measured more than just a few networks, a comprehensible visualization should be preceded by the projection of the pairwise similarities into a few dimensions.

A method suitable for such a problem is Ordinal Multidimensional Scaling (Ordinal MDS). The aim of MDS in general is to place items (in our case, networks) in a low-dimensional, preferably 2D or 3D, space so that the pairwise similarity matrix reproduced from the location of items in this space represents the original similarity matrix as well as possible. In Ordinal (or non-metric) MDS the aim is to produce a solution in which the rank order of similarities optimally matches the rank order of the Euclidean distances between the items; the numerical value of the reproduced similarity measure is not important. The result of the MDS, the coordinates of items in the projection, can be used to visualize pairwise similarities. For more detailed descriptions of MDS see Bartholomew et al. (2008), Chapter 3 and Cox and Cox (2010).

2.4. Step 2: Checking the consistency of similarities across groups

For the set of network items to be clusterable, their similarity matrices across the various groups should show a consistent pattern. This means that, roughly, in different groups, the same pairs of items should be similar and the same should be dissimilar. Items having an erratic pattern in their similarities to other items would disturb the clustering process and should be left out. How to assess the consistency of this pattern? If an item is identically related to the other items across all groups, the rank order of similarities between this item and the other items will be the same across groups. Thus, for item k, we are considering row k of the similarity matrices for all groups, and we require the rank order of the cells in this row to be similar across the groups. In other words, if r_{gkm} denotes the rank of network m in the order of similarities to k for group g, then the rows $(r_{gk1}, \ldots, r_{gkK})$ should be approximately equal in all groups g.

To quantify the consistency of network similarities across the groups, we apply a nonparametric measure, Kendall's coefficient of concordance, or Kendall's W (Kendall and Babington Smith, 1939; Legendre, 2005). The level of concordance of a network k is calculated by comparing the rank orders of the similarities of the remaining $K-1$ networks to k in the G groups. If a network item is perfectly consistent in its relation to the rest of the items, that is, the order of items based on their similarity to the network in question is the same in all groups, Kendall's W takes the value of 1. Smaller values indicate lower consistency with the index being close to 0 if the order of similarities is basically random across the groups.

Formally, we can write Kendall's W for a given network k in G groups as

$$W_k = \frac{12\sum_{m \neq k}^{K}(R_m - \bar{R})^2}{G^2((K-1)^3 - (K-1))}, \tag{3}$$

where G denotes the number of groups across which we compare the rank order of similarities, K is the number of network items (including k), R_{km} is the sum of the ranks of the similarity between network k and another network m over all G groups,

$$R_{km} = \sum_{g=1}^{G} r_{gkm}, \tag{4}$$

and $\bar{R}$ is the average sum of ranks given the number of groups and networks (note again that K includes k),

$$\bar{R} = \frac{1}{2}GK. \tag{5}$$

For each network item k, we calculate Kendall's W_k and inspect their values. We use these values not to test a null hypothesis, but to assess descriptively the homogeneity of the groups with respect to how this item relates to the other items. If many items have a very small W_k (though there is no strict rule for what counts as "small"), then the G groups are very heterogeneous in how the networks relate to each other, and chances for finding a clustering that would work well in all groups are low. However, it may be possible to make similarities more consistent by dropping from the analysis network items with low W; but this will affect the W_k values between the remaining networks. Therefore we do it in a recursive fashion: in each step drop the network with the lowest W, then recalculate the coefficients, now without the excluded network. After dropping a small number of items, it may be possible to obtain a restricted set of networks, all of which have a sufficiently high consistency measure.

Deciding on a stopping rule in this procedure is a qualitative decision that does not only depend on the values and the distribution of Kendall's W, but also on judgment about item content. The conclusion of this step in the analysis might be that finding a clustering of networks that is valid in all groups is not feasible. In such a case, a different strategy might be necessary (for example,

[5] Some of the statistical techniques applied below take distances rather than similarities as an input. For sake of clarity, we only use "similarity" throughout the arguments, but note that the two concepts carry the same information. For instance, the Jaccard distance can be expressed simply as 1 − *Jaccard index*.

involving the classification of the groups based on the clustering of the items in each of them), which we do not treat in the present paper. We assume that at the end of Step 2, a set of items was obtained with sufficient homogeneity.

2.5. Step 3: Finding group-level and global clusters in multiple groups

As a result of the previous steps, we now have similarity matrices, raw and reproduced from Ordinal MDS, between a potentially reduced number of K' networks in all G groups. Although the similarity structures will usually still differ between groups, these differences were reduced in Step 2 so that the groups may be regarded as being more homogeneous. In the current step, we explore the cluster structures of the network items in each group separately. For doing so, we can apply a suitable method of Cluster Analysis on the similarity matrices. Here we use Ward hierarchical clustering which has been found to produce well-interpretable solutions in various situations (Bartholomew et al., 2008:24). Ward's minimum variance method is an agglomerative hierarchical clustering procedure. In our example the method starts with K' clusters (each network item individually) and groups them one by one so that in each step the two clusters having the lowest between-cluster variance are merged.

In this step, we also attempt to define a cluster structure that will be a candidate for a global grouping. There may be several ways for doing this, but we have found it useful to simply calculate the mean similarity matrix from the group-wise similarity matrices, and use the clusters obtained from that as a potential global solution. Note that this is justifiable only if the group-level clusterings were relatively similar to each other – which we have promoted by dropping inconsistent network items in Step 2.[6] In the next section we provide an option for assessing how well the global cluster structure fits the group-level ones.

2.6. Step 4: Checking the fit of the global solution

Our aim is to find a single clustering of network items that reasonably represents the similarity structure in all groups. Therefore, before we use the global clustering as a basis for creating composite networks, we need to assess how well it fits the cluster solution of each group separately. First, a visual inspection is presented comparing the group-wise similarities within clusters to those between clusters; ideally, for all groups the former should be high and the latter should be low. Second, we use the Rand index, a measure for comparing two cluster solutions based on whether they group each pair of items similarly (Rand, 1971; Hubert and Arabie, 1985). The Rand index for comparing a global clustering with the clustering obtained individually from group g can be formulated as

$$R_g = \frac{\text{no. of agreements}}{\text{all pairs of networks}} = \frac{a+b}{\binom{K'}{2}}, \qquad (6)$$

where a and b are the number of agreements, that is, the pairs of network items that belong to the same cluster (a) or to different clusters (b) in both solutions and $\binom{K'}{2}$ is the number of pairs that can be made of our K' network items. The Rand index takes the value 1 if cluster memberships in the two compared solutions are identical, and 0 if there is no pair of items that are grouped the same way in both clusterings. The Rand index thus can be used to measure how similar the group-level clusterings are to the global structure we consider imposing on them. If the fit is satisfactory, we may define the same set of composite network measures in all of our groups. If not, we might drop some of the groups or some of the network items, and restart the whole procedure.

2.7. Step 5: Defining composite networks

The final step in the procedure is to construct a small number of binary networks as a combination of network items that were grouped together by our analysis. There may be several ways of doing so, but here we apply an approach that can be regarded as taking a "relaxed" union of ties in networks that belong to the same cluster. We say a tie exists in a composite network belonging to a cluster of items if the given tie is present in at least t of the constituent networks. If $t = 1$, this defines the union; if t is equal to the cluster size, it is the intersection of the networks in the cluster; but t could also be in between the extremes. There may be more than one good solutions of this step, and it is up to researchers to determine the value of t that best fits their substantive interest. The choice may be based on the number of items in each cluster, prior knowledge about the density of the composite network dimension, expectations about its similarity (e.g., as measured by the Jaccard index) to other networks external to the ones that were used to obtain the clusterings, etc. The defined composite networks then can be used as separate network dimensions in the analyses of substantive questions. An overview of the whole procedure suggested here is presented in Fig. 1.

3. An application: peer-perception dimensions in Hungarian high schools

To demonstrate the procedure proposed above, we examine two waves of multiplex social network data from first-grade high-school classrooms in Hungary. The full dataset contains self-reported information about affective relations (*friendship*, *liking*, *dislike*, *hate*), shared activities, attribute, behavior and role perceptions, and bully-victim relations between the members of 43 classrooms. Here we use 21 interpersonal perception networks from a random subset of 18 school classes (see Appendix B), for the possibility of future cross-validation. To these 21 network dimensions in 18 groups, we apply the five-step procedure outlined above. The analysis identifies three main relational dimensions in the data: positive attributions, social role attributions and negative attributions. We use these three clusters to construct four composite networks representing these latent dimensions, and show that they significantly complement the information obtained from standard measures of affective relations, such as friendship or dislike.[7]

[6] One could think about weighting the groups based on, for example, their size. When group sizes are different, however, one must address if (and why) the new information added by each group is proportional to group size. We think this is a question by itself. When groups do not vastly differ in size (which is the case in the analyzed sample), it may be better not to weight them, but honor the contribution of each group equally.

[7] The data analyzed here, along with the *R* scripts which help to reproduce the study (or apply the procedure to different datasets) is publicly available at https://www.stats.ox.ac.uk/~snijders/CAMN_scripts_data.rar [anonymized for review, materials provided along with the manuscript]. The analysis was carried out in version 3.2.0 of the *R* environment. For calculating Kendall's *W* we used the *kendall* function in the *irr* package (v0.84). Multidimensional Scaling was done by the *isoMDS* function in package *MASS* (v7.3-40). Hierarchical Cluster Analysis results are from the *hclust* function with Ward's method of package *MASS*, applied to Jaccard distances. The Rand index was calculated by the *randIndex* function in the *flexclust* package (v1.3-4). In addition to the base *R* functions for plotting, we relied on the functionalities of the packages *ape* (v3.2) (for cluster trees), *vioplot* (v0.2) (for violin plots), *venneuler* (v1.1-0) (for Venn diagrams) *lattice* (v0.20-31), and *gridExtra* (v0.9.1) (for heatmaps).

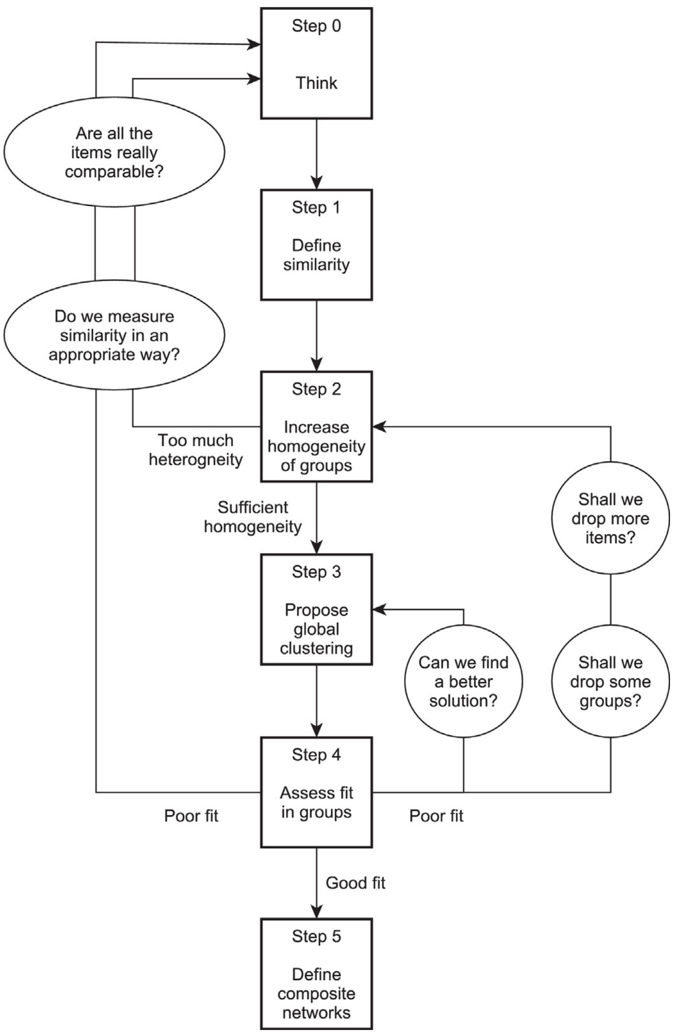

Fig. 1. An overview of the procedure for the cluster analysis of multiplex networks.

3.1. Data

The data we analyze is from a longitudinal network survey conducted in Hungarian high-school classrooms in the period of 2010–13. The data was collected as part of the "Wired into each other" network study by the Research Center for Educational and Network Studies (RECENS).[8] In the current study, we use the first two waves of the RECENS dataset. These observations frame an intense six-month period from the first year in high school when classroom members become familiar with their new environment, get to know each other, and start to develop shared views about which behaviors or opinions are acceptable in the class and which are to be rejected.

The evolution of the value systems (as used in Coleman, 1961) of classrooms goes hand in hand with changes in adolescents' perceptions about their classmates, and may also affect the relationships between the different dimensions of peer perceptions in the community. As a consequence, the comparison of our results from the two waves may highlight certain general tendencies in the ongoing social dynamics.

For our exploratory analysis, we selected a random subsample of 18 classrooms so that each school, classroom training program (vocational, technical, or grammar),[9] and town size (capital, large town and small town) is independently represented by at least 2 classes (see Appendix B). These contextual variables capture regional and school-level differences and the systematic variation in students' background between training programs (Dobos et al., 2011), which may all be related to the structure of perceptual dimensions in the classrooms.

The size of the studied classrooms varies between 26 and 38 students, with an average of 32 over the two waves. Girls are in a majority in most of the classes (14 out of 18), their average proportion is 63%. This is higher than the country-level gender ratio in high-schools at the time of the study (KSH, 2011). The mean age of students was 15.3 years at the time of wave 1 (with an average within-class standard deviation of 0.57 years). The average response rate per classroom is 90% in wave 1 and 85% in wave 2. Further background information about the RECENS high-school dataset is provided in Appendix A; for a more detailed description of the analyzed subsample, see Appendix B.

3.2. Step 0: The network items and their expected grouping

Our analysis focuses on those network items from the RECENS dataset which are related to students' perceptions about their classmates. Specifically, we examine the similarities between 21 networks measuring different perceived attributes, behaviors, and social roles in the 18 classrooms.[10] Table 1 lists the network items we use in the current study: the translation of their original formulation in the questionnaire and the short labels by which we refer to them in the following. Further, it indicates if items identical or similar to ours can be found in the psychological tests we discussed in Section 1.3, namely PNI (Wiggins and Winder, 1961), RCP – the Revised Class Play (Masten et al., 1985), PEI (Pekarik et al., 1976), or ASSESS (Prinz et al., 1978).

As can be seen in Table 1, many of the perceptual dimensions we study here show up in some form in the earlier sociometric measures. Some of our network items are related to the Sociability/Likeability (*funny*, *clever*, *kind*), Withdrawal (*shy*), Hostility (*smug*), Aggression (*gossipy*, *fights*), and Social Competence (*argues*, *protects*, *dispute*, *decides*, *organize*) factors of the mentioned tests.

However, there are also quite a few items which do not have a precedent in these studies. The differences can be explained by the sociological focus of the RECENS dataset: perceptions connected to dyadic trust, social roles/leadership, social status, and physical attractiveness were important network dimensions in this study. Based on these earlier findings and the substantive content of the items, we can propose an a priori classification for the 21 analyzed items. The expected grouping may or may not be confirmed by the data, but it provides guidance in the decisions we need to make during the investigation.

When thinking about clusters among our network items, we have to keep in mind that unlike the cited psychological studies, we focus on the measurement of networks and not of individuals. This leads to important differences in interpretation. For example, let us assume that the items *teacher's pet* and *organize* have a high

[8] The survey was supported by the Hungarian Research Funds (ref. no. K81336).

[9] In the Hungarian public education system, secondary schools often run classrooms with different training programs in a single grade.

[10] The analysis was also replicated with the two positive and two negative affection networks included. The positions of the *friendship* and *liking* network items were ambiguous in the similarity structure depending on whether the clustering was based on raw similarities or those reproduced from the MDS projection. This suggests that the overlaps between *friendship*, *liking*, and the rest of the networks should be examined separately, and so these two items were dropped from the procedure. The inclusion or exclusion of the *dislike* and *hate* items did not alter the rest of the results, and for consistency, these were also omitted from the analysis reported in the paper.

Table 1
Description of the analyzed perception network items.

	Label	Item description	Expected grouping	Test with similar item
1.	funny	He/She is funny	Positive attributes	PEI, PNI, RCP
2.	clever	He/She is clever	Positive attributes	PEI, PNI
3.	kind	He/She is kind	Positive attributes	PEI, PNI
4.	shy	He/She is shy, reserved	Negative attributes	PEI, PNI, ASSESS
5.	smug	He/She is stuck-up, smug	Negative attributes	PEI, RCP
6.	gossipy	He/She is gossipy	Negative attributes	PEI, PNI, ASSESS
7.	fights	He/She often gets involved in fights	Negative attributes	PEI, PNI, ASSESS
8.	nerd	He/She is a nerd	Negative attributes	–
9.	t. pet	He/She is a teacher's pet	Negative attributes	–
10.	look down	I look down on him/her	Negative attributes	–
11.	argues	He/She dares to argue with teachers	Leadership	ASSESS
12.	protects	He/She protects the weak	Leadership	ASSESS, RCP
13.	dispute	He/She would be able to solve disputes	Leadership	ASSESS, RCP
14.	decides	He/She decides what we do after school	Leadership	ASSESS, RCP
15.	organize	He/She could organize a class trip	Leadership	ASSESS
16.	look up	I look up to him/her	Leadership	–
17.	trust	I trust my secrets with him/her	Trust	–
18.	help	I could ask him/her for help	Trust	–
19.	pretty	He/She is handsome/pretty	Attractiveness	–
20.	would date	I would date him/her	Attractiveness	–
21.	money	He/She has a lot of money	Social status	–

Abbreviations: PNI – Peer Nomination Inventory; RCP – Revised Class Play; PEI – Peer Evaluation Inventory; ASSESS – Adjustment Scales for Sociometric Evaluation of Secondary-School Students.

indegree correlation but a low Jaccard index in one of the classrooms. From the former fact only, we would conclude that students who are despised for trying to have too good relations with teachers are nonetheless appreciated by their peers as efficient organizers – perhaps because it is easier to get things done in school for people who are recognized by teachers – and we might be led to a conclusion about ambiguous dyadic perceptions, negative on the one hand and positive on the other. However, the low Jaccard similarity between these items would alter the interpretation by showing that the perceptions of teacher's pet and good organizers are held by different sets of classmates, differentiated perhaps by their school attitudes.

Thus, we can say that dyadic similarity measures, by retaining more information available in the networks than degree-level measures, take into account the structure of the studied groups, the fact that communities can contain subgroups that may have quite different opinions about certain peers. This further underlines that unlike the earlier studies, we are not trying to identify typical student profiles based on received nominations, but rather types of perceptions that usually go together. Such an approach will present a natural way for constructing aggregated network measures.

Based on earlier findings and intuition, we can assign our network items into a few groups. The first three networks (*funny*, *clever*, *kind*) clearly represent perceptions about general positive characteristics of social behavior. We may call this group positive attributions. The following seven items in Table 1 capture negative perceptions about peers (*shy*, *smug*, *gossipy*, *fights*, *nerd*, *teacher's pet*, and *look down*), hence we label them as negative attributions.[11]

The third, larger group, titled leadership, contains perceptions that are related to being active, independent, a leader in social situations (*argues*, *protects*, *dispute*, *decides*, *organize*, and *look up*). The network items *trust* and *help* form the fourth group, and they may be understood as two specific dimensions of general trust in dyadic relations. As a fifth category, there are two more items (*pretty* and *would date*) which represent the dimension of romantic attractiveness. The last perception network, *money*, is not similar to any of the other items as it represents impressions about the social standing or status of peers.

[11] It is to be noted here that the classification of items, just as in any application of Cluster Analysis, is based on multivariate similarities. This means that it is not necessary for all pairwise similarities within groups to be equally high, only higher than between groups. As an example, *shy* and *fights* being in the same group is not a contradiction, neither substantively nor technically. Choosing an appropriately high combination threshold (see Section 2.7) ensures that the interpretation of ties in the composite networks will not be ambiguous.

3.3. Step 1: The similarity structure of peer-perception networks

First we calculate the Jaccard index between each pair of the 21 networks in every classroom using the formula from Section 2.3. The values observed in the 18 classes range from 0 (wave 1–2) to 0.53 (wave 1) and 0.66 (wave 2), with a mean of 0.11 and 0.12. These numbers tell us that the pairwise overlap between the items in the sample is typically low, but varies greatly depending on the classroom and the items compared. In the second step of the procedure, we will aim exactly to identify those network items that contribute most to the variability between classrooms, and only then we shall turn to examine the patterns in between-item similarities. Here we do not further describe the raw Jaccard index values between our networks.

Before we proceed, however, it is worthwhile to visualize the average similarities using ordinal MDS, as described in Section 2.3. Fig. 2 presents the 2-dimensional projection at the two time points. Stress in both cases was below 15% which suggests the two dimensions provide a reasonable representation of the mean similarities (Bartholomew et al., 2008:65). Items are identified in the figure by their labels from Table 1 (and their colors represent their *a priori* grouping).

It is visible that most of our items categorized as negative attributes (labels in red) are well separated from the rest, with the sole exception of the *fights* item. On the other hand, positive attributes (green), leadership-related attributions (blue), and trust (light blue) are all generally close to one another, suggesting significant overlap between these networks. Only the *argues* and *decides* items are distant from their group in both waves.

The two items we associated with attractiveness are placed far from each other: *pretty* is close to the just described "busy area", while *would date* is on the periphery at both observations. The last item, *money*, is closest to *fights*, *argues* and *would date* in wave 1, but more distant from all items in wave 2.

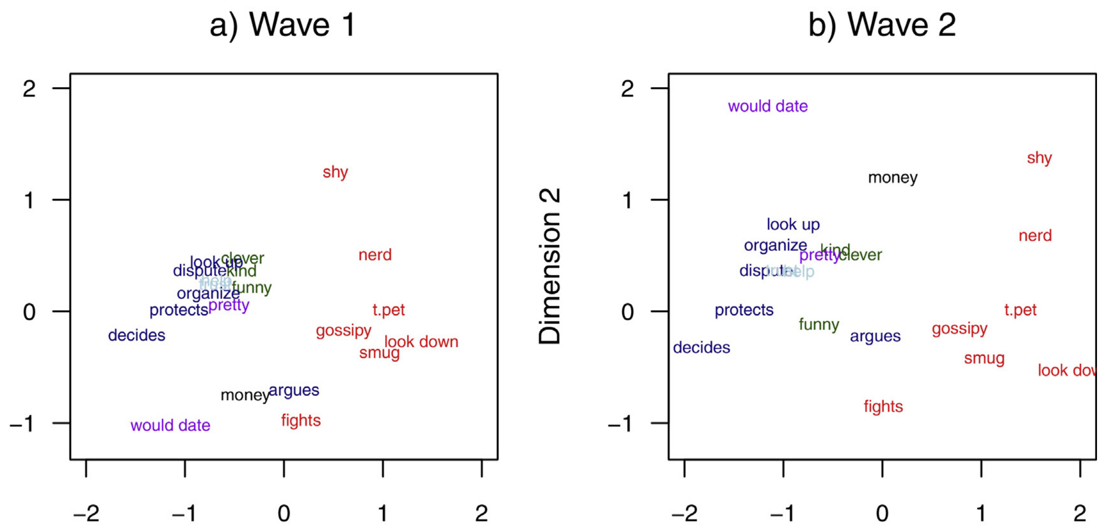

Fig. 2. Visual representation of average similarities between the 21 network items at wave 1 (a) and 2 (b). *(Coordinates obtained from the 2-dimensional MDS projection of mean Jaccard index values.)*

The inspection of the figure shows that our prior expectations for the grouping of networks are roughly correct, and suggests that we might find confirmation for the existence of at least two groups: one of negative attributes and one of positive and leadership attributions and trust. However, this first impression may not be entirely correct if the classrooms are highly heterogeneous with regard to the similarities between items – we shall turn to this issue in the following.

3.4. Step 2: The consistency of the similarity structure across classrooms

As described in Section 2.4, we use Kendall's W to assess how consistently each of the network items are related to the rest of the items across the classes.[12] Fig. 3a and b presents the distribution of W for all of the 21 networks in wave 1 and 2, respectively. The items on the left side of each plot have the lowest, those on the right the highest concordance coefficients. The vertical positioning of the labels serves only to emphasize the rank order of the W values. Label sizes are proportional to the average density of items in the classes (the class-wise densities of the networks are reported in Appendix A).

W ranges between 0.2 and 0.85 in both waves. The figures also show that most dense networks typically have higher W values. The correlation between average density and W is 0.65 in wave 1 ($p < 0.01$) and 0.53 in wave 2 ($p = 0.01$). A simple explanation for this tendency is that in case of sparse networks, chance may have a larger impact on the relative frequency of ties overlapping with other networks. However, we also see some sparse network items with larger coefficients (e.g. *trust*, *dispute*, *organize*, *protects*), which highlights that our choices of excluding items should not be solely determined by their density.

An acceptable distribution of concordance coefficients can be reached after dropping four network items: *fights*, *money*, *argues*, and *would date*. The distribution of Kendall's W after stepwise excluding these networks is shown in Fig. 3c and d. At this step, all W values are larger than 0.4, which is a notable increase from 0.2. Some remaining items have a higher W in the reduced item set than in the initial set.

[12] The W values reported here are corrected for ties in rank orders. However, this makes little difference, since exactly equal Jaccard indices are rare.

Note that the stepwise nature of the procedure redefined the position of *argues* in wave 2 with respect to the other items: initially it had the fifth smallest W value (Fig. 3b), but in the fourth step it was removed (and *shy* retained). Further, the correlation between average density and Kendall's W in the reduced item set is 0.60 in wave 1 ($p < 0.01$) and 0.43 in wave 2 ($p = 0.08$), smaller than in the case of all 21 networks. This shows that the removal affected the networks with the lowest density, those most prone to chance overlap with the other networks.

Dropping more than four networks generates worse solutions than the ones presented here. The items we classified as negative attributions are next on the list for potential exclusion, but excluding some of them largely decreases the consistency measures of the rest. The next natural stopping point in the procedure is where all networks from the negative attribute group are removed. We find this an unnecessary loss of substantial information. Finally, we draw attention to the fact that the four removed network items are the same which were placed far from the others and which were positioned quite differently for waves 1 and 2 in Fig. 2. Based on the presented results, we continue the analysis with the remaining 17 networks.

3.5. Step 3: Three dimensions of peer perceptions: the global solution

We apply Ward Hierarchical Cluster Analysis to study the structure of similarities among the remaining network items (see Section 2.5). This is done for each classroom separately and then for a similarity matrix containing the average Jaccard index for each pair of items. Fig. 4 presents the cluster trees based on the mean similarities at the two observations. The colors highlight the cluster membership of items in well-fitting solutions for wave 1 (a) and wave 2 (b).

Three clusters seem appropriate in the former case, five in the latter. However, it is apparent that the groupings are almost identical at the two time points. At the top of the figures, four items form a group in both waves: *pretty*, *funny*, *clever*, and *kind*. This cluster is similar to the group of positive perceptions from Section 3.2, with the addition of the *pretty/handsome* item. We shall refer to this cluster of network items as *positive attributions*.

At the bottom of the plots, five items are grouped together in wave 1: *nerd*, *shy*, *teacher's pet*, *gossipy*, *look down*, and *smug*. This cluster may be labeled as the group of *negative attributions*.

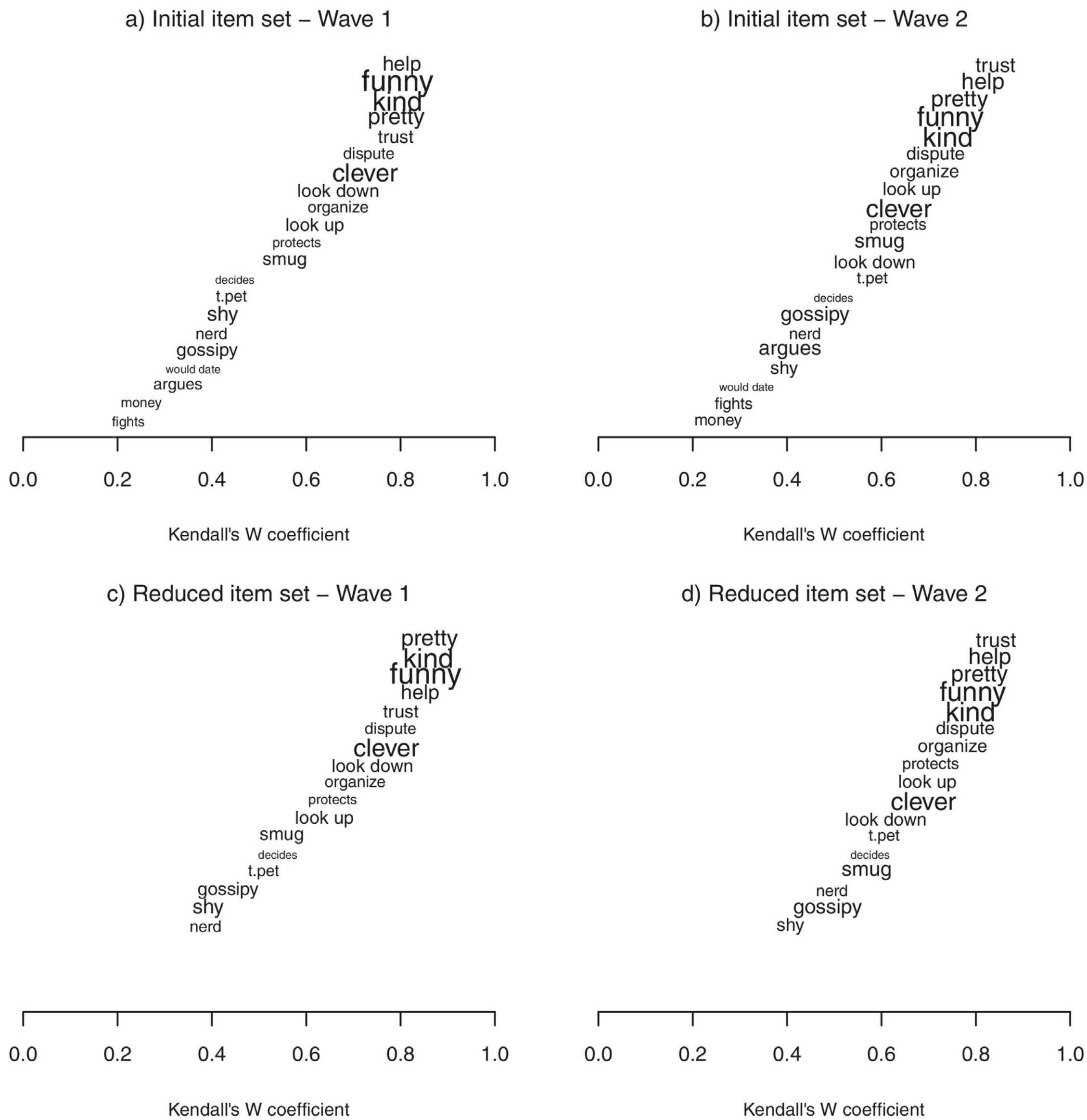

Fig. 3. Distribution of Kendall's W for the full and reduced item set in wave 1 (a, c) and wave 2 (b, d). *(W values were calculated based on 18 classes; the horizontal axis measures Kendall's W, while the items are similarly ordered along the vertical axis; label sizes are proportional to the average density of items in the classrooms.)*

Although *nerd* and *shy* emerge as a separate cluster in wave 2, they are still clearly close to the other negative items.

Based on the first observation, the remaining seven networks form a single cluster. This group contains most of the items we referred to earlier as attributions of some kind of social leadership (*protects*, *decides*, *dispute*, *organize*, *look up*), but also those related to dyadic trust (*trust*, *help*). We may thus use the more general term of *social role attributions* for the items in this cluster.[13]

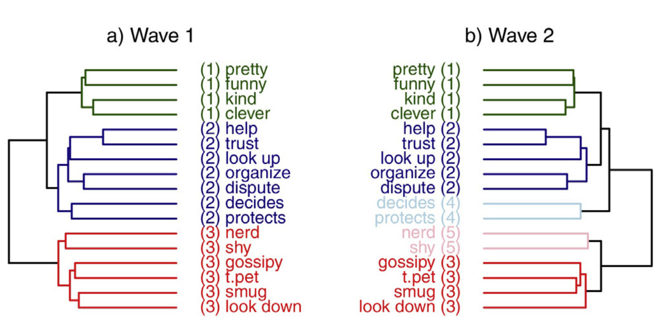

Fig. 4. Cluster trees from the average similarities in wave 1 (a) and 2 (b). *(The colors and numbers represent the resulting clusters.)*

In wave 2, *protects* and *decides* are together, like in wave 1, but form a separate cluster. We treat this discrepancy between the waves as a nuance for two reasons. First, *protects* and *decides* are the sparsest networks in our sample (see Figs. 3 and 11) so that their similarities to other items may be more affected by small changes in network ties. Second, the two items still belong clearly to the groups of positive and role attributions.

It has to be emphasized that the overall similarity structure among our items is remarkably stable at the two ends of a six-month period, while the networks themselves do change to an appreciable extent (see Fig. 12 in the Appendix). Some ties are

[13] Note that the roles in our reduced item set all seem to bear a positive value for the community. Of course, social roles can be negative or ambivalent, as well. Consider the case of *fights* or 'argues with the teacher', items which we dropped due to their inconsistent position among the other items in the studied classrooms. Here we use the term "social roles" to indicate positive social roles, but different item pools may allow further distinctions between various types of roles.

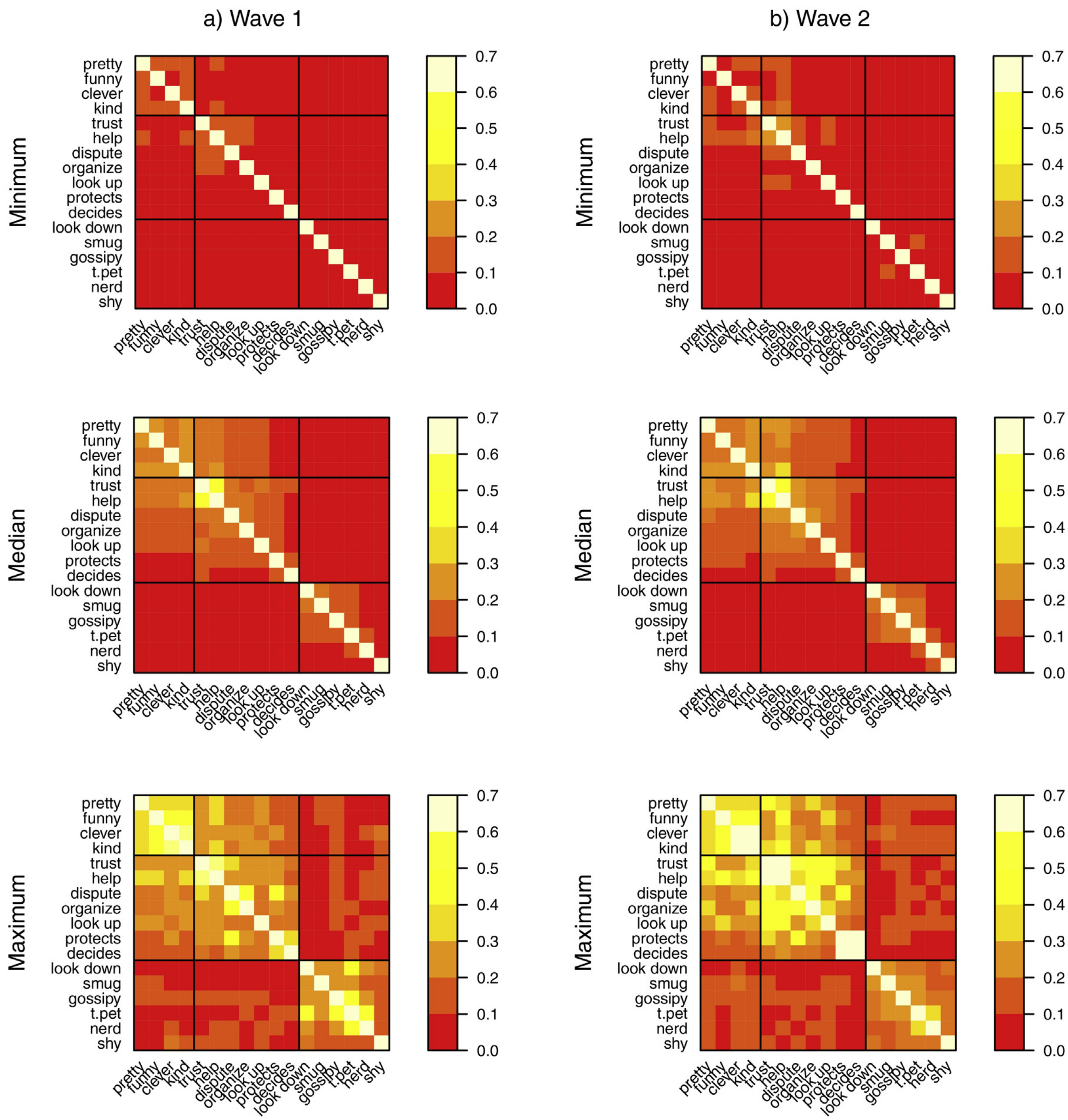

Fig. 5. The group-wise minimum, median, and maximum Jaccard index values among the 17 clustered items in wave 1 (a) and 2 (b). *(The diagonal values were set to the off-diagonal maximum of the six matrices. Note that in the six plots, the items are held constant, and so they provide an intuition about the distribution of the Jaccard index across the studied classrooms.)*

created, some are dropped between the waves, and the Jaccard similarities also increase and decrease. Our key finding is that the studied networks change in a way that their relative similarities remain essentially unchanged, leading to almost the same clustering solution.[14]

We use the three-cluster result from the first wave for both observations, in order to have a consistent basis for the definition of aggregated networks. However, before we proceed to define new network measures, we need to check if the clusters obtained from the mean similarities at the first time point are acceptable classifications for all of the classrooms in both waves.

[14] On the full sample of 43 classrooms, the resulting clusters at the two time points are identical.

3.6. Step 4: The fit of the global solution across classes

In this section, we assess how well the proposed three-cluster solution fits the individual classrooms. Note that in Step 2 of the procedure some network items were dropped in order to increase the homogeneity of classrooms. This means that we already did some work ahead to increase the chances for finding a well-fitting solution. There are numerous ways to measure the fit of a clustering solution. We present results from just a few simple tests.

First of all, we inspect the distribution of the pairwise similarities between the networks in the studied classrooms. If the proposed clustering is indeed reasonable for all or most of the groups, similarities should be high within clusters and low between clusters. The distribution of the Jaccard indices is represented in Fig. 5. The six heatmaps show the minimum, median, and maximum of the values observed between the items across the 18 classes at the two time points. Brighter cells refer to higher levels of similarity. The colors

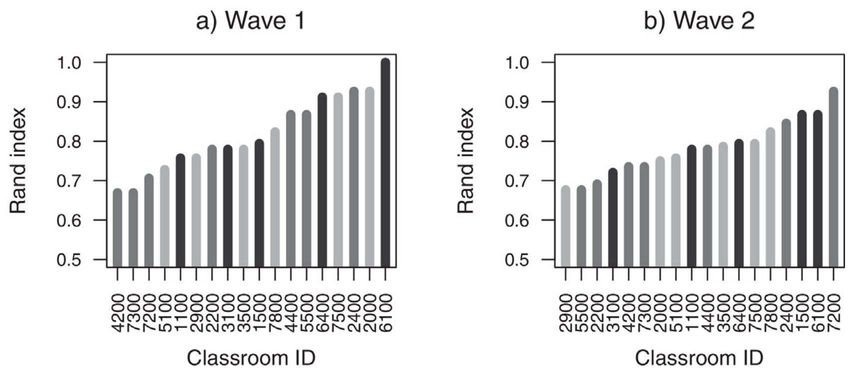

Fig. 6. Comparison of class-level clusterings to the global solution in wave 1 (a) and 2 (b). *(Classroom-level three-cluster solutions are compared with the clustering in Fig. 4a. The colors of the bars represent the three classroom training programs.)*

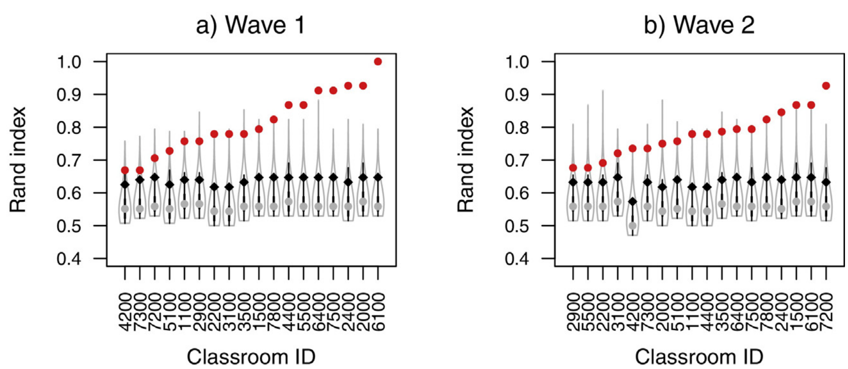

Fig. 7. A permutation test for the comparison of class-level and global clusterings in wave 1 (a) and 2 (b). *(Classroom-level three-cluster solutions are compared with the clustering in Fig. 4a. Null-distributions are generated by 10,000 random permutations of the cluster membership vectors in each case. Dots at the top represent the observed similarity values, the lower gray circles the means of the simulated distributions, and black rectangles the 95% critical values for one-sided t-tests.)*

are comparable across the six images in that the same color refers to the same degree of similarity in all of them. The minimum similarity gives the most unfavorable, the maximum the most favorable case for the comparison. The ideal pattern would be bright diagonal blocks and dark off-diagonal blocks.[15]

We can see in the figure that the three diagonal block areas marking our three clusters are generally brighter than the surrounding cells. Although there is considerable overlap between items in the positive and social role attribution clusters, the distinction between the two groups is clear (although *trust* and *help* might belong to either cluster). This provides some confirmation that the proposed clustering is a reasonable representation of the similarities in the 18 classrooms.

Next we compare he three-cluster solutions obtained per classroom with the three-cluster global solution proposed in the previous section. Fig. 6 shows the distribution of the Rand index (see Section 2.6) in the studied classrooms at the two observations. The results show that the global classification of networks is reasonably similar to those we can obtain directly from the class-level data. In all cases, more than two thirds of all pairs of items are similarly classified in the compared clusterings, and this is above 75% for more than half of the classrooms in both waves. The shade of the bars in the figure signal the training program of classrooms: light gray – vocational, medium gray – technical, dark gray – grammar. There is no significant tendency for classrooms of any training program to have a more "prototypical" clustering than the others.

The agreement expressed by the Rand indices can be tested by comparing the values found with those that would be found for random partitions with the same cluster sizes. Fig. 7 displays the Rand index values from the previous figure (the top row of data points), along with null distributions based on 10,000 random permutations of the clusterings in each classroom. It is clear that even in the worst cases, the observed values are higher than the 95% critical

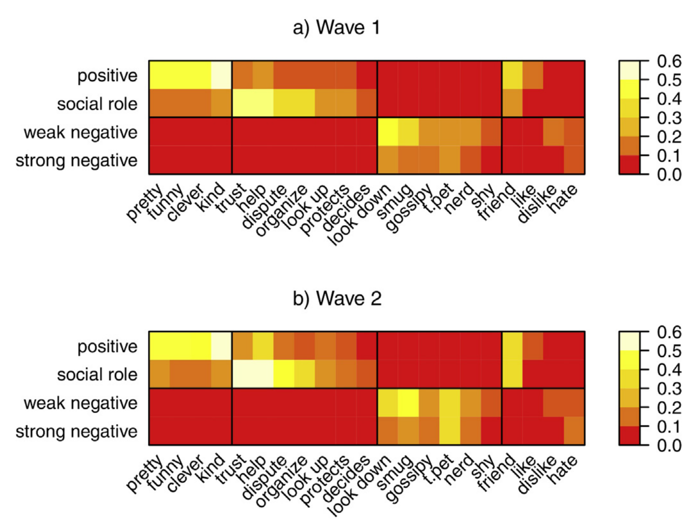

Fig. 8. Median values of the Jaccard index between the composite networks and the original and affective items in wave 1 (a) and 2 (b).

[15] The diagonal elements of the similarity matrices, that is the similarity of an item to itself, are obviously 1. However, for the purpose of this presentation they were replaced by the value of 0.7 which is the highest off-diagonal similarity, observed in the second wave (see the bottom right image).

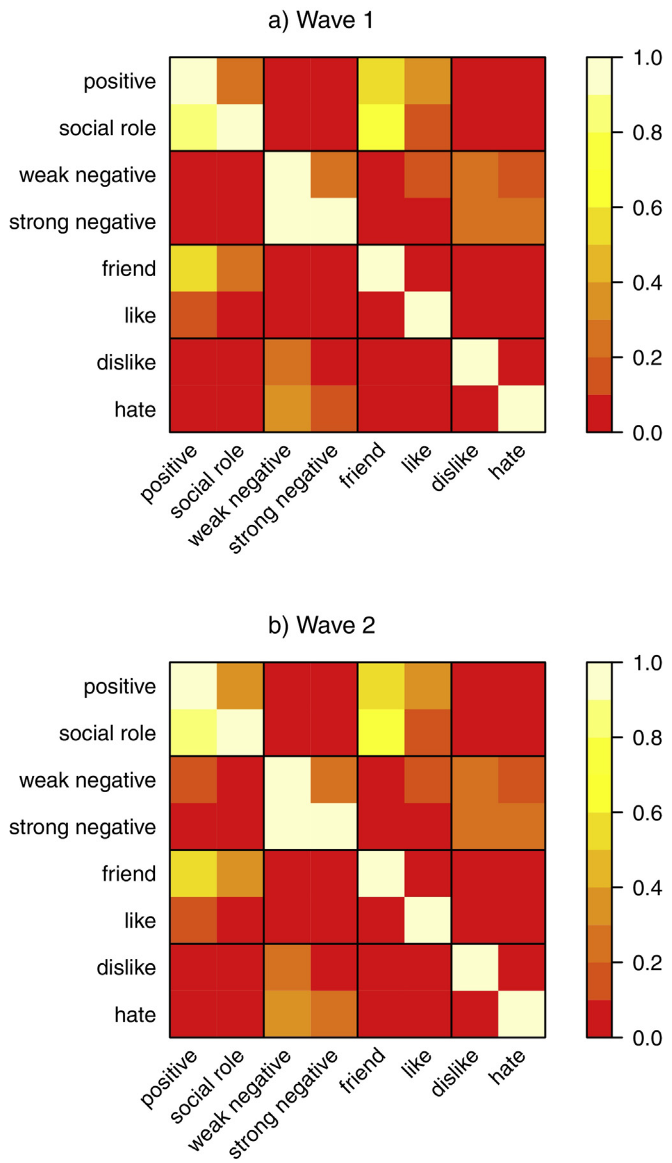

Fig. 9. Average overlap between the composite networks and the affective items in wave 1 (a) and 2 (b). *(Cell values represent the proportion of overlapping ties relative to the row networks density.)*

values of the null distributions (black rectangles). This means that our global solution performs better in all classrooms than similar but random clusterings.[16]

One can think of plenty of other ways for testing fit, especially considering the challenges of clustering in multiple groups. However, the focus of this paper lies elsewhere, so for the moment we are satisfied with the demonstrated evidence which suggests that our three-cluster solution is reasonable for the studied set of classrooms. We will return to certain issues of goodness of fit in the discussion.

3.7. Step 5: Defining positive, negative, and social role attributions

Now that we have established that the proposed three-cluster grouping is a reasonable representation of the similarities between the network items in the studied classrooms, we can turn to the definition of composite network measures. As described in Section 2.7, we use a simple "at-least-*t*-ties" rule to combine the tie-level information present in the networks of a given cluster. A practical question here is the choice of the threshold *t*. A too low value of *t* may lead to a too general interpretation of ties in the resulting aggregated network. For example, if we take the union of ties ($t = 1$) from the networks in our negative attribution cluster, then a tie in our composed network could mean a perception that the receiver is either shy or smug – clearly very different attributions. Increasing *t* reduces the prevalence of these problems. However, with a *t* too high the definition of a tie may become too strict, and the density of the composed network too low. The possibility of measurement errors also advises against extreme values of *t*. Clearly, there is no universal recipe for choosing a combination threshold, and it is up to the researcher to decide on its appropriate value.

For the present dataset, we propose the definition of four composite networks: one of positive, one of social role attributions, and two variants for negative attributions. In case of the first two item groups, we choose the combination threshold $t = 2$. For reasons discussed below, two networks are defined for the cluster of negative attributions, one at a threshold of 2 and another at 3. Thus the latter network contains only a subset of the ties in the former, and represents stronger negative attributions. The definitions of our four composed networks are summarized in Table 2. The two negative networks could also be transformed to one weighted network, of weak and strong negative attributions.

To have a first impression about the new measures and to be reassured that the clustering procedure was successful, Fig. 8 presents two heatmaps similar to those used above. Here, however, the four composite networks are compared with the 17 items used in the cluster analysis and with the four affective networks available in the dataset. The brightness of the cells in the figures is based on the median Jaccard index values between each row and

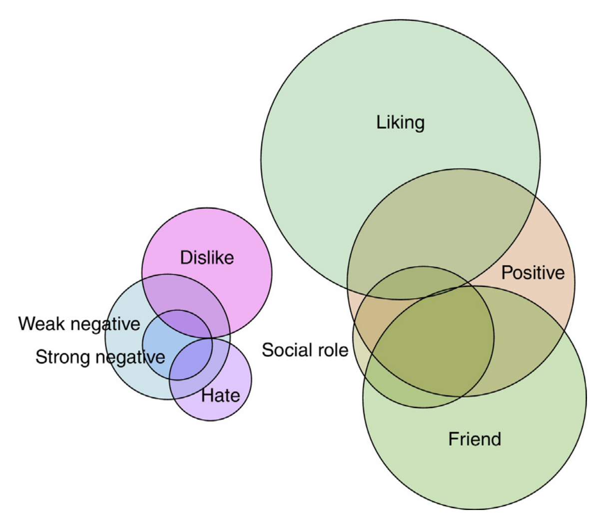

Fig. 10. The structure of relations between the aggregated and affective network measures. *(The sets were placed algorithmically to best represent the overlaps observed between the networks; the area of each set part is proportional to the frequency of ties that fall within; the overlaps between the four positive and four negative networks were disregarded; calculations are based on the average distribution of network ties across the 18 classrooms in wave 1.)*

[16] We also assessed in the same way whether the five-cluster solution from the mean Jaccard values in wave 2 (Fig. 4b) is a better proposal for the classrooms in both waves. The overall levels of fit are approximately equal in both cases, but there are significant classroom-level differences. This is expected as the two solutions presented in Fig. 4 are very similar; but this simple test provides further support that the three-cluster structure we impose on every classroom is close to a (locally) optimal solution given the data at hand.

Table 2
The definition of the four composite networks.

Composite network	Combination rule	Group of original networks
Positive attributions	At least 2 ties from	pretty, funny, clever, kind
Social role attributions	At least 2 ties from	trust, help, dispute, organize, look up, protects, decides
Weak negative attributions	At least 2 ties from	look down, gossipy, smug, nerd, t.pet, shy
Strong negative attributions	At least 3 ties from	look down, gossipy, smug, nerd, t.pet, shy

column item over the groups. The main results discussed here are found in every studied classroom.

The figures confirm that the composed networks have the highest overlap with the network items based on which they were defined – this shows that the success of the clustering procedure. Further, one cannot miss the minimal overlap between positive and negative items. Finally, it has to be highlighted that positive and social role attributions tend to coincide with positive affections (friendship and liking), while to some extent this is true for negative attributions and negative affections (dislike and hate). We study these connections in more detail in the following section.

3.8. *Comparing the aggregated attribution networks with affective relations*

How are the defined composite networks related to the more common relational measures in the field? This question needs to be answered in order to assess how the inclusion of these (or similar) new network dimensions in future studies could expand our understanding of social dynamics in adolescence. Now we examine how

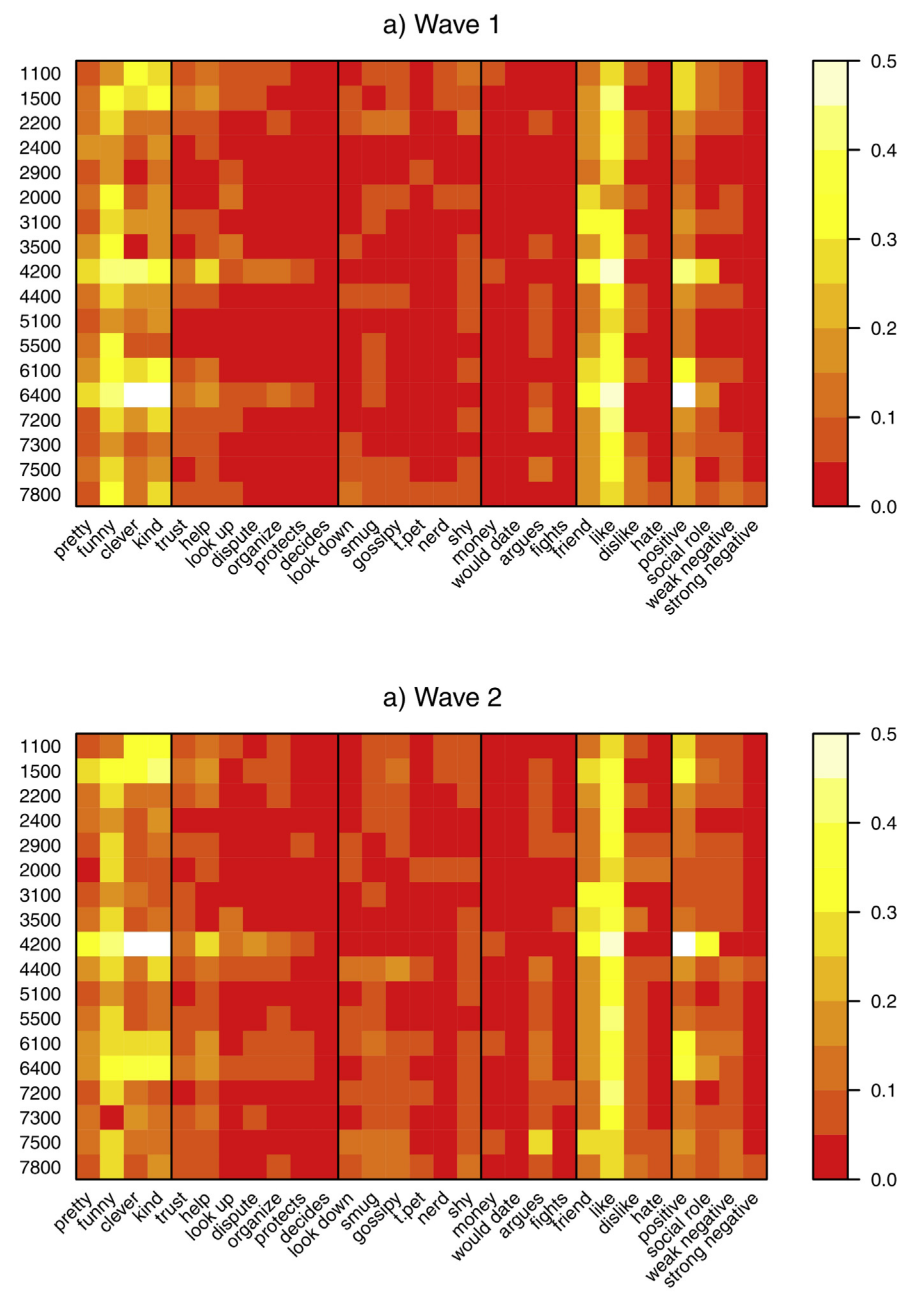

Fig. 11. The density of all used networks in the subsample classrooms at wave 1 (a) and 2 (b).

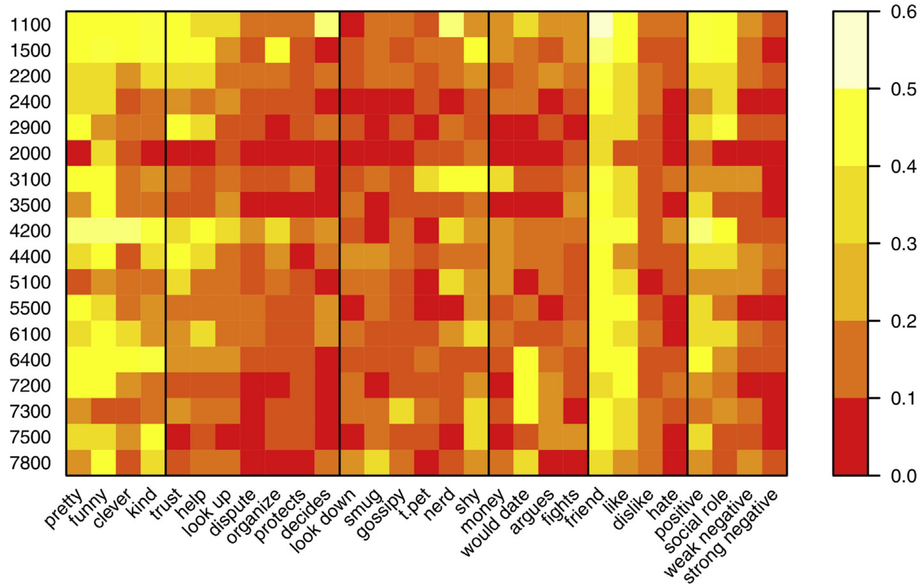

Fig. 12. The stability of ties in all used networks between wave 1 and 2. *(Stability is measured by the Jaccard index between the wave 1 and wave 2 instances of each network item; larger values represent greater stability; e.g. 0.5 means that the number of stable ties equals the number of changing ties in a network.)*

our composite networks are related to four basic networks of affection that are more regularly studied in the context of adolescent research: *friendship*, *liking*, *dislike*, and *hate*.[17]

Fig. 9 provides a more detailed picture of the overlaps between the eight networks at the two time points. Unlike the ones above, these heatmaps show the asymmetric relations between the items: each cell represents the amount of overlapping ties between the row and column networks, relative to the total number of ties in the *row* network (compare with Section 2.3). The figures are based on the mean values of this overlap index across the 18 classes.

First of all, we have to note that plots from wave 1 and 2 looking strikingly similar, suggesting that the structure of the relations between these networks remained remarkably stable over the six-month period, despite the considerable changes in network ties (see Fig. 12 in Appendix B). Second, we can see in the figure that positive and social role attributions are asymmetrically related, as shown by the 2-by-2 cell block in the top left corner. Most social role ties are coupled with positive attributions, but many positive attribution ties exist without social role attributions. This is true for all classrooms in both waves.

Further, the two positive composite networks have interesting connections to friendship and liking. Looking at the third 2-by-2 cell block from the left in the first row, we can see social role attributions more often run parallel with friendship than do positive perceptions. On the contrary, more of the positive attribution ties coincide with liking than do social role attributions. These findings hold for 16 out of the 18 classes in both waves. Comparing to the number of friendship and liking ties (first block in the third row) tells us that a larger proportion of friendship ties are also social role or positive attributions than liking ties. This is again true in 16 classrooms at both observations.

Results for the negative networks vary more strongly between classrooms. Still, it seems that more strong negative attributions are also hate ties than weak negative attributions (see the last 2-by-2 block in row 2). We find this tendency in 11 and 12 classes in wave 1 and 2, respectively. This means that by choosing a higher threshold for negative attributions, we retain a disproportionate amount of ties which are associated with hate. The same difference between strong and weak negative attributions is not so prevalent in the case of dislike: only in case of 6 (wave 1) and 3 (wave 2) classes can we observe a similar pattern.

Somewhat on the contrary, it is typical that a larger proportion of dislike and hate ties are paired with weaker rather than stronger negative attributions (see the second block in the bottom row). In general, hate more often comes with the two other forms of negativity than does dislike. These findings are true in 12 and 14 classrooms in wave 1 and 2, respectively.

We provide a concise overview of the results discussed so far in Fig. 10. The circles represent ties in the eight networks; their sizes are proportional to their average density in the 18 classrooms. Moreover, the size of each intersection area (pairwise and three-way) is approximately proportional to the average amount of ties in the given category.[18] We only present one figure as the results from the two waves are almost identical (see Fig. 9).

For the sake of the illustration, we disregarded the overlap between the "two worlds" of negative and positive networks. This helped to make the visual representation of the remaining overlaps more realistic. We have seen in Fig. 9 that positive–negative overlaps are very scarce: only a little ambivalence is captured by the eight measures (although the affective networks are mutually exclusive, there were no constraints on the other 17 items used to construct the new network measures).

It is clear that positive and social role attributions are coupled with friendship or liking in the vast majority of cases. This may be somewhat surprising for positive attributions, but remember that we used a threshold of 2 when constructing this measure, which means that a tie requires positive perceptions in at least two specific dimensions. Indeed, with a lower cut point at $t = 1$, we would see more positive attributions without positive affection (we do not present these results here).

[17] See Appendix A for the measurement of these relations in this specific study.

[18] Approximately, because Venn diagrams of more than two sets cannot exactly represent the relative areas of cell parts in most cases. However, we used the algorithm programmed in the *venneuler* R-package for placing the sets to ensure the image is as close to the observed values as possible. A stress-type measure (see Section 3.3) shows that this two-dimensional figure represents the relative size of the set parts very well ($p < 0.05$).

Social role attributions are typically strongly embedded in friendships in the studied sample. Although they sometimes also occur in liking relations, it is very rare that social roles are attributed without the presence of any positive affection. On the contrary, weak and strong negative attributions quite often exist without dislike or hate – though there is a significant level of overlap.

4. Conclusions from the example

In this paper, we presented an analytical strategy for reducing the number of network dimensions in highly multiplex network data. Applying statistical methods that are already well established for monadic variables, we demonstrated that different types of social relations may be classified based on their pairwise similarities. The resulting clustering may reveal important latent dimensions of relationships in a given context. We also showed a simple yet flexible approach to construct composite network measures. This is primarily meaningful when the items are clustered based on their similarity on the level of network ties. Statistical models for multiplex networks can be more easily applied to the aggregated networks compared to the original, large set of dimensions.

The application of the presented procedure to our observation of 21 interpersonal networks from 18 high-school classrooms at two time points resulted in the identification of three well-interpretable groups of perceptions. These we labeled as positive, negative, and social role attributions. Whilst there was substantial change in peer nominations, the global similarity structure among the relational dimensions remained remarkably stable between the two waves.

The stability of the overall clustering points to two interesting phenomena. First, it provides evidence that attributions in the same cluster are connected also dynamically: the ties change but the cluster structure of attributions remains stable. Second, although the group-level cluster solution changes between the waves for many classes, these changes seem to cancel out when we average over all classes. This proves that it is indeed a promising direction to look for the latent structures connecting several network items, and that when doing this it is helpful to consider a set of groups rather than one particular group.

Based on some simple tests, we concluded that the proposed global clustering of items fits the similarity structure of the individual classrooms reasonably well, independent of the academic level as indicated by the classroom training program, which is known to be strongly associated with the socio-economic background, academic achievement, and aspirations of students. Apparently, the clustering found for the network items taps underlying dimensions that are stronger than variations between classrooms in subcultures and value systems.

Relying on the identified three-cluster solution for the items, we created four aggregated network measures. A combination threshold of 2 was applied to compose a positive and a social role attribution network. We found that these two composite networks are asymmetrically related in the studied classrooms: positive attributions seem to represent general perceptions which often exist without role attributions, while the assignment social roles almost always coincides positive perceptions. Further, social role attributions also appear to be closely linked to friendship, which highlights the functional features of positive relations at school. Finally, our results suggest that positive attributions (or similar measures) may not only be instrumental in explaining how positive affections develop but also in studying how affective relations become stronger, e.g. how liking evolves into friendship.

In the case of negative attributions, we defined two composite networks at the thresholds of 2 and 3. Thus we acquired networks representing "weaker" and a "stronger" negative perceptions. We found in many of the classrooms that strong negative attributions more often coincide with strong negative affections than do weaker attributions. However, it is notable that many negative attributions, both weak and strong, exist between students who did not report any negative feelings (dislike or hate) about each other. These results exemplify that defining a set of ordered composite network measures, or in more general terms, defining weighted networks as a result of the procedure is also a viable option.

5. Discussion

The proposed procedure for the cluster analysis of multiplex networks opens novel possibilities and also presents new challenges for social network research. While statistical methods of dimension reduction have proven useful for the construction of reliable monadic measures in the social sciences, they have hardly been applied to dyadic data with the aim to define aggregated network measures. The main intended contribution this paper is the combination of a set of already available methods and measures to create composite networks. We believe that this is an innovative application which requires some new practices and rules of thumb further to be developed. Therefore, here we give a detailed account of some critical decisions and questions brought up by the procedure.

First, we have to note that the choice between similarity measures obviously affects the results and interpretations greatly. The use of the Jaccard index, a dyadic measure, fits well to the aim of combining information from several networks on the level of social ties. As explained in Section 1.3, actor-level similarity, expressed for instance by indegree correlation, can also be a meaningful basis for classifying the networks. However, comparisons based on actor-level measures do not take into account the dyadic essence of network structure. Therefore, actor-level similarity measures may be useful primarily when the ultimate goal is the classification of actors based on their position in a group, and not the construction of aggregated networks.

On the side, we may note that perhaps a more common application of Cluster Analysis to networks would be the classification of networks based a set of aggregate structural measures, such as density, hierarchy, centralization, and so on. However, such an approach seems more valuable for descriptive studies about the structural similarities between different types of networks than for the aim of finding common dimensions for networks measured in the same groups in highly multiplex datasets.

As a second and related point, we have to mention that for assessing network similarities, different rules of thumb should be used than for inter-item correlations of monadic variables in scale construction or Factor Analysis. For instance, a correlation of.5 between scale items may be labeled as rather low, but a Jaccard index of the same magnitude means that there are exactly as many overlapping ties in two networks as ties that are present in only one of the two – which is definitely a high level of similarity. In light of this, it is important to keep in mind that what counts as low or high similarity depends on the measure we use.

Third, plenty of available options were deliberately left unexplored in case of goodness of fit assessment. On the one hand, further tests can include the application of additional measures, such as the cophenetic correlation coefficient (Rohlf and Sokal, 1981). On the other, the main challenge here is empirical and theoretical: the general problem of finding a single clustering that fits the data from multiple groups (in our case, classrooms). There is no simple general solution to this issue, and it deserves separate and more thorough treatment.

A few words need to be said, however, about one related issue. Obviously, the chances for finding a well-fitting global clustering depend on the homogeneity of the groups with regard to the inter-item similarities. The second step of the procedure (see Section 2.4) aims at increasing this homogeneity. Yet, it still is possible that the studied groups are systematically different from each other. With many groups it may thus be a viable approach to classify the groups themselves based on the results from the cluster analysis of the network dimensions, and so identify the general types of communities in our sample.

By following the same approach, we may not only learn about the studied groups, but also about the network items which cause most of the differences (like *argues*, *money*, *would date*, and *fights* in the presented analysis). Why are these networks heterogeneous in their relations to the other items? Do they have a certain consistent meaning in some groups and a different one in others? Are the differences related to sociologically relevant factors? Is it possible that the heterogeneities are due to omitted network items? The last question leads us to the next point.

Fourth, in any multiplex network study, the results will depend on the set of network items being considered, and we cannot be sure that we measured all of the important items connected to the relational dimensions of interest. We may be unaware of the number of latent relational dimensions in a given social context. In which ways are the results of the procedure affected by the item pool that we use? At this point, it is difficult to give conclusive answers to this question. We can only say that since the classification of network items in our approach is based on relative similarity (i.e., two items are more similar to each other than to the other items), the omission of a set of networks measuring something different would probably not affect the results strongly.

It is likely, however, that the relative number of items belonging to each category has an impact on both the results in each group and the similarity of clusters across multiple groups. In the example presented, there were fewer negative than positive items, and networks in the former group showed lower consistency across groups. This may also be related to density: if the networks measuring a relational dimension are sparse, we may need more of them to acquire robust results. However, it is important to note that low density in itself is not a cause of low consistency across groups. The reason for an item to be dropped at this step of the procedure is that it does not belong to any of the clusters. This suggests that in order to arrive at good measures for rare social relations (i.e., with low average degrees), researchers should look for and employ a larger number of specific items than in cases of more common ties.

The measurement of bullying is a good example here. Specific questions may facilitate a better identification of bully-victim relations than more general ones, but they may yield very sparse networks. However, with a sufficiently large number of items chances are better for the identification of a separate cluster of bullying relations across groups. This cluster then might be used to construct more valid, and perhaps multiple, measures of bullying relations. The conclusion of this exercise might also be that specific forms of bullying cannot be separated from other social relations that have their own specificity, such as teasing, rivalry, or dominance. In any case, however, the results would contribute to our understanding of bullying among children.

Further, quite naturally, with few items one cannot distinguish between the possible sub-dimensions in a given set of networks. That is, with fewer positive items, we probably would not have been able to distinguish between positive attributions and social role attributions; on the other hand, measuring more negative items might have revealed interesting substructures in the negative attribution cluster. Leaving out key network items in a given dimension is likely to make the identification of all dimensions more difficult, as boundaries between clusters may become more fuzzy. However, we need to collect further experience with the structure of network similarities before we can give more specific answers to the questions raised here.

In spite of the uncertainties due to the lack of experience with the procedure, the first results seem to be substantively meaningful and relevant. Based on our findings, studies into the co-evolution of positive attributions and affections of different strength promise to clarify how friendships develop in school, and may unveil a so far omitted variation in adolescents' friendship relations. The closer examination of the part that social roles play in friendship may be important in understanding why (some) friends are more influential when it comes to academic achievement, aspirations, risky behaviors, and so on. Our results also showed there is more negativity in student relations than what is captured by the simple questions of dislike and hate. This suggests that by using an appropriate set of measures for negative perceptions we may be able to acquire a more complete picture of negative relations in a social setting. Such data could contribute greatly to studies of social exclusion, status competition, intergroup conflict, and bullying – in or outside of the school setting.

In summary, this paper proposed an approach for dimension reduction in highly multiplex network data in multiple groups, and demonstrated on a real-life example that this approach is viable and leads to well-interpretable results. Some parts of the procedure need to be further developed. Also, more examples and experience are necessary to set up rules of thumb that can guide researchers when applying this procedure, and explore the effects of imperfect measurement on the results. Despite these issues, however, the proposed analytical strategy holds some potential to help uncovering the latent dimensions of relationships in different social contexts. Further, elaborating and applying this approach in research promises to relax the dependence on the measurement of complex types of relations by a single, often loosely defined item, as it is usually done for friendship. Studying multiple relational items with a well-specified meaning may improve the validity of our interpretations and may help us develop better measures of social relations.

Appendix A. The RECENS high-school dataset

The data analyzed in the paper come from a 4-wave longitudinal network survey conducted in 43 Hungarian high-school classrooms in the period of 2010–13. The data was collected as part of the "Wired into each other" network study by the Research Center for Educational and Network Studies (RECENS). The classrooms in the dataset were distributed in 7 schools from the capital, one large town, and two smaller towns in Hungary. In each participating school, all classes in the first-grade cohort (14–15 year-olds) of 2010–11 were followed for the first three of their four years in high school. Students answered self-administered pen-and-paper questionnaires four times during this period: twice in the first year (two months after the start of first grade and six months later), then once per year in the next two (1.5 and 2.5 years after the first wave).

The questionnaires in each classroom and wave were filled out in the course of a 45-minute class at school, under the supervision of at least two trained researchers participating in the study. Before each wave, students and their parents received information about the survey and were asked for participation and consent. Network ties were reported by students using a classroom-level roster, that is, the number of ties to classmates was not limited, but ties to people outside one's own class were not permitted. Affective relationships were measured on a 5-point scale: friendship, liking, neutrality, dislike, and hate, where choices were mutually exclusive. In addition to these items, questions about bullying and victimization networks, shared activities and perceptions of peers' attributes, behavior, social roles, and status were asked in the questionnaires. Altogether, the dataset contains information on various social ties between students along 40 dimensions in each classroom. In the present study, we used 21 of these network dimensions representing perceptions about a selection of attributes, behaviors, and roles (see Section 3.2).

This paper analyzes the first two waves of the RECENS dataset. Wave 1 was collected in October 2010, only about a month after the start of the first high-school year, and wave 2 in April 2011. These observations frame an intense six-month period when classroom members become familiar with their new environment, get to know each other, and start to develop shared views about which behaviors or opinions are acceptable in the class and which are to be rejected. The evolution of the value systems (as used in Coleman, 1961) of classrooms goes hand in hand with changes in adolescents' perceptions about their classmates, and may also affect the relationships between the different dimensions of peer perceptions in the community. As a consequence, the comparison of our results from the two waves may highlight certain general tendencies of the ongoing social dynamics.

Appendix B. The analyzed 18-classroom subsample

For our exploratory analysis, we selected a random subsample of 18 classrooms so that each school, classroom training program (vocational, technical, or grammar),[9] and town size (capital, large town and small town) is independently represented by at least 2 classes (see Tables 3 and 4 for an overview of the subsample). These contextual variables capture regional and school-level differences and the systematic variation in students' background between training programs (Dobos et al., 2011), which may all be related to the structuring of perceptual dimensions in the classrooms. Thus, our subsample retains the variation of the full sample in some important contextual aspects, while it remains small enough to be conveniently explored. The selection of at least two cases in each category also provides protection against idiosyncrasies in classroom-level results. Finally, using a subset of classrooms from the full dataset leaves open the possibility of future cross-validation.

Table 3
The distribution of classrooms in schools, locations, and training programs in the subsample and the full sample.

School	Number of classrooms (in subsample/in full sample)			
	Grammar	Technical	Vocational	Total
Bigtown 1	2/5	–	–	2/5
Bigtown 2	–	2/4	2/6	4/10
SmalltownA 1	1/3	0/1	1/3	2/7
SmalltownB 1	–	2/4	–	2/4
SmalltownB 2	–	1/2	1/3	2/5
Capital 1	2/4	–	–	2/4
Capital 2	–	2/4	2/4	4/8
Total	5/12 (42%)	7/15 (47%)	6/16 (38%)	18/43 (42%)

Table 4
Identification of classrooms in the analyzed subsample.

School	Classroom IDs per training program					
	Grammar		Technical		Vocational	
Bigtown 1	1100	1500	–	–	–	–
Bigtown 2	–	–	2200	2400	2900	2000
SmalltownA 1	3100	–	–	–	3500	–
SmalltownB 1	–	–	4200	4400	–	–
SmalltownB 2	–	–	5500	–	5100	–
Capital 1	6100	6400	–	–	–	–
Capital 2	–	–	7200	7300	7500	7800

Basic descriptive information about the studied classrooms is presented in Table 5. The subsample we analyze here consists of 5 grammar, 7 technical, and 6 vocational school classes. The size of the classrooms varies between 26 and 38 students, with an average of 32 over the two waves. Girls are in a majority in most of the classes (14 out of 18), and their average proportion is 63%. It has to be noted that across all three training programs, girls are over represented in our sample compared to their national share in the time period of the study: 69%, 62%, 60% against 57%, 49%, 38% in grammar, technical, and vocational training respectively (KSH, 2011, 2012, 2013). However, our classrooms still show considerable heterogeneity in gender composition: the lowest observed proportion of girls is 34%, while the highest is 80%. The differences in the ratio of boys and girls are not associated with training program or classroom size in our subsample.

The average age of students was 15.3 years at the time of wave 1 (with the mean of within-class standard deviations being 0.57 years). There are significant differences between training programs with regard to the classroom-level means and variances of age: vocational school classes have the highest average age and also the largest within-class variance, then come technical classes, followed by the grammar training program (p-values of all pairwise tests are below 0.05). This is in line with the general tendency in Hungarian education: between 2005–2007, the proportion of students who failed and had to repeat a grade was more than three times as high in vocational (5%) and twice as high in technical training (3%) than in grammar classes (1.5%) (Fehérvári, 2009:7). Later studies report that the gap further widened since then, leading to a 7%:1% ratio of fails between vocational and grammar classrooms in 2008/2009, with one out of every six vocational students failing first grade in high school (Dobos et al., 2011:171).

It can further be noted that there is some change in classroom compositions, due to joiners and leavers, between the two waves. The turnover rate is higher for technical and vocational (mean = 4.2 students) than for grammar classes (mean = 0.8 students, $p = 0.01$ for the test of zero difference). However, it is likely that most of this fluctuation is due the swapping of problematic, school-avoiding students between schools – as suggested by a high positive correlation (0.77, $p < 0.01$) between the number of changes in the classroom roster and the number leaving and joining students who were absent at wave 1 and 2 respectively. Classroom sizes and the ratio of boys and girls are not altered substantially by the changing composition.

The differences between training programs are reflected also in participation rates. While on average 90% and 85% of students answered the questionnaires in wave 1 and 2 respectively, there is a large disparity between vocational classes and the two other types (see Table 6). In grammar and technical classrooms, typically less than 10% of students were absent from school at the time of the data collections or chose to opt out from the study, but the same rate is over 20% in vocational classes on average (the difference is significant at $p < 0.01$).

Based on field interviews with the head teachers of the vocational classrooms, several students who were absent were notorious for avoiding school, and some even faced to be expelled from the given institution because of this. Indeed, if we exclude absent students who left their classroom (and probably their school) by the third data collection, absence rates drop quite a bit – as it can be seen in Table 6.

The problem of absent later drop-outs is clearly one about drawing the boundary of the networks. Do these students belong to the classroom community even though they are mostly not at school? We decided to keep them for the present analysis, for two reasons. On the one hand, we argue that students who are absent more often than not are not active members of their class community. As a consequence, it is not important for us to know how they evaluate their

Table 5
Classroom-level descriptive statistics for the subsample.

Class	Training program	Size	Girls %	Age mean (std. dev.)	Size change w1 to w2	Leavers	Joiners	Girls %Δ
1100	Grammar	32	66%	15.1 (0.28)	−1	1	0	−0.01
1500	Grammar	29	76%	15.1 (0.37)	−1	2	1	−0.01
3100	Grammar	29	48%	14.9 (0.37)	0	0	0	0.00
6100	Grammar	33	82%	15.1 (0.37)	0	0	0	0.00
6400	Grammar	36	72%	15.0 (0.34)	0	0	0	0.00
2200	Technical	31	71%	15.2 (0.48)	−5	5	0	0.02
2400	Technical	29	62%	15.1 (0.48)	2	2	4	−0.01
4200	Technical	35	54%	15.0 (0.33)	1	0	1	0.01
4400	Technical	38	79%	15.1 (0.47)	−1	1	0	−0.01
5500	Technical	35	40%	15.4 (0.51)	0	2	2	−0.06
7200	Technical	31	61%	15.5 (0.72)	−3	3	0	−0.01
7300	Technical	31	65%	15.6 (0.76)	−2	5	3	0.01
2900	Vocational	32	59%	15.2 (0.48)	3	4	7	0.01
2000	Vocational	30	47%	16.0 (0.92)	−3	5	2	0.05
3500	Vocational	37	41%	15.7 (0.67)	1	1	2	−0.01
5100	Vocational	33	79%	15.5 (0.81)	1	1	2	−0.02
7500	Vocational	33	67%	15.6 (1.03)	−7	9	2	−0.05
7800	Vocational	35	69%	15.9 (0.96)	−6	6	0	−0.03

Table 6
Absence rates among all students and active students in the subsample classrooms.

Class	Training program	All absent (%) w1	Active absent (%) w1	All absent (%) w2	Active absent (%) w2
1100	Grammar	6%	6%	3%	3%
1500	Grammar	3%	3%	0%	0%
3100	Grammar	3%	3%	3%	3%
6100	Grammar	6%	6%	3%	3%
6400	Grammar	3%	3%	0%	0%
2200	Technical	6%	3%	8%	4%
2400	Technical	3%	0%	6%	3%
4200	Technical	0%	0%	0%	0%
4400	Technical	0%	0%	0%	0%
5500	Technical	9%	6%	23%	13%
7200	Technical	32%	22%	21%	12%
7300	Technical	3%	0%	14%	4%
2900	Vocational	9%	3%	29%	7%
2000	Vocational	20%	8%	30%	10%
3500	Vocational	27%	10%	45%	9%
5100	Vocational	18%	13%	35%	21%
7500	Vocational	36%	28%	35%	19%
7800	Vocational	23%	16%	10%	4%

classmates (all of whom they might not even know). Additionally, their missing responses do not pose a problem for the analysis.

On the other hand, these people, as rarely as they may contact them in or outside of school, can still exert significant influence on their classmates' interpersonal evaluations, especially as extreme role models. Therefore, taking into account the nominations they receive from classmates may help to better identify the structure of perceptual dimensions among the active members of the classrooms. To summarize, we argue that the higher rate of absence in vocational classes should not lead to biased results regarding the active part of the classes, and so systematic differences between classrooms should not arise from this phenomenon.

References

Asher, S.R., Coie, J.D. (Eds.), 1990. Peer Rejection in Childhood. Cambridge University Press.
Bartholomew, D., Steele, F., Galbraith, J., Moustaki, I., 2008. Analysis of Multivariate Social Science Data. Routledge.
Batagelj, V., Bren, M., 1995. Comparing resemblance measures. J. Classif. 12 (1), 73–90.
Berndt, T.J., 1982. The features and effects of friendship in early adolescence. Child Dev. 53, 1447–1460.
Borgatti, S.P., 1994. Using MDS to infer relative status from dominance matrices. Connections 17 (2), 75–77.
Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. UCINET for Windows: Software for Social Network Analysis.
Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G., 2009. Network analysis in the social sciences. Science 323 (5916), 892–895.
Bower, E.M., 1960. Early Identification of Emotionally Handicapped Children. Thomas, Springfield, IL.
Brandes, U., Lerner, J., Nagel, U., 2011. Network ensemble clustering using latent roles. Adv. Data Anal. Classif. 5 (2), 81–94.
Carroll, C., 2006. Canonical correlation analysis: assessing links between multiplex networks. Soc. Netw. 28 (4), 310–330.
Cillessen, A.H., Rose, A.J., 2005. Understanding popularity in the peer system. Curr. Dir. Psychol. Sci. 14 (2), 102–105.
Cillessen, A.H., Schwartz, D., Mayeux, L.,2011. Popularity in the Peer System. Guilford Press.
Coie, J.D., Dodge, K.A., Coppotelli, H., 1982. Dimensions and types of social status: a cross-age perspective. Dev. Psychol. 18 (4), 557.
Coie, J.D., Dodge, K.A., Kupersmidt, J.B., 1990. Peer group behavior and social status. In: Asher, S.R., Coie, J.D. (Eds.), Peer Rejection in Childhood. Cambridge University Press.
Coleman, J.S., 1961. The Adolescent Society. Free Press, New York.
Cox, T.F., Cox, M.A., 2010. Multidimensional Scaling. CRC Press.
Dabkowski, M., Breiger, R., Szidarovszky, F., 2015. Simultaneous-direct blockmodeling for multiple relations in Pajek. Soc. Netw. 40, 1–16.
Davis, J.A., 1968. Statistical analysis of pair relationships: symmetry, subjective consistency and reciprocity. Sociometry 31 (1), 102–119.
De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S., Arenas, A., 2013. Mathematical formulation of multilayer networks. Phys. Rev. X 3 (4), 041022.
Dekker, D., Krackhardt, D., Snijders, T.A., 2007. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. Psychometrika 72 (4), 563–581.
Dobos, K., Gloviczki, Z., Halász, G., Kaposi, J., Kovács, I.V., Lannert, J., Loboda, Z., Setényi, J. (Eds.), 2011. Jelentés a magyar közoktatásról – 2010. Oktatáskutató és Fejlesztö Intézet, Budapest.
Doreian, P., Batagelj, V., Ferligoj, A., 2005. Generalized blockmodeling. In: Structural Analysis in the Social Sciences, vol. 25. Cambridge University Press.
Faust, K., 2005. Using correspondence analysis for joint displays of affiliation networks. Models Methods Soc. Netw. Anal. 7, 117–147.

Faust, K., Skvoretz, J., 2002. Comparing networks across space and time, size and species. Sociol. Methodol. 32 (1), 267–299.
Fehérvári, A., 2009. Kudarcok a szakiskolában. Szakképzési Szemle 25 (1), 23–44.
Fienberg, S.E., Meyer, M.M., Wasserman, S.S., 1985. Statistical analysis of multiple sociometric relations. J. Am. Stat. Assoc. 80 (389), 51–67.
Fischer, C.S., 1982. What do we mean by 'friend'? An inductive study. Soc. Netw. 3 (4), 287–306.
Hanneman, R.A., Riddle, M., 2005. Introduction to Social Network Methods. Riverside, University of California.
Homans, G.C., 1950. The Human Group. Harcourt, Brace, New York.
Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classif. 2 (1), 193–218.
Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. 44, 223–270.
Kadushin, C., 2012. Understanding Social Networks: Theories, Concepts, and Findings. Oxford University Press.
Kendall, M.G., Babington Smith, B., 1939. The problem of *m* rankings. Ann. Math. Stat. 10 (3), 275–287.
Krackhardt, D., 1988. Predicting with networks: nonparametric multiple regression analysis of dyadic data. Soc. Netw. 10 (4), 359–381.
KSH, 2011. Oktatási adatok – 2010/2011. Stat. Tükör 5 (21), 1–4.
KSH, 2012. Oktatási adatok – 2011/2012. Stat. Tükör 6 (23), 1–5.
KSH, 2013. Oktatási adatok – 2012/2013. Stat. Tükör 7 (32), 1–5.
Lazega, E., 2001. The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership. Oxford University Press, New York, NY.
Legendre, P., 2005. Species associations: the Kendall coefficient of concordance revisited. J. Agric. Biol. Environ. Stat. 10 (2), 226–245.
Lusher, D., Koskinen, J., Robins, G., 2012. Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications. Cambridge University Press.
Marsden, P.V., Campbell, K.E., 1984. Measuring tie strength. Soc. Forces 63 (2), 482–501.
Masten, A.S., Morison, P., Pellegrini, D.S., 1985. A revised class play method of peer assessment. Dev. Psychol. 21 (3), 523–533.
Newcomb, A.F., Bukowski, W.M., Pattee, L., 1993. Children's peer relations: a meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. Psychol. Bull. 113 (1), 99–128.
Pekarik, E.G., Prinz, R.J., Liebert, D.E., Weintraub, S., Neale, J.M., 1976. The pupil evaluation inventory. J. Abnormal Child Psychol. 4 (1), 83–97.
Prell, C., 2011. Social Network Analysis: History, Theory and Methodology. Sage.
Prinz, R.J., Swan, G., Liebert, D., Weintraub, S., Neale, J.M., 1978. ASSESS: adjustment scales for sociometric evaluation of secondary-school students. J. Abnormal Child Psychol. 6 (4), 493–501.
Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. 66 (336), 846–850.
Roethlisberger, F.J., Dickson, W.J., 1939. Management and the Worker. Cambridge University Press.
Rohlf, F.J., Sokal, R.R., 1981. Comparing numerical taxonomic studies. Syst. Biol. 30 (4), 459–490.
Sampson, S., 1969. Crisis in a cloister (Unpublished Ph.D. Dissertation). Cornell University.
Skvoretz, J., Faust, K., 2002. Relations, species, and network structure. J. Soc. Struct. 3 (3), 139–145.
Snijders, T.A., Lomi, A., Torló, V.J., 2013. A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. Soc. Netw. 35 (2), 265–276.
Snijders, T.A., Van de Bunt, G.G., Steglich, C.E., 2010. Introduction to stochastic actor-based models for network dynamics. Soc. Netw. 32 (1), 44–60.
Sokal, R., Michener, C., 1958. A Statistical Method for Evaluating Systematic Relationships. University of Kansas Science Bulletin, University of Kansas.
Szell, M., Lambiotte, R., Thurner, S., 2010. Multirelational organization of large-scale social networks in an online world. Proc. Natl. Acad. Sci. 107 (31), 13636–13641.
Wasserman, S., Faust, K., 1989. Canonical analysis of the composition and structure of social networks. Sociol. Methodol. 19 (1), 1–42.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.
Wasserman, S., Faust, K., Galaskiewicz, J., 1990. Correspondence and canonical analysis of relational data. J. Math. Sociol. 15 (1), 11–64.
White, H.C., Boorman, S.A., Breiger, R.L., 1976. Social structure from multiple networks. I. Blockmodels of roles and positions. Am. J. Sociol. 81 (4), 730–780.
Wiggins, J.S., Winder, C., 1961. The peer nomination inventory: an empirically derived sociometric measure of adjustment in preadolescent boys. Psychol. Rep. 9 (3), 643–677.