

Individuality in Social Networks

Daniel Borchmann and Tom Hanika

1 Introduction

Social networks form an integral part of human societies, and their study has been at the core of social science for a long time. It is only recently that mathematical methods have entered the stage, mainly because social networks are now made more explicit than ever due to the availability of social media. This has allowed classical mathematical instruments from graph theory and elsewhere to be applied to social networks—with astonishing results.

One of the first breakthroughs in understanding social networks by means of properties of their graph representations is due to the seminal work by Watts and Strogatz [26]. Here the authors introduce the notion of *small world networks*, encompassing the two simple graph properties of *average shortest path length* and *average local clustering coefficient*. Based on these properties, a graph is said to be a *small world network* if the average shortest path length is small and if the average local clustering coefficient is large. The second seminal result in that direction is the work by Barabási and Albert [3], where social networks are characterized as graphs whose *degree distribution* follows a low-degree *power-law distribution*. It turns out that, surprisingly, both small world networks and power-law distributions describe social networks to a large degree.

D. Borchmann

Chair of Automata Theory, Technische Universität Dresden, Dresden, Germany

e-mail: daniel.borchmann@tu-dresden.de

T. Hanika (✉)

Knowledge & Data Engineering Group, Research Center for Information System Design (ITeG),
University of Kassel, Kassel, Germany

e-mail: tom.hanika@cs.uni-kassel.de

In the wake of the results around small world networks, a plenitude of graph-related properties have been reinterpreted as properties of social networks, a popular example being the interpretation of *cliques* in social networks as social groups. However, despite a comparably vast body of research, characterizing all relevant aspects of social networks in terms of mathematical properties of their graph representation has not been achieved to a satisfactory degree. In particular, graphs exist on which existing measures cannot differentiate further, but which intuitively represent qualitatively different social networks.

In this work we want to consider another facet of bipartite social networks, which, as far as we can see, has not been investigated in the literature. This facet is *individuality* in social networks, and by this we intuitively mean the number of unique groups of users a social network has. Note that despite the fact that individuality is concerned with individual users of a network, the measure of individuality we want to investigate in this work is a property of the whole network. It should thus not be confused with notions such as *centrality* or *betweenness*, which apply only to individual vertices instead.

To define the uniqueness of a group of users, we consider the uniqueness of its *milieu* in the given bipartite social network. This intuition of individuality strongly depends on the actual definition of “milieu,” a notion that has been discussed in the social sciences before. However, we shall define and employ in this work an interpretation of this word that is different from the one usually used [24].

In a classical representation of social networks as graphs, two users are linked by an edge if and only if they “know” each other in this network. Then the notion of a milieu of a particular user could just be represented as the *neighborhood* of this user in this graph. In this work, however, we want to take up a different stand by representing bipartite social networks as *formal contexts*. These are structures originating from the theory of *formal concept analysis* [9, 27] that allow general investigations of data sets comprising of objects with certain attributes. Using formal contexts, we shall represent a social network as a collection of users with certain properties, where the actual choice of the properties is a matter of modeling. In this way, we can represent various aspects of a social network in a uniform manner.

The main goal of this work is to illustrate that our new notions of individuality are both natural and meaningful. To this end, we shall examine these measures on various real-world data sets, providing evidence that our definitions are reasonable. Even more, we shall show that networks that are similar in terms of their small world character can vary widely when it comes to individuality, suggesting that our new notion expresses properties of social networks that are not covered by the standard notions.

The paper is structured as follows. After revisiting some existing research on mathematical investigations of social networks in Sect. 2, we shall have a closer look on how to represent social networks as formal contexts in Sect. 3. Thereafter, we shall present our notion of *individuality* in Sect. 4, together with the two auxiliary measures of *individuality distribution* and *average milieu size*. An experimental investigation of these new notions follows in Sect. 5. We close with Outlook in Sect. 6.

2 Related Work

Formal concept analysis originated as a subfield of mathematical order theory, more precisely of lattice theory [4]. Lattice theory itself has already been applied to social network analysis, in particular to understanding the clique distribution (among others) in social networks, for example, in [7]. In this work concept lattices were used to analyze the relations between cliques.

Cliques indeed will play a major role in our considerations, and, as already mentioned, cliques have been investigated in the realm of social network analysis before. For example, the clique distribution of social networks was investigated in [28], where the focus was on empirically studying the connection between the power-law distribution of network nodes and the density of cliques. The authors showed to what extent the clique size distribution can be used to estimate the clique density in a social network. In [10] the authors proposed a method to efficiently estimate the distribution of clique sizes from a probability sample of network nodes. However, both works considered uni-modal social networks only. Previous work that also considered clique distributions in bi-modal networks is [22], where it is shown that medium sized cliques are more common in real-world networks than triangles. However, here only cliques in the projected graph were considered, and not in the original bipartite graph.

To the best of our knowledge, individuality in social networks as we consider it in this article has not been studied before as a property of social networks. The only relevant prior work is from the second author [2], on which this article greatly expands.

3 Social Networks as Formal Contexts

Formal concept analysis deals at its core with the representation of complete lattices through *formal contexts*. These are structures $\mathbb{K} = (G, M, I)$ where G and M are sets and $I \subseteq G \times M$ is a binary relation. The standard interpretation of formal contexts is that the set G is a set of *objects*, the set M is a set of *attributes*, and $(g, m) \in I$ signifies that g has the attribute m .

Indeed, modeling bipartite social networks as formal contexts is straightforward: consider a social network and identify within this network two sets U and A . We think of the set U as the set of (*interesting*) *users* of the network and of the set A as the set of (*relevant*) *attributes* of the users in U . Note, however, that this interpretation of U as a set of users and A as a set of attributes is only one among many possible ones, and there is no restriction on the type of elements contained in these sets.

After having identified the sets U and A , a formal context *representing* a social network is of the form (U, A, I) where $(u, a) \in I$ for $u \in U$, $a \in A$ only if user u has attribute a . This representation is also closely linked to considering *bi-modal*

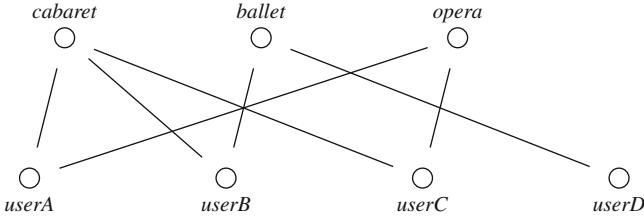


Fig. 1 Small motivational example, called the *music interest social network* (misl)

social networks, i.e., social networks that give rise to a bipartite graph. The benefit of choosing formal contexts over bipartite graphs is that in the former case we can apply methods from formal concept analysis to obtain further insights.

The particular choices of the user set U and the attribute set A are modeling decisions, and finding these sets may not at all be straightforward. For the set U one usually collects all real users of the framework, but other choices—depending on the particular application in mind—are possible. The set A of attributes can contain usual features such as *likes*, *posts*, and *gender*, but can also contain rather “unnatural” features such as other users. In this case, one could define, say, that some user u “has” some other user v as a feature if and only if they are linked in the original social network.

A small example of a social network is given by the bipartite graph in Fig. 1. A formal context representing this network is

\mathbb{K}_{misl}	cabaret	ballet	opera
userA	×		×
userB	×	×	
userC	×		×
userD		×	

In formal contexts we can define two natural *derivation operators* as follows. Let $A \subseteq G$ be a set of objects. Then the set A' of *common attributes* of A is defined as

$$A' := \{m \in M \mid \forall g \in A: (g, m) \in I\}.$$

Dually, for a set $B \subseteq M$ of attributes, we define the set B' of *satisfying objects* of B as

$$B' := \{g \in G \mid \forall m \in B: (g, m) \in I\}.$$

Note that although both operators are denoted by $'$, there is usually no danger of confusion, as it is clear from the context whether we are dealing with a set of objects or a set of attributes.

A pair (A, B) is called a *formal concept* of \mathbb{K} if and only if $A' = B$ and $B' = A$. The set A is then called the *extent* and B is called the *intent* of the formal concept (A, B) , respectively. Indeed, for each set $A \subseteq G$, the set A is an extent of \mathbb{K} if and only if $A'' = A$. The set of all formal concepts of \mathbb{K} is denoted by $\mathfrak{B}(\mathbb{K})$.

Let us point out the connection of formal concepts to cliques in bipartite graphs: for any formal context \mathbb{K} emerging from a bipartite graph, every formal concept of \mathbb{K} corresponds to a maximal bi-clique in the graph and vice versa.

On the set of all formal concepts $\mathfrak{B}(\mathbb{K})$ we can define a natural order as follows. Let $(A_1, B_1), (A_2, B_2) \in \mathfrak{B}(\mathbb{K})$. Then we say that (A_2, B_2) is *more general* than (A_1, B_1) , in symbols $(A_1, B_1) \leq (A_2, B_2)$, if and only if $A_1 \subseteq A_2$. While this definition looks rather asymmetric at first, it turns out that $(A_1, B_1) \leq (A_2, B_2)$ if and only if $B_2 \subseteq B_1$. Moreover, the relation \leq is an *order relation*, and $\mathfrak{B}(\mathbb{K})$ together with \leq forms a complete lattice, the *concept lattice* of \mathbb{K} . Conversely, one of the first results of formal concept analysis states that every complete lattice is isomorphic to the concept lattice of some formal context. In this way, formal concept analysis acts as a *representation theory* of complete lattices. Formal concept analysis also allows to link lattice theory to *relational data sets*, as the latter can naturally be represented as formal contexts. In this way, formal concept analysis makes accessible methods from lattice theory for the study of relational data tables.

4 Individuality of Social Networks

We have motivated our notion of individuality by the uniqueness of user milieus. Clearly, this motivation strongly depends on the particular interpretation of the word “milieu”, and it is the purpose of this section to provide a formal definition for it. Indeed, modeling a social network by a formal context suggests an immediate definition that is both simple and, as we find, convincing.

Let $\mathbb{K} = (U, A, I)$ be a formal context representing a social network. Then for each user $u \in U$ we define the *milieu* of u simply as the set $\{u\}'$ of attributes common to u . Moreover, if $V \subseteq U$ is a set of users, then the *milieu of V* is the set of attributes common to all users in V , i.e., V' . Using this definition of user milieus, we want to measure the individuality of a social network \mathbb{K} by the amount of milieus that occur in \mathbb{K} . Indeed, we shall be a bit more careful here, and propose a notion of *k-group individuality* as a measure to quantify the number of milieus that occur in \mathbb{K} as the milieu of groups of size k , in the sense of how many of the milieus occurring in our social network \mathbb{K} define groups of size exactly k , compared to the number of all groups of size k . Then, the more individuality a social network contains, the more individual groups of a certain size can be defined through their milieu. Conversely, if a social network is quite homogeneous, then defining certain subgroups of individuals by their milieu is improbable.

This approach can naturally be rephrased in terms of formal concept analysis: measuring individuality in \mathbb{K} for user groups of size k is the question of how many

subsets $V \subseteq U$ with $|V| = k$ can be expressed in terms of $V = B'$ for some $B \subseteq A$. In other words, we ask for the number of *extents of size k* in \mathbb{K} and use this number to measure the k -group individuality in \mathbb{K} . The following definition captures this idea.

Definition 1 Let $\mathbb{K} = (U, A, I)$ be a formal context. Define the set $\text{Ext}_k(\mathbb{K})$ as the set of extents of \mathbb{K} of size k , i.e.,

$$\text{Ext}_k(\mathbb{K}) := \{V \subseteq U \mid V = V'', |V| = k\}.$$

Then the k -group individuality $\text{gi}_k(\mathbb{K})$ of \mathbb{K} is

$$\text{gi}_k(\mathbb{K}) := \frac{|\text{Ext}_k(\mathbb{K})|}{\min\{\binom{|U|}{k}, 2^{|A|}\}}. \quad (1)$$

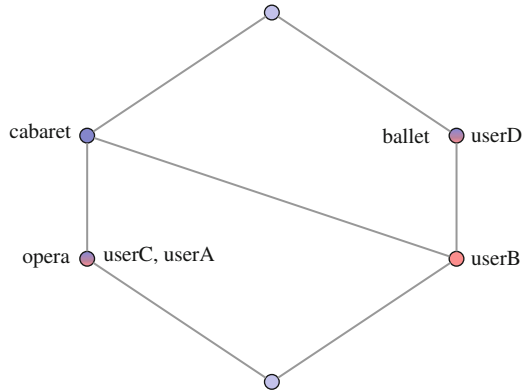
Note that we also normalize by the factor $\min\{\binom{|U|}{k}, 2^{|A|}\}$, because this is the maximal number of k -groups definable by their milieu, and thus allows comparability between individuality of different networks. The used normalization is not optimal, as for k larger than 1 the value of $\text{gi}_k(\mathbb{K})$ rapidly decreases. However, so far the authors are not aware of other normalization approaches.

On a side note, one may also consider the dual measure taking the intents of size k , which would help to measure and describe the individuality of a social network from the attribute point of view.

In terms of measuring the individuality in a social network, the value $\text{gi}_1(\mathbb{K})$ is of particular interest, as this is the percentage of users in this network uniquely determinable by their milieu. In this case, we shall also talk about the *user individuality* $\text{ui}(\mathbb{K}) = \text{gi}_1(\mathbb{K})$ of a social network \mathbb{K} .

Using our example from Fig. 1, we first compute the extent sets. As we see in Fig. 2, the concept lattice consists of four elements (apart from the top and bottom ones), and consequently there are four different extents. Indeed we obtain

Fig. 2 Formal concept lattice for \mathbb{K}_{misc}



$$\begin{aligned} \text{Ext}_1 &= \{\{\text{userB}\}\}, \\ \text{Ext}_2 &= \{\{\text{userB}, \text{userD}\}, \{\text{userA}, \text{userC}\}\}, \\ \text{Ext}_3 &= \{\{\text{userA}, \text{userB}, \text{userC}\}\}. \end{aligned}$$

Therefore, $\text{gi}_1(\mathbb{K}_{\text{misn}}) = \frac{1}{4}$, since only one user has a unique interest that is not covered by another user. We also obtain $\text{gi}_2(\mathbb{K}_{\text{misn}}) = \frac{1}{3}$, demonstrating that in this network the individuality of “pairs” of users is higher than for individual users. Finally, $\text{gi}_3(\mathbb{K}_{\text{misn}}) = \frac{1}{4}$, showing that there is only one group of size three.

The network would be changed considerably if userC would have liked ballet instead of cabaret. In this context, which we want to call *misn'*, there would be three extents of size one and therefore $\text{gi}_1(\mathbb{K}_{\text{misn}'}) = \frac{3}{4}$. Additionally, the number of extents of size two would be four, resulting in $\text{gi}_2(\mathbb{K}_{\text{misn}'}) = \frac{2}{3}$. In short, by not being a copy of the interest of userA, userC can shift the individuality of the network massively by one interest change.

A remark on computing k -group individuality is in order. From the very definition of $\text{gi}_k(\mathbb{K})$, it seems as if computing this value requires to iterate through all subsets of G of size k and check whether they are closed under \cdot'' . However, using methods from formal concept analysis, the overall effort can be reduced to compute only extents of size *at most* k . More precisely, the algorithm of Next-Closure [9] is able to enumerate closed sets of arbitrary closure operators in a particular order. Exploiting the fact that \cdot'' is a closure operator allows us to compute all extents of \mathbb{K} with only polynomial overhead. Furthermore, Next-Closure can be extended to compute only extents of size at most k , further reducing the overall computation costs. A drawback is that Next-Closure cannot be extended to only compute extents of size k , a disadvantage that is not of profound severity, since k -group individuality is usually computed for values $k = 1, 2, \dots, \ell$ up to some limit $\ell \in \mathbb{N}$.

Note that group individuality also allows detecting the presence of large homogeneous groups, i.e., groups of users with the same milieu. Clearly, such a group of size k exists if and only if $\text{gi}_k(\mathbb{K}) > 0$. In other words, the set

$$\text{gid}(\mathbb{K}) := \{k \in \mathbb{N} \mid \text{gi}_k(\mathbb{K}) > 0\}$$

can be seen as a quantity for the *individuality distribution* in the social network represented by \mathbb{K} .

Finally, another aspect of group individuality that we want to consider in this work is the question of how much *information* is necessary to define the milieu of a group of size k . In terms of our modeling of social networks as formal contexts, we reformulate the question to ask how many attributes are necessary *on average* to define a unique group of size k that is itself identifiable through its unique milieu. This gives rise to the following definition.

Definition 2 Let \mathbb{K} be a formal context and let $k \in \text{gid}(\mathbb{K})$. Define the *k-group average milieu size* $\text{ams}_k(\mathbb{K})$ of \mathbb{K} as

$$\text{ams}_k(\mathbb{K}) := \frac{1}{|\text{Ext}_k(\mathbb{K})|} \cdot \sum_{V \in \text{Ext}_k(\mathbb{K})} |V'|$$

For $k \notin \text{gid}(\mathbb{K})$ the value of $\text{ams}_k(\mathbb{K})$ is not defined. It may be set to 0 in those cases if this permits further calculations.

Average milieu size can be naturally linked to *robustness* of group individuality: to deprive a group of k users of being definable in terms of their milieu, on average $\text{ams}_k(\mathbb{K})$ attributes have to be removed from the social network. Consequently, if there are more than $\text{ams}_k(\mathbb{K})$ attributes removed from the network, substantial changes in the k -group individuality should be expected. Verifying this intuition is not within the scope of this work, and is left for future work.

5 Experimental Results

To illustrate our definitions of measuring individuality in social networks, we shall investigate seven different real-world social networks, introduced in Sect. 5.1. We shall see in Sect. 5.2 that all these social networks are indeed small world networks. In Sects. 5.3, 5.4, and 5.5, we examine group individuality, group individuality distribution, and average milieu size of these networks. Finally, we discuss our findings in Sect. 5.6.

5.1 Data and Modeling

In the following we provide short descriptions of the used data sets. The graph properties of all mentioned graphs are summarized in Table 1.

Table 1 Investigated (bi-)partite graphs and their properties

Graph	Vertices in U	Vertices in V	Edges	Edge-density
G_{CM}	40	25	95	0.095
G_{FB}	899	522	7089	0.015
G_{ALNM}	111	134	480	0.032
G_{PLNM}	607	209	5361	0.042
G_{APLNM}	79	188	903	0.061
G_{NB}	1495	367	1746	0.003
G_{SW}	18	14	89	0.35

5.1.1 Club Membership Network (CM) [14]

This data set consists of a bipartite graph describing the affiliations of a set of corporate executive officers to a set of social organizations. This graph consists of 65 vertices representing 40 persons (U_{CM}) and 25 organizations (V_{CM}), as well as 95 edges connecting them. In the following we shall denote this graph by $G_{CM} = (U_{CM} \cup V_{CM}, E_{CM})$.

5.1.2 Facebook-Like Forum Network (FB) [17]

This data set was created by using data from an online community of students from the University of California, Irvine. By using a forum and posting messages to various topics, the students and the topics constitute a bipartite social network. This network consists of a set of 899 users (U_{FB}) and a set of 522 topics (V_{FB}) as well as 7089 edges relating a topic to a user. We shall refer to the resulting graph as $G_{FB} = (U_{FB} \cup V_{FB}, E_{FB})$.

5.1.3 Lange Nacht der Musik (LNM) [20]

This data set stems from an annual cultural event organized in the city of Munich in 2013, the so-called *Lange Nacht der Musik* (Long Night of Music). The corresponding network consists of two bipartite graphs and their intersection. All three of them make use of the same set of vertices, consisting of 1159 users (U_{LNM}) and 212 distinct performances (V_{LNM}).

The first graph records for some users their attendance to performances. We refer to this *attendance graph* by $G_{ALNM} = (V_{ALNM} \cup U_{ALNM}, E_{ALNM})$, where $V_{ALNM} \subseteq V_{LNM}$ and $U_{ALNM} \subseteq U_{LNM}$.

The second graph represents the preferences of some users for where to go during the event. We call this graph the *preference graph* and refer to it in the following as $G_{PLNM} = (V_{PLNM} \cup U_{PLNM}, E_{PLNM})$, where $V_{PLNM} \subseteq V_{LNM}$, and $U_{PLNM} \subseteq U_{LNM}$.

Finally, by intersecting the vertex sets of G_{ALNM} and G_{PLNM} and restricting E_{ALNM} accordingly, we obtain a new graph G_{APLNM} that is the graph of performance attendances where the preferences of the users were known beforehand.

5.1.4 Norwegian Board Members (NB) [21]

This data set was compiled to investigate interlocking directorates among 384 public limited companies in Norway. This network consists of 367 companies (V_{NB}), the set of their 1495 directors (U_{NB}), 1746 edges connecting them (E_{NB}). We shall refer to this bipartite graph by $G_{NB} = (U_{NB} \cup V_{NB}, E_{NB})$.

5.1.5 Southern Women (SW) [24]

A systematic collection observing the social activities of 18 individual women (U_{SW}) over a 9-month period. In this time they attended 14 events (V_{SW}). We shall refer to this graph data set by $G_{\text{SW}} = (U_{\text{SW}}, V_{\text{SW}}, E_{\text{SW}})$.

5.2 Small World Network Properties

Graphs arising from social networks empirically satisfy the *small world network property* (SWP), i.e., they expose specific characteristics in terms of local clustering and global separation [5, 25, 26]. With exception of the LNM and NB networks, it is well known that all the networks mentioned in the previous section satisfy SWP to a certain extent. It is the purpose of this section to remind the reader of what those specific characteristics are and what particular values they exhibit on the corresponding networks.

In dealing with networks based on bipartite graphs, so-called *bi-modal networks*, it is common to employ *projections* to obtain the so-called *uni-modal* social networks that allow arbitrary links between vertices. While this approach may result in unforeseeable difficulties [29, 30], we shall nevertheless employ it in our work. The main reason for this is comparability: the methods from [26] only apply to uni-modal networks, and projections were used to turn bi-modal networks into uni-modal ones.

Given a bipartite network $G = (U \cup V, E)$, we obtain the *projection* $G^U = (U, E^U)$ of G by the following rule: whenever two users $u_1, u_2 \in U$ share a common neighbor in G , i.e., $\{u_1, v\}, \{u_2, v\} \in E$ for some $v \in V$, then an edge in the projected network G^U will connect them, i.e., $\{u_1, u_2\} \in E$. Then G^U is an undirected graph that corresponds to a uni-modal social network.

Since many observations of network properties are inherited from the network's *degree distribution* [13], it is common to validate the SWP of given networks against a so-called *null model*: to confidently claim that a graph indeed represents a small world social network, the values for local clustering and social separation in the null model should not be larger than in the original network. Here a *null model* for a uni-modal projection of the bipartite social network is represented by a graph that possesses an identical vertex degree distribution but otherwise consists of random connections between the vertices only. To obtain such a null model, we employ the algorithm from [11], which shuffles the edges of the original projection of the bipartite social network while preserving the degree of every vertex. In order to obtain a valid null model, i.e., independent from the edges of the input graph, we shuffle for at least 100 times the number of edges in the input graph [15].

In the following we shall explain in detail how global separation and local clustering are measured by means of *average shortest path length* and *average local clustering coefficients*.

5.2.1 Average Shortest Path

A *path* from u to w in a graph is a sequence of $n \in \mathbb{N}$ vertices successively connected by edges. The *length* of such a path is n . A *shortest path* between nodes u and v is a path of minimal length that starts at u and ends at v .

A social network possessing the small world property must exhibit an *average shortest path length* (ASP) that is low compared to the size of the network. For example, the follower graph of twitter has an average path length of about 4.17 [16], the internet router network has a value of 9.51 [23], and the southern women data set has a value of 1.09 [8].

The results we obtained in our experiment are listed in Table 2. All mentioned bipartite networks exhibit a low average shortest path length in their projected graphs. The numbers vary from 2.01 for the attendance network of LNM to 1.09 in the Southern Women data set. Moreover, in almost all cases the corresponding null model features about the same average shortest path length, as expected for small world social networks, with the only exception being the Norwegian Board Membership graph. For this network the value increases by about 15%. The exceptionality of NB among all data sets will prevail in the later measures.

5.2.2 Average Local Clustering Coefficient

Intuitively, a social network possesses a high local clustering, i.e., users that are connected to a particular user are also likely to be connected themselves. Local clustering in networks is measured by introducing a particular quantity called the *average local clustering coefficient* (ALCC) [26], and every social network must have a comparably high value for this parameter.

The average local clustering coefficient for a graph G can be calculated using the local clustering coefficients C_i for every v_i by $\text{alcc}(G) = \frac{1}{n} \cdot \sum_{i=1}^n C_i$, where

$$C_i = \frac{2 \cdot |\{\{v_j, v_k\} \in E \mid v_j, v_k \in N_i\}|}{|N_i| \cdot (|N_i| - 1)}$$

and $N_i := \{v \in V \mid \{v_i, v\} \in E\}$ is the *neighborhood* of v_i in G .

Table 2 Average shortest path lengths (ASP) and average local clustering coefficients (ALCC), alongside the values in a corresponding null model (NM)

Graph	# Edges	Density	ASP	ASP:NM	ALCC	ALCC:NM
G_{CM}	259	0.86	1.14	1.14	0.93	0.92
G_{FB}	123,231	0.30	1.70	1.70	0.69	0.62
G_{ALNM}	1145	0.19	2.01	1.93	0.52	0.31
G_{PLNM}	78,415	0.42	1.63	1.63	0.74	0.70
G_{APLNM}	586	0.20	1.58	1.58	0.71	0.64
G_{NB}	421	0.01	1.34	1.55	0.20	0.01
G_{SW}	138	0.90	1.09	1.09	0.94	0.93

To get a feeling of what certain values of ALCC actually mean for social networks, let us look at some examples: the aforementioned internet router network has an ALCC of 0.03, see [26]. Hence, it would not be considered as a small world social network. In comparison, the twitter followers network has an ALCC of 0.3 [16], which is bigger, but yet not high. Thus, twitter is a social network in which the small world property is not pronounced that much. A good example for a social network with a strong small world property is the one formed by actors using their common movies, which has an ALCC of 0.79, see [26].

Table 2 shows the values of ALCC of the projections of our data sets and of a corresponding null model. Here we observe values between 0.20 for NB and 0.94 for SW, and the values in the null model are lower than in the original networks.

5.2.3 Summary

The investigated data sets clearly exhibit small world network character, with exception of the Norwegian Board Member network, because of its low average local clustering coefficient. Nonetheless, this is a social network, since it is derived from real social data, showing that the heuristic of small world networks has its limits when it comes to identifying social networks. Because of this, it will be even more interesting to see the results for our new individuality measures on this network.

A drawback of our approach to identify small world networks is the usage of projections to obtain uni-modal networks from bi-modal ones. Indeed, in the literature bi-modal social networks are rarely analyzed without transforming them into uni-modal networks, since there are only few methods that can be directly applied to the former. With our new individuality measures we therefore hope to provide a reliable new measure that can be directly applied to bi-modal networks.

5.3 Group Individuality

We present in Table 3 and Figs. 3 and 4 the values of k -group individuality for our data sets for $k = 1, 2, 3, 4$. The largest value of 1-group individuality can be found for the NB data set with 0.96. This was not expected due to the very low value of

Table 3 Experimental results for g_i^k for $k = 1, 2, 3, 4$

Graph	g_1^1	g_2^1	g_3^1	g_4^1
G_{CM}	0.64	0.04	0.01	0.00
G_{FB}	0.70	0.02	0.00	0.00
G_{ALNM}	0.91	0.02	0.00	0.00
G_{PLNM}	0.69	0.02	0.00	0.00
G_{APLNM}	0.81	0.10	0.00	0.00
G_{NB}	0.96	0.00	0.00	0.00
G_{SW}	0.39	0.08	0.02	0.00

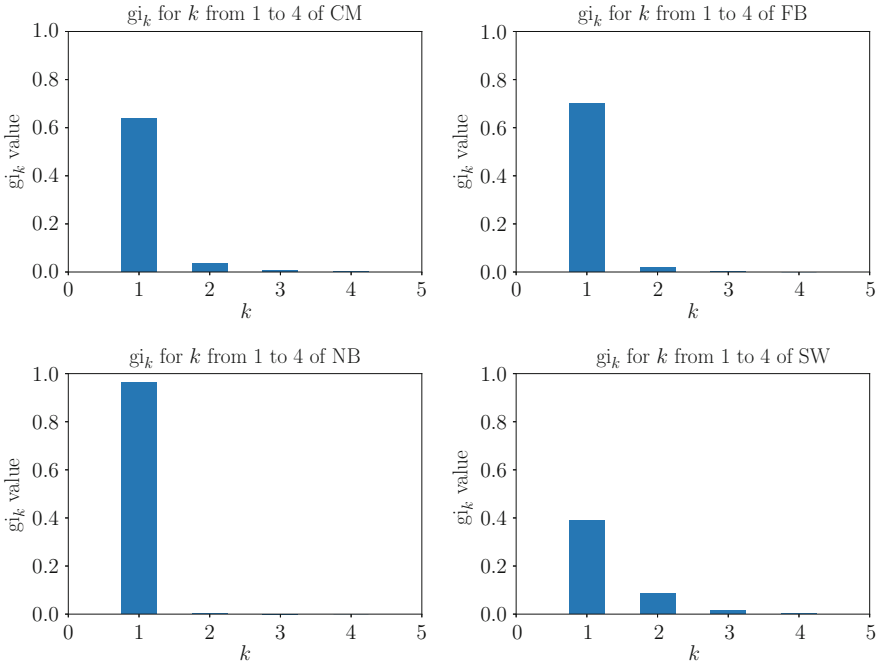


Fig. 3 Group individuality (g_i^k) for CM, FB, NB, and SW data sets (from top left to right bottom)

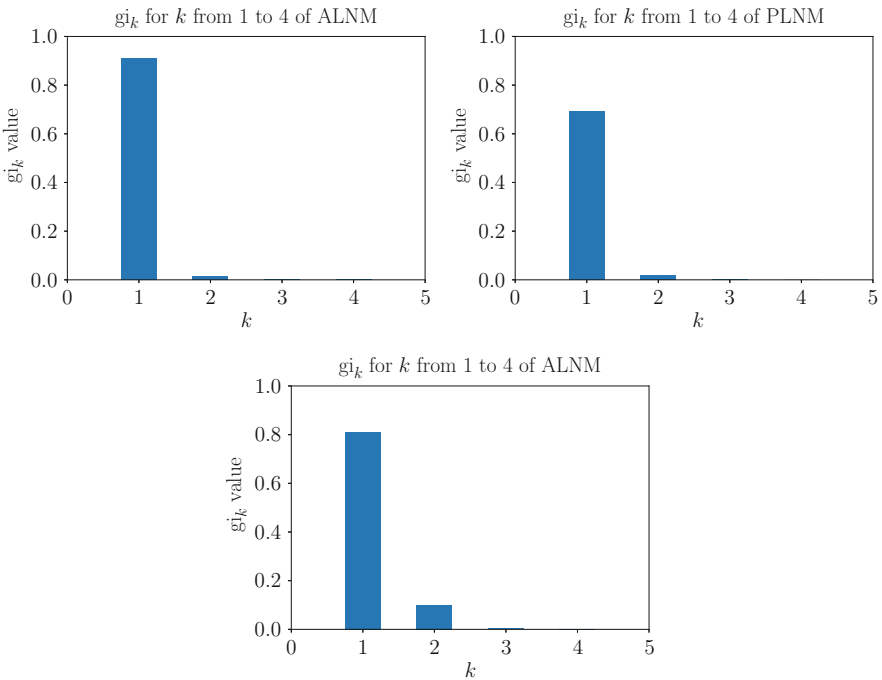


Fig. 4 Group individuality (g_i^k) for CM, FB, NB, and SW data sets (from top left to right bottom)

ASP, which would imply many common neighbors and therefore high probability for similar neighborhoods. A first guess could account the very low value of ALCC for this, which is untypical for small world networks. Yet, if we consider ALNM (ALCC of 0.52) and APLNM (ALCC of 0.71), we also observe very high values for 1-group individuality. Hence, in our experiments ALCC does not seem to be associated with 1-group individuality. This observation also carries over to 2-group individuality.

In general, no correlation of 1-group individuality with ALCC, ASP, or the size of the social network can be found in our results. This is particularly clear for the networks FB and CM, whose k -group individuality is similar, but which are very different in size. For 2-group individuality, the APLNM network shows the highest value with 0.10, followed by SW with 0.08. Indeed, these two data sets illustrate that there seems to be no connection between ALCC, ASP, or network size with the k -group individuality, and there is also no indication that k -group individuality depends in any way on the edge density of the network. Moreover, the amount of deviation of the null model to a data set cannot be connected to k -group individuality: the data sets of NB and APLNM are counterexamples to this, as both are similar in their k -group individuality, but differ significantly in their deviation to their null models. To sum up, all this substantiates our original intuition that group individuality is a completely new and independent measure for social networks.

As can be seen from the values of group individuality, this measure allows us to differentiate between the various networks by exhibiting qualitatively different values. Moreover, we can see that in all cases increasing the value of k results in k -group individuality to decrease significantly. This is indeed expected behavior from the definition of group individuality, since the denominator in gi_k is growing rapidly with k . However, from the perspective of understanding social networks, the low values of gi_k for $k \geq 1$ might itself be seen as a necessary property for a small world network: the formation of large groups definable by their milieu is something that can hardly be expected. Indeed, large values of k -group individuality for values $k > 1$ are usually a sign for *artificiality*: it is easy to generate a formal context, and hence a bi-modal network, with k -group individuality of 1 for $k > 1$, examples being *fixed row density contexts* [6]. Those formal contexts, however, possess a lot of symmetry and are thus highly artificial. On the other hand, in most of the investigated data sets we can still observe some non-zero values for $k = 2, 3$, and those values could represent intrinsic properties of the underlying network. Thus the *presence* of larger groups definable by their milieu could also be associated as a necessary property of social networks.

5.4 Individuality Distribution

In contrast to group individuality, group individuality distribution cannot be visualized in the usual manner, since the latter is a set instead of a simple number. Instead,

for every network represented by a formal context \mathbb{K} , we computed $gi_k(\mathbb{K})$ for every k from 1 up to the number of users G in the data set. We then identify the value $k_{\max} < |G|$ such that $k_{\max} \in \text{gid}(\mathbb{K})$, i.e.,

$$k_{\max} = \max(\text{gid}(\mathbb{K}) \setminus \{|G|\}).$$

To visualize $\text{gid}(\mathbb{K})$, we then plot its *indicator function* $\mathbf{1}_{\{i \in \mathbb{N} \mid i < k_{\max}\}}(\text{gid}(\mathbb{K}))$. The results are shown in Fig. 5.

The first thing we can read off from the diagrams is of course the corresponding values of k_{\max} , the size of the biggest individual group. Furthermore, the density of lines in the plot signifies the existence of groups of various sizes in the network: the more lines are present, the more groups of different sizes exist that are definable through their milieu. From this perspective of its individuality distribution, we perceive PLNМ as special, compared to the other networks, because its individuality distribution appears to be very dense. This is also the case for the SW network, because with fourteen users the value for k_{\max} of twelve is also very large. Moreover, comparing PLNМ with a data set of comparable size like FB, the structural difference between these networks can be spotted easily: for PLNМ the parameter k_{\max} is double as large as for FB. Therefore, even though both networks have similar values for ASP, ALCC, and even for user individuality, the PLNМ network seems more interesting than the FB network with respect to their individuality distributions. Indeed, we consider networks with a large value of k_{\max} to be more *interesting* from this point of view.

A more thorough examination reveals that none of the networks exhibits large gaps in their individuality distribution. This is a bit surprising, because one may have expected the existence of very large individual groups and also big gaps in the individuality distribution to the smaller groups, but this is not the case. In general, except for SW, no network exhibits big individual groups definable through their milieu compared to the number of its users.

Finally let us point out that although APLNM is a sub-network of ALNM, their individuality distributions are very similar, although their sizes differ significantly. Based on this observation, one could conjecture that a large part of the individuality of ALNM is already contained in APLNM, or put differently, that most of the individuality of ALNM comes from the APLNM sub-network. However, this conjecture requires further study that is not within the scope of this work.

5.5 Average Milieu Size

The results for our experiments on average milieu size are presented in Table 4 and Figs. 6 and 7. For every data set we computed ams_k for $k = 1, \dots, 7$. Indeed, for comparing these results with the ones in Sect. 5.3, a maximal value of $k = 4$ would have been sufficient. Yet we observed an interesting peak for the CM data

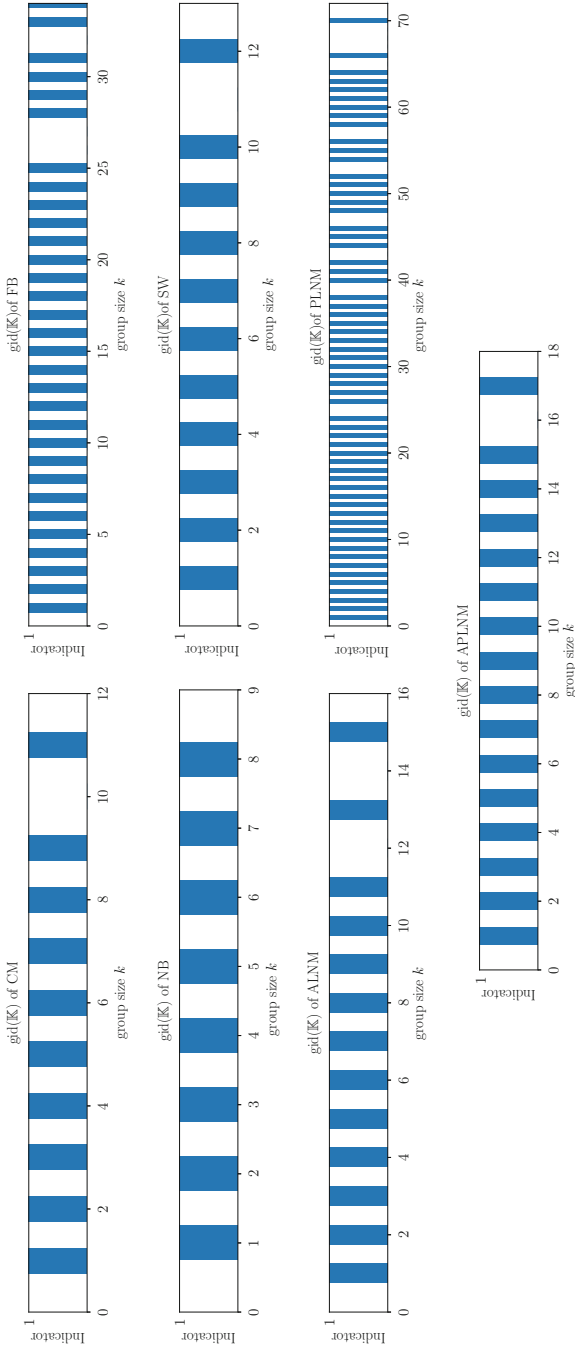


Fig. 5 Individuality distribution of all considered data sets

Table 4 Values of ams_k for $k = 1, \dots, 7$

Graph	ams_1	ams_2	ams_3	ams_4	ams_5	ams_6	ams_7
G_{CM}	4.25	3.50	2.50	1.75	1.00	3.00	2.00
G_{FB}	9.80	3.33	2.55	2.18	1.98	1.69	1.59
G_{ALNM}	4.58	2.17	1.45	1.47	1.25	1.33	1.14
G_{PLNM}	11.5	5.82	4.69	4.07	3.62	3.33	3.13
G_{APLNM}	13.7	5.04	3.39	2.65	2.21	1.91	1.69
G_{NB}	4.97	1.31	1.12	1.00	1.00	1.00	1.0
G_{SW}	7.14	5.23	3.85	2.70	2.43	2.00	2.0

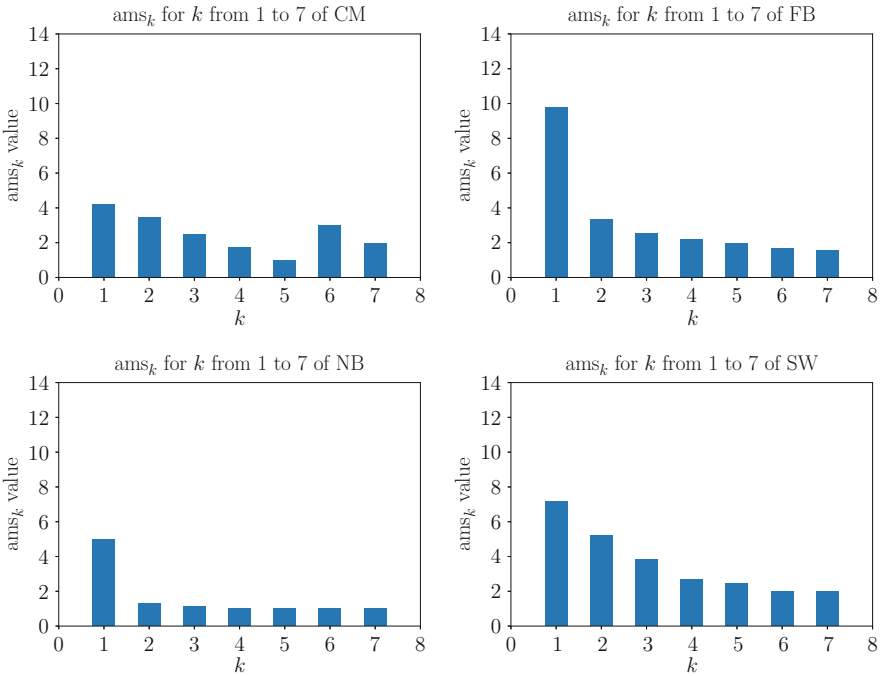


Fig. 6 Average milieu size (ams_k) for CM, FB, NB, and SW data sets (from *top left* to *right bottom*)

set, so we decided to show the result as seen. Additionally, we bounded for all plots the maximal value of their y-axis to fourteen to make comparison between the data sets easier.

Comparing the properties of the previous sections, as shown in Tables 1 and 2, to the values of average milieu size in our data sets, again no immediate correlation is visible, suggesting the independence of the introduced measure. In particular, a high value of group individuality does not imply anything on the average milieu size and vice versa. Moreover, the plot for the CM network reveals that, surprisingly, average milieu size does not necessarily need to be monotone in k , as suggested by the other plots.

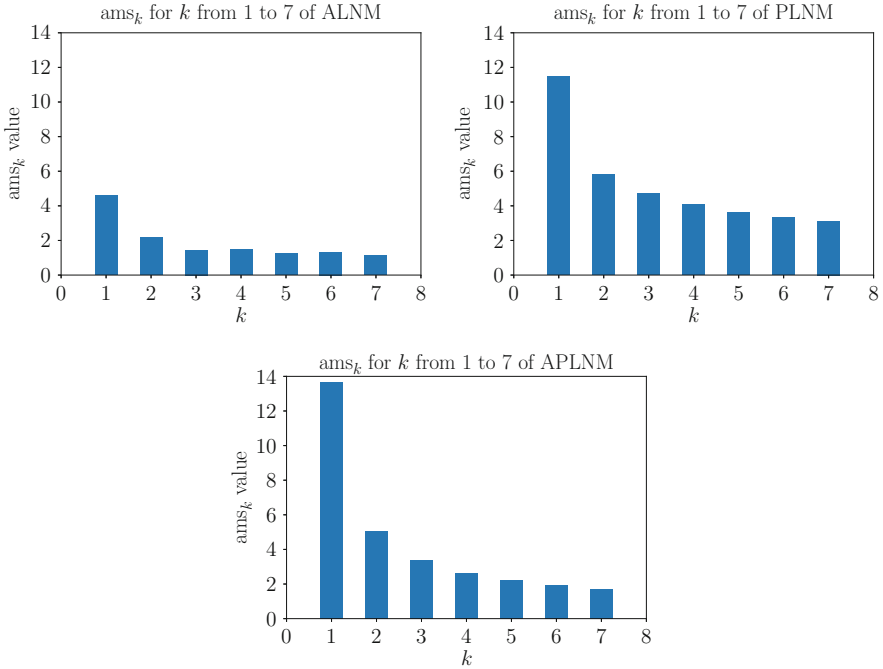


Fig. 7 Average milieu size (ams_k) for ALNM, PLNM, and APLNM data sets (from *top left* to *right bottom*)

Among the plots, the one for the NB network stands out for its low average milieu size for groups of size $k > 1$: on average all k -groups have about one attribute in common. Therefore, groups of two or more users rarely have an attribute in common. It is important to point out that in particular the average clustering coefficient is not able to represent this fact: compared to the NB network, the ALNM network has similar average milieu sizes, but a significantly larger value for ALCC.

An interesting observation in the plots is the difference between average milieu size for groups of size 1 compared to larger groups: there is usually a steep decline from the value of ams_1 to the one of ams_2 , say. One may consider a ratio between these values as a measure of how different the milieus of users are compared to those of larger groups of users.

Finally, as explained in Sect. 4, average milieu size can be perceived as a measure for the robustness of the number of k -groups. For this we observe that the APLNM data set reached a value of 14 for ams_1 , and hence the milieus of milieu-definable users consist on average of 14 attributes. We consequently conjecture the robustness of the user individuality in the APLNM network to be very high, but leave an experimental validation of this hypothesis for future work.

5.6 Discussion and Interpretation

The measures introduced in this work clearly represent facets of individuality in social networks, and it was the purpose of this Sect. 5 to demonstrate the usability and benefit of these quantities. To this end, we discussed various cases where the classical notion of small world networks suggests that two social networks were very similar, but where group individuality and its distribution revealed great structural differences.

The authors can only conjecture the reasons that lead to the observed results. For example, the very high user individuality in the NB network may be explained by strict rules for appointing board members. Especially the very low average milieu sizes for $k > 1$ lead to the impression that there are certain policies in place preventing “clubs” across boards.

The LNM data sets are somewhat special, since they are all intertwined. For example, each of them shows a high user individuality. Using k -group individuality, one may deduce that single users that were tracked during the event were in general more individual in their actions than the ones planning their evening. For the APLNM network, where both attendance and preferences were known, we observe values of user individuality between the ones of ALNM and PLNM. Yet, 2-group individuality is significantly larger in APLNM than in ALNM and PLNM. An interpretation could be that people who planned the evening beforehand are more likely to spend the evening in pairs of two.

To summarize, we claim that the benefit from having an instrument like group individuality is apparent. Furthermore, we assert that there is no method, known to the authors, to get comparable information from a social network.

We want to close this section with a note on the practicability of our approach. We refrained from giving concrete running times for our experiments, mostly because our implementations of the proposed algorithms are preliminary. Showing these values may nevertheless be worthwhile, in particular for arguing that our approach can be applied in practice. Because of this, we show the running times of all our experiments in Table 5. As can be seen from these numbers, the computation times for our data sets never posed a serious problem for the feasibility of our

Table 5 Running times of individual experiments, all times in seconds

	gi	gid	ams
G_{CM}	10.06	9.94	9.84
G_{FB}	170.67	212.94	210.18
G_{ALNM}	13.21	14.04	13.00
G_{PLNM}	1330.58	2302.80	2327.90
G_{APLNM}	13.64	16.00	15.70
G_{NM}	138.56	134.51	140.36
G_{SW}	10.10	10.69	10.44

approach. Moreover, all these running times can be greatly improved by using optimized implementations specifically designed for the fast computation of all formal concepts [1, 18].

6 Conclusions and Outlook

It was the purpose of this work to introduce a new measure on social networks that incorporates the notion of individuality in social networks, an approach that has not been examined before. For this we made use of ideas from formal concept analysis to provide a notion of milieu definability. Based on this, we developed in a natural way the notions of *group individuality*, *individuality distribution*, and *average milieu size*. Conducting experiments on real-world data sets, we were able to show that these new measures were both independent of previously known metrics like ASP and ALCC and allowed differentiating further otherwise similar networks. To sum up, we claim to have shown that the measures of individuality introduced in this work are both natural and meaningful.

This work has only started the study of our individuality measures, and it has not reached its end. For example, so far we have investigated individuality only on real-world networks, where this notion has a natural interpretation. However, we have not even started to look at individuality in networks that do not stem from real-world networks, and we do not know what values of individuality to expect there. In a similar vein, one could ask in how far group individuality is suitable to distinguish real-world networks from artificial ones.

Another aspect that requires further research is the scaling factor for k -group individuality. To improve comparability, we divide the number of extents of size k by $\binom{|G|}{k}$, the theoretical maximal number of such extents. Due to this scaling, k -group individuality is always between zero and one. However, this maximum is never achieved in practice and results in almost-zero values of k -group individuality for larger values of k , making those values virtually useless. Finding a better approach to scale k -group individuality is subject to further investigations.

In our experiments, the running times of our algorithms never posed a problem. However, for larger networks, measuring group individuality can represent a serious challenge: our methods require in the worst case the computation of the whole concept lattice of the representing formal context, and this lattice can be exponentially large. This somehow limits the usefulness of our approach, and further investigations are necessary to explore the possibilities of measuring group individuality of real-world networks.

The networks we have considered in this paper were bi-modal networks from the start, and the actual modeling of finding a suitable attribute set was not an issue. However, for uni-modal networks, finding a suitable set of attributes for a contextual representation may be difficult. To what extent group individuality can be adapted to this kind of networks remains an open problem and is subject to future research.

To establish the small world character of our used data sets, we employ the approach of using null models—something we have not yet done for our individuality measures. One of the main reasons for this is that generating null models for bi-modal networks has received attention from the research community only recently [19], and a proper evaluation is still missing.

A particular kind of social network that is not covered with our contextual representation are the so-called *tripartite* networks, sometimes also called *folksonomies* [12]. The corresponding structure in formal concept analysis is the one of a *triadic formal context*, and generalizing group individuality to those structures is also a promising line for future research.

Acknowledgements We thank R. Schaller and B. Ludwig for providing the LNM data set, which is a great addition to this work. Daniel Borchmann gratefully acknowledges support by the Cluster of Excellence “Center for Advancing Electronics Dresden” (cfAED). The computations presented in this paper were conducted by conexp-clj, a general-purpose software for formal concept analysis (<https://github.com/exot/conexp-clj>). Finally, we would like to thank the reviewers for their insightful comments on this work.

References

1. Andrews, S.: A ‘best-of-breed’ approach for designing a fast algorithm for computing fixpoints of Galois connections. *Inf. Sci.* **295**, 633–649 (2015). doi:10.1016/j.ins.2014.10.011
2. Atzmueller, M., Hanika, T., Stumme, G., Schaller, R., Ludwig, B.: Social event network analysis: structure, preferences, and reality. In: Proceedings of IEEE/ACM ASONAM. IEEE Press, Boston, MA (2016)
3. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
4. Birkhoff, G.: Lattice Theory. Colloquium Publications, vol. 25, 3rd edn. American Mathematical Society, New York (1967)
5. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Complex networks: structure and dynamics. *Phys. Rep.* **424**(4–5), 175–308 (2006). ISSN: 0370-1573
6. Borchmann, D., Hanika, T.: Some experimental results on randomly generating formal contexts. In: Huchard, M., Kuznetsov, S. (eds.) CLA. CEUR Workshop Proceedings, vol. 1624, pp. 57–69. CEUR-WS.org (2016)
7. Freeman, L.C.: Cliques, Galois lattices, and the structure of human social groups. *Soc. Netw.* **18**(3), 173–187 (1996). ISSN: 0378-8733
8. Freeman, L.C.: Finding social groups: a meta-analysis of the southern women data. In: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, pp. 39–97. National Research Council, The National Academies, Washington, DC (2002)
9. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg (1999)
10. Gjoka, M., Smith, E., Butts, C.: Estimating clique composition and size distributions from sampled network data. In: 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 837–842. IEEE, Toronto (2014)
11. Gkantsidist, C., Mihail, M., Zegura, E.: The Markov chain simulation method for generating connected power law random graphs. In: Proceedings of the 5th Workshop on Algorithm Engineering and Experiments, vol. 111, p. 16. SIAM, Philadelphia (2003)
12. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: Discovering shared conceptualizations in folksonomies. *J. Web Semant.* **6**(1), 38–53 (2008)

13. Kolaczyk, E.D.: *Statistical Analysis of Network Data: Methods and Models*. Springer Series in Statistics, p. 386. Springer, New York (2009)
14. KONECT (2016) Club membership network dataset
15. Maslov, S., Sneppen, K.: Specificity and stability in topology of protein networks. *Science* **296**(5569), 910 (2002)
16. Myers, S.A., Sharma, A., Gupta, P., Lin, J.: Information network or social network?: The structure of the Twitter follow graph. In: *Proc. WWW (Companion)*, pp. 493–498. ACM, Seoul (2014). ISBN: 978-1-4503-2745-9
17. Opsahl, T., Panzarasa, P.: Clustering in weighted networks. *Soc. Netw.* **31**(2), 155–163 (2009)
18. Outrata, J., Vychodil, V.: Fast algorithm for computing fixpoints of Galois connections induced by object-attribute relational data. *Inf. Sci.* **185**(1), 114–127 (2012). doi:10.1016/j.ins.2011.09.023
19. Saracco, F., Di Clemente, R., Gabrielli, A., Squartini, T.: Randomizing bipartite networks: the case of the World Trade Web. *Sci. Rep.* **5**, 10595 (2015)
20. Schaller, R., Harvey, M., Elswailer, D.: Detecting event visits in urban areas via smartphone GPS data. In: *Advances in Information Retrieval. Proc. ECIR*. Springer, Cham (2014)
21. Seierstad, C., Opsahl, T.: For the few not the many? The effects of affirmative action on presence, prominence, and social capital of women directors in Norway. *Scand. J. Manag.* **27**(1), 44–54 (2011)
22. Slater, N., Itzschack, R., Louzoun, Y.: Mid size cliques are more common in real world networks than triangles. *Netw. Sci.* **2**(3), 387–402 (2014). doi:10.1017/nws.2014.22
23. Vázquez, A., Pastor-Satorras, R., Vespignani, A.: Internet topology at the router and autonomous system level. In: *CoRR* (2002). cond-mat/0206084
24. Wasserman, S., Faust, K.: *Social Network Analysis. Methods and Applications. Structural Analysis in the Social Sciences*. Cambridge University Press, New York (1994)
25. Watts, D.J.: Networks, dynamics, and the small-world phenomenon. *Am. J. Sociol.* **105**, 493–527 (1999)
26. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* **393**(6684), 440–442 (1998). ISSN: 0028-0836. doi:10.1038/30918
27. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) *Ordered Sets: Proceedings of the NATO Advanced Study Institute*, pp. 445–470, Banff, 28 August–12 September 1981. Springer, Dordrecht (1982). ISBN: 978-94-009-7798-3
28. Xiao, W.-K., et al.: Empirical study on clique-degree distribution of networks. *Phys. Rev. E* **76**(3), 037102 (2007)
29. Zweig, K.A.: How to forget the second side of the story: a new method for the one-mode projection of bipartite graphs. In: *International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2010*, pp. 200–207, Odense, 9–11 August 2010. doi:10.1109/ASONAM.2010.24
30. Zweig, K.A., Kaufmann, M.: A systematic approach to the one-mode projection of bipartite graphs. *Soc. Netw. Anal. Min.* **1**(3), 187–218 (2011). doi:10.1007/s13278-011-0021-0