



# Group Measures and Modeling for Social Networks

Vincent Levorato

## ► To cite this version:

Vincent Levorato. Group Measures and Modeling for Social Networks. Journal of Complex Systems, 2014, 2014, pp.Article ID 354385. 10.1155/2014/354385 . hal-01105054

**HAL Id: hal-01105054**

**<https://hal.science/hal-01105054>**

Submitted on 19 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Research Article

# Group Measures and Modeling for Social Networks

**Vincent Levorato**<sup>1,2</sup>

<sup>1</sup> CESI, 959 Rue de la Bergeresse, 45160 Olivet, France

<sup>2</sup> LIFO, University of Orléans, 45067 Orléans, France

Correspondence should be addressed to Vincent Levorato; [vlevorato@cesi.fr](mailto:vlevorato@cesi.fr)

Received 20 May 2014; Revised 27 August 2014; Accepted 29 August 2014; Published 30 September 2014

Academic Editor: Juan Luis Cabrera

Copyright © 2014 Vincent Levorato. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Social network modeling is generally based on graph theory, which allows for study of dynamics and emerging phenomena. However, in terms of neighborhood, the graphs are not necessarily adapted to represent complex interactions, and the neighborhood of a group of vertices can be inferred from the neighborhoods of each vertex composing that group. In our study, we consider that a group has to be considered as a complex system where emerging phenomena can appear. In this paper, a formalism is proposed to resolve this problematic by modeling groups in social networks using pretopology as a generalization of the graph theory. After giving some definitions and examples of modeling, we show how some measures used in social network analysis (degree, betweenness, and closeness) can be also generalized to consider a group as a whole entity.

## 1. Introduction

Network modeling is an area of research which covers several domains like computer sciences, physics, sociology, or biology. In social networks modeling, graphs are often used to describe the links representing relationships or flows between entities [1]. Based on graph theory, the studies consider in most cases individuals as single elements, a group being formed by several persons interacting with each other. Most of the few works on modeling groups in social networks consider a group as a combination of persons [2], not as a whole entity. As social network analysis leads to centrality notion and others sociometric features, what about group centrality?

The centrality of a vertex in a graph is widely used to determine the relative “importance” of this vertex within the network [3]. Centrality measures enable us to find users who are extensively involved in relationships with other network members. There are different centralities such as degree centrality, betweenness centrality, or closeness centrality. The problem we face is the following: analyzing a vertex can be done with this kind of measure, but if we analyze a group of persons using the same measure, we will have no particular emergence of characteristics as the union property of the neighborhoods in a graph is preserved. As social networks

are complex networks [4–6], emergence of phenomena can occur [7], and the behavior of a group of persons can be different from the “sum” of the person behaviors composing the group. Some work tried to capture the different scales of a network, and a group can be viewed as a community [8]; thus, in our opinion, graph theory only is inadequate to model all complex interactions occurring in a social network. Some group-based measures based on graphs have been proposed before [2, 9, 10], and multilayer analysis is a topic widely studied [11–13] combining multiple graphs; however, to our knowledge, there is no general framework coupling the modeling of a group of entities with a multirelational network. Based on this observation, we propose to use a mathematical framework for group modeling in social networks including several relation types and new group-based measures from a topological point of view.

This paper is structured in two parts: first we introduce pretopology formalism, and we give the definitions and some illustrations for group modeling in a social network, which are reminders of a previous work [14]. Then we define new group measures with pretopology generalizing known classical methods. The reader of this paper is supposed to be familiar with graph theory and some concepts from mathematical topology.

## 2. Modeling Groups in Social Networks

Before giving theoretical definitions, we first illustrate the problem of a realistic group modeling by giving a simple example: we would like to represent interactions between four persons. A small social network with a simple symmetrical friendship relation will help in illustrating our approach (Figure 1).

Let focus on John and Ed. John has three friends: Tim, Ben, and Ed, while Ed has two friends: Ben and John. As we notice, it is very easy to know who are the friends of a person. But now, we want to know whose people are the friends of the group  $\{\text{John}, \text{Ed}\}$ . In this case, having a look at the nonoriented graph induced by the network, we see Tim in relation with John, and Ben in relation with John and Ed. As we can reach Tim and Ben from John and Ed, we could say that Tim and Ben are friends with John and Ed. However, in reality, Ed and Tim are not friends. We need to set a constraint, saying, for instance, the friends of the group  $\{\text{John}, \text{Ed}\}$  must be friends with each person of the group. So, we finally have only Ben as a friend of the group  $\{\text{John}, \text{Ed}\}$ . The operation we made in this small example is defining the neighborhood of an element and, more generally, of a set of elements. This example can be extrapolated to large known online social networks like Facebook or Twitter. Thus, we need a theory able to precisely model the neighborhood of sets: here we thought of pretopology formalism.

Pretopology is a mathematical modeling tool for the concept of proximity suitable for discrete spaces [15, 16]. It provides also powerful tools for structure analysis and automatic classification [17]. Pretopology theory generalizes topology theory and has weaker axioms than topology, allowing us to analyze transformation procedures step-by-step and handle a multirelation description. As pretopology formalism is based on set theory, a group of individuals is now considered as a set, allowing us to consider a group as a whole independent entity.

**2.1. Pretopological Notions.** The pretopology formalism comes from classical topology but has weaker axioms. The “closure” process is not necessary idempotent and only the extensivity axiom is required. Reference [18] gave the first definition of a pretopological space. Fréchet spaces [19], Kuratowski’s closure axioms [20], and Čech closure operator [21] have been preponderant in definition of a pretopological space and of the pseudoclosure map (Definition 1 and Figure 2).

**Definition 1 (pseudoclosure).** Let consider a nonempty set  $E$  and  $\mathcal{P}(E)$  which designates all the subsets of  $E$ . A pseudoclosure map, denoted by  $a(\cdot) : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ , is a map meeting the following conditions:

- (1)  $a(\emptyset) = \emptyset$ ,
- (2)  $\forall A \in \mathcal{P}(E), A \subseteq a(A)$ .

A pretopological space is a set  $E$  endowed with a pseudoclosure map.

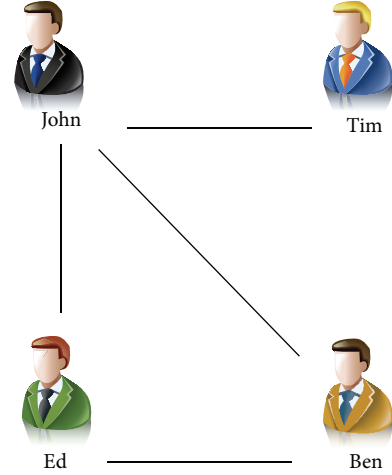


FIGURE 1: A small social network with friendship relation.

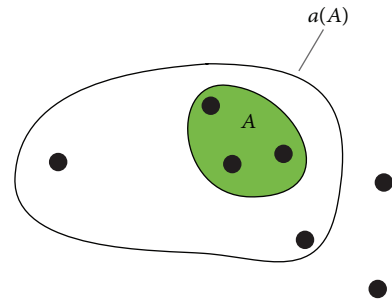


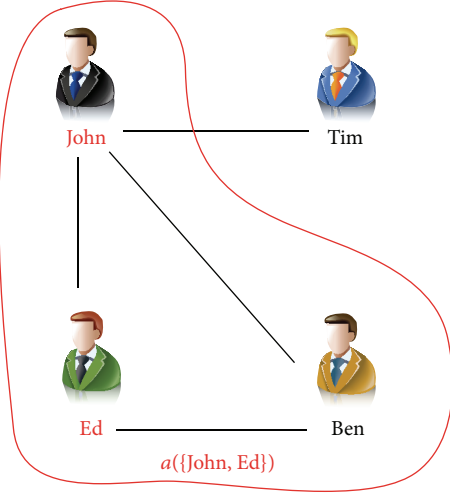
FIGURE 2: Pseudoclosure of  $A$ .

**Definition 2 (pretopological space).** A pretopological space is a pair  $(E, a)$ , where  $E$  is a nonempty set and  $a$  is a pseudoclosure map  $a(\cdot) : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ .

The pseudoclosure operator is useful to model dynamic processes in the considered space  $(E, a)$ . In social networks, we can say that elements belonging to  $a(A)$  are close to  $A$ , allowing us to determine direct neighbors. For each pseudoclosure, we absorb new elements  $(A \subseteq a^2(A) \subseteq a^3(A) \subseteq \dots \subseteq a^k(A))$  which are more and more “distant.” Hence, we are able to model complex dynamics like, for instance, information diffusion process. The definition of the neighborhood can also be based on several different relations (see [16] for the complete mathematical definitions).

To understand this approach, let us return to our network example with the four friends and define the problem by applying the pretopology formalism. We build a pretopological space  $(E, a)$  with four elements in  $E$  and we build  $a(A)$  with  $A \in \mathcal{P}(E)$  as

$$\begin{aligned}
 a(A) \\
 &= A \cup \text{people who are friends with each person of } A.
 \end{aligned}
 \tag{1}$$

FIGURE 3: Pseudoclosure of  $\{\text{John}, \text{Ed}\}$ .

With the same assumptions, we make the same operations as in the previous example:

- (i)  $a(\{\text{John}\}) = \{\text{John}, \text{Tim}, \text{Ben}, \text{Ed}\}$ ,
- (ii)  $a(\{\text{Ed}\}) = \{\text{Ed}, \text{John}, \text{Ben}\}$ ,
- (iii)  $a(\{\text{John}, \text{Ed}\}) = \{\text{John}, \text{Ed}, \text{Ben}\}$  (Figure 3).

As we notice,  $a(\{\text{John}, \text{Ed}\}) \neq a(\{\text{John}\}) \cup a(\{\text{Ed}\})$ . We have also  $\{\text{John}\} \subseteq \{\text{John}, \text{Ed}\}$  but not  $a(\{\text{John}\}) \subseteq a(\{\text{John}, \text{Ed}\})$ . Pretopology helps us define interactions which are not convenient to describe with other formalisms.

In a network, diffusion process modeled by the pseudoclosure can stop its progression. Such a configuration is called a closure [22] (Figure 4).

**Definition 3** (closed subset and closure). Let  $(E, a)$  be a pretopological space:

- (i) a subset  $B$  of  $E$  such as  $a(B) = B$  is called a closed subset of  $E$  for  $a(\cdot)$ . The notation  $F(\cdot)$  corresponds to the closure operator,
- (ii) an elementary closed subset, denoted as  $F_x$ , is the closure of a one element set  $\{x\}$  of  $E$ . The notation  $F(\{x\})$  is equivalent.

Closure is very important because of the information it gives about the “influence” or “reachability” of a set, meaning, for example, that a set  $A$  can influence or reach elements into  $F(A)$ , but not further.

**2.2. Modeling Group Interactions.** First, we have to define a social network with pretopology concepts. A social network is a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of binary or valued relations [23].

In pretopology, we can generalize this definition by the fact that a (social) network is a family of pretopologies on a given set  $E$  (Figure 5) [24].

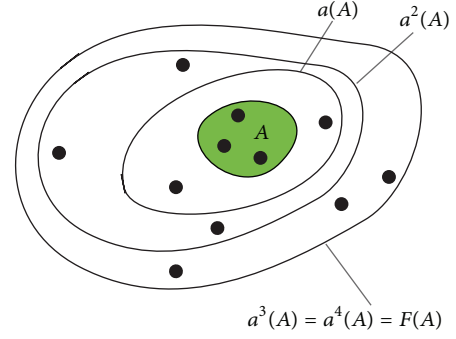


FIGURE 4: Iterated application of the pseudoclosure map leading to the closure.

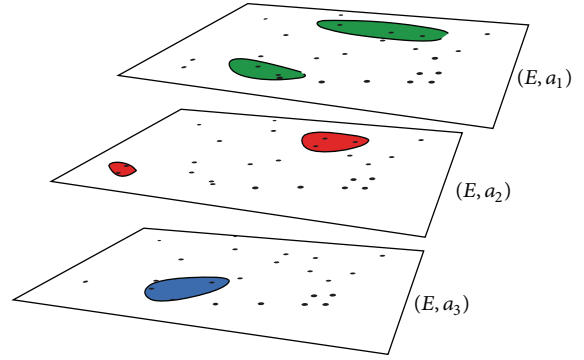


FIGURE 5: Social network defined with pretopology and composed of three different pseudoclosures.

**Definition 4** (network). Let  $E$  be a nonempty set.

Let  $I$  be a countable family of indexes.

Let  $\{a_i, i \in I\}$  be a family of pretopologies on  $E$ .

The family of pretopological spaces  $\{(E, a_i), i \in I\}$  is a (social) network on  $E$ .

Different kinds of relations with different natures can be represented: for instance, we can model a social network containing work colleagues and geographic relations. We can build a pseudoclosure answering to a modeling problem, having, for example, a person close to another if they are friends, working together, and living close to each other (using a given distance threshold). The following paragraphs give to the reader examples of how we can build pseudoclosures depending on the data or information we want to use for our modeling.

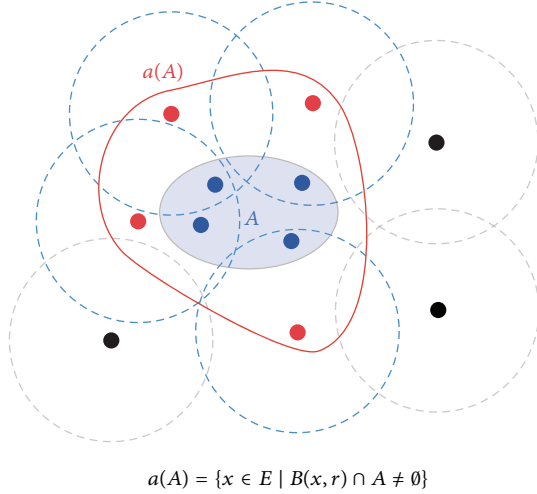
**Example 5** (metric space). We represent the relations between people with an euclidean distance (the people are represented as nodes), considering, for instance, the geographical location of each person.

Let  $E$  be endowed with a metric defined by a distance  $d$ . Let  $r$  be a positive real. For each element  $x$  of  $E$ ,  $B(x, r)$  is a ball with the center  $x$  and a radius  $r$  defined by

$$B(x, r) = \{y \in E \mid d(x, y) \leq r\}. \quad (2)$$

A pseudoclosure  $a(\cdot)$  can be built with  $B(x, r)$ :

$$\forall A \in \mathcal{P}(E), a(A) = \{x \in E \mid B(x, r) \cap A \neq \emptyset\}. \quad (3)$$

FIGURE 6: Pseudoclosure of  $A$  in a metric pretopological space.

The pseudoclosure  $a(A)$  is composed of all elements of  $A$  and all elements  $y \notin A$  such that  $y$  is within a distance of at most  $r$  from at least one element of  $A$  (Figure 6).

*Example 6* (binary space). In social networks, we are confronted to nonmetric relations, representing qualitative information (friendship relation, etc.). In such a space, the elements of  $E$  are bound by  $n$  reflexive binary relations  $R_i$  with  $i \in \mathbb{N}^*$ .  $R$  is not necessarily symmetric. We define  $\forall i \in \{1, \dots, n\}$ :

$$\begin{aligned} R_i(x) &= \{y \in E \mid xR_i y\}, \\ R_i^{-1}(x) &= \{y \in E \mid yR_i x\}. \end{aligned} \quad (4)$$

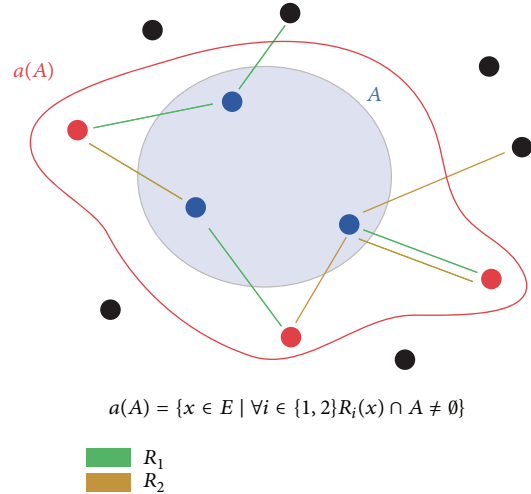
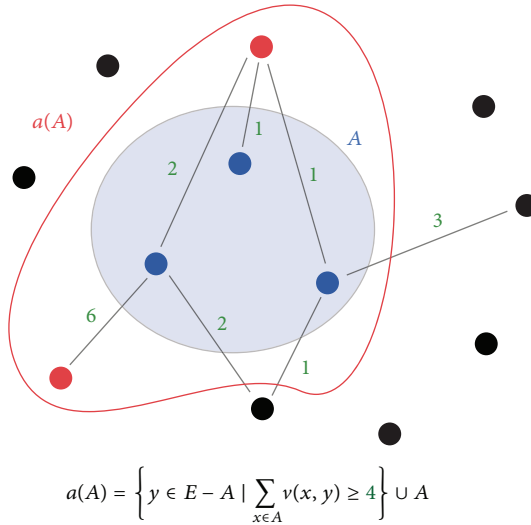
As the relation  $R$  is reflexive,  $x$  belongs to  $R(x)$  and to  $R^{-1}(x)$ . We can construct two pseudoclosures:

$$\begin{aligned} \forall A \in \mathcal{P}(E), \\ a(A) &= \{x \in E \mid \forall i \in \{1, \dots, n\}, R_i(x) \cap A \neq \emptyset\}, \\ \forall A \in \mathcal{P}(E), \\ a(A) &= \{x \in E \mid \forall i \in \{1, \dots, n\}, R_i^{-1}(x) \cap A \neq \emptyset\}. \end{aligned} \quad (5)$$

These pseudoclosures are not equivalent when  $R$  is not symmetric. The pseudoclosure  $a(A)$  is composed of  $A$  ( $R$  is reflexive) and of all elements  $y$  which have  $R_n$  (or  $R_n^{-1}$ ) relations with, at least, one element of  $A$ . An illustration of this kind of space with two binary relations is illustrated in Figure 7. In this example, an element  $x$  of  $E$  belongs to pseudoclosure of  $A$  if  $x$  is in relation with an element of  $A$  according to relation  $R_1$  and  $R_2$ .

*Example 7* (valued space). In order to model certain problems, binary relations are not sufficient. We therefore need to have a value (integer, real, function, ...) on the links. In this kind of space, the elements of  $E$  are bound by a valued relation. For instance, we can define an integer value  $v$  on relations as

- (i)  $E \times E \rightarrow \mathbb{N}$ ,
- (ii)  $(x, y) \rightarrow v(x, y)$ .

FIGURE 7: Pseudoclosure of  $A$  in a binary pretopological space.FIGURE 8: Pseudoclosure of  $A$  in a valued pretopological space.

We can build a pseudoclosure giving different results between singletons and sets ( $E - A$  and  $A^c$  are equivalent notations corresponding to the complementary of  $A$ ):

$$\forall A \in \mathcal{P}(E), \quad a(A) = \left\{ y \in E - A \mid \sum_{x \in A} v(x, y) \geq s \right\} \cup A, \quad (6)$$

where  $s$  is an integer.

The pseudoclosure  $a(A)$  is composed of  $A$  and of all elements  $y$  where the sum of valued edges between some elements of  $A$  and  $y$  is greater than the threshold  $s$ . Figure 8 gives an illustration of this space with  $s = 4$ . This kind of modeling can be used in social networks where weighted relations are necessary and illustrates the interest of the pretopology modeling. Indeed, this example shows that group behavior is different than the “sum” of individuals composing

it. In Figure 8, the person at top (bold) is absorbed because he knows 2 persons (value = 1) a little and better another one (value = 2), so he can be considered as a friend of group  $A$ . If we take each individual of  $A$  saying this external individual, the person, at top, is friend of  $A$ ; if the value of a link is superior to  $s = 4$ , he will not be taken into account.

**2.3. Example of a Complex Group Interactions Model.** We show how we can model social networks with complex interactions. The following example is taken from [25] study, a dataset which represents several relations between monks in a cloister and has been often used in sociology studies. It concerns the social interactions which have been collected as numerous sociometric rankings. A lot of relations are coded, but we focus on two of them: esteem and disesteem. Relations are nonsymmetric and weighted with three integer values, from 1 to 3 (3 indicates the highest or first choice and 1 the last choice). We built our model on the following assumptions: a person  $y$  is close to another person  $x$  if

- (i) there is an esteem relation from  $y$  to  $x$  according to a chosen threshold,
- (ii) there is no disesteem relation from  $x$  to  $y$  according to a chosen threshold.

This proximity definition makes sense if we consider group coalitions: if a person asks other people “who wants to join my group?”, people who have the greatest esteem for him should join his group more than the others. But this person will not accept in his group people he does not esteem, even if these people have esteem for him. We set two valued nonsymmetric relations with  $E$  the set containing individuals:

- (i)  $E \times E \rightarrow \mathbb{N}$ ,
- (ii)  $(x, y) \rightarrow \text{est}(x, y)$ ,
- (iii)  $(x, y) \rightarrow \text{disest}(x, y)$ ,

with  $\text{est}$  being the esteem relation and  $\text{disest}$  the disesteem relation.

We build our pseudoclosure as

$$\begin{aligned} \forall A \in \mathcal{P}(E), \\ a(A) = \left\{ y \in E - A \mid \sum_{x \in A} \text{est}(y, x) \right. \\ \left. \geq \alpha \wedge \sum_{x \in A} \text{disest}(x, y) < \beta \right\} \cup A \end{aligned} \quad (7)$$

with  $\alpha, \beta \in \mathbb{N}$ .

With  $\alpha = 3$  and  $\beta = 1$ , we put a strong restriction on the neighborhood: if we apply this pseudoclosure on a set  $A$ , people would be in  $a(A)$  if they have esteem with one or several persons of  $A$  (according to  $\alpha$ ) and if no people of  $A$  have disesteem for them (see Figure 9).

In this model, the question of finding the largest group following the rules of our neighborhood can be treated by building  $\mathcal{F}_e$ , the family of all elementary closed subsets of  $E$  (see Definition 3.). When closure is applied to each singleton

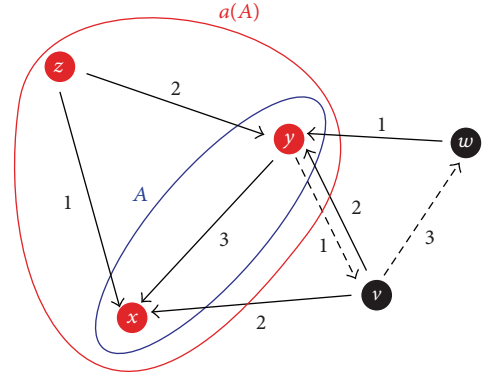


FIGURE 9: Pseudoclosure of  $A$  with  $\alpha = 3$ ,  $\beta = 1$ . Esteem relation is plain and disesteem relation is dashed.

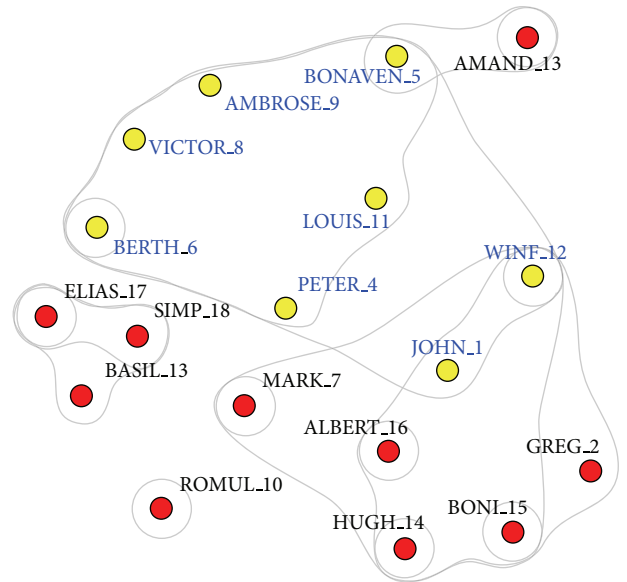


FIGURE 10: Largest group found according to the neighborhood definition.

of  $E$ , it reveals that VICTOR.8 is the person who can rally the biggest amount of people in the network (Figure 10).

To highlight the interest of this formalism for centrality measures, using the network sample Figure 9, we give the example of the degree centrality (see Section 3.1 for definitions). First, degree definition in weighted directed graph changes from the classical definition, as we can, instead of counting incoming or outgoing arcs, sum all the weights of incoming or outgoing arcs. In this example, there are two relations, giving two graphs. The set  $A$  contains two vertices  $x$  and  $y$ : as we can easily compute degree values for each vertex with each relation (e.g., we choose the incoming arcs giving  $\deg_{\text{Est-in}}(x) = 6$ ,  $\deg_{\text{Est-in}}(y) = 5$  by summing weights or only use the classical degree definition giving  $\deg_{\text{Est-in}}(x) = 3$  and  $\deg_{\text{Est-in}}(y) = 3$ ), and even for the set  $A$  ( $\deg_{\text{Est-in}}(A) = 8$  by summing weights or  $\deg_{\text{Est-in}}(A) = 5$  using classical degree definition), but we cannot directly answer the problem we are working on, meaning representing a degree for people who



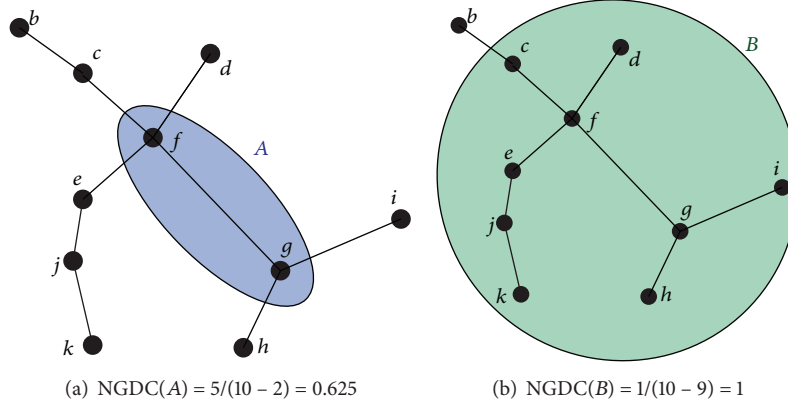


FIGURE 11: Bias in normalized extended degree centrality group measure.

wants to join a group. We should put others rules like “an esteem arc is taken into account if its weight is superior to  $\alpha$ ” and we should also consider disesteem to obtain the same results we can have directly with pretopology which are, in this case,  $\deg(x) = |a(\{x\}) - \{x\}| = 1$ ,  $\deg(y) = |a(\{y\}) - \{y\}| = 0$ , and  $\deg(A) = |a(A) - A| = 1$ . Different kinds of rules can be transparently inserted into the pseudoclosure map to represent a special type of interaction (logic conditional statements, functions, Markov processes, etc.).

General examples of social networks modeled with pretopology and complex interactions modeling have been presented. Building a pseudoclosure map depends on the network nature and problem representation and can need several pseudoclosures to be combined to obtain specific results. Several previous works have investigated in modeling complex systems with pretopology and have shown some interesting results in modeling the impact of geographic proximity on scientific collaborations [26], structuring and clustering data [17], modeling pollution phenomena [27], analyzing the evolution of a communicable disease [28], or analyzing communities of the Web [29]. This section showed the interest of modeling groups in multirelational social networks by applying pretopology. In the next section, existing measures used in social network analysis are adapted in order to give more general and relevant results for group measurement.

### 3. Group Measures for Social Network Analysis

In social network analysis, choosing the “good” measure depends on the problem we are working on and on the properties we would like to highlight. Measures are used to describe the features of actors within social networks as well as to indicate personal importance of individuals [2]. The centralities are some of these measures and are usually applied to individual actors. However, there are many situations when it would be advantageous to be able to apply them to a group or a set of actors. The proximity of actors would be dependent on each actor neighborhood but also on a neighborhood group formed by the actors. It would be useful to identify

emergent groups in a network (such as cliques or structural equivalence in graphs). In another way, we would like to construct groups that have maximal centrality, for example, to maximize information diffusion through a particular group. Another thing is the ability to optimize the efficiency of a group by removing actors who would not or slightly reduce the centrality of the group.

Moreover, the current original and extended centralities measures have some weakness [30] that the group versions resolve:

- (i) classical centralities are rarely used in multirelational networks, as the complex networks consist of more than one type of relationship,
- (ii) there is a lack of applicability in weighted networks,
- (iii) disconnected networks are not handled by methods based on the shortest paths (if a given node cannot be reachable by one of the rest nodes, it results in not relevant outcomes).

Furthermore, if groups are seen as sets in a pretopological space with complex neighborhoods, groups cannot be measured with classical centralities. Some work goes beyond simple graph theory using hypergraph formalism [31], but it is still not sufficient to model multirelational and dynamic networks [32, 33].

In this section, group measures are not normalized because it can bring some bias in the measures. If we take the normalized group degree centrality (NGDC) in [2] study, normalizing the measure benefits larger group (Figure 11) and gives erroneous results. Constructing groups with maximal centrality becomes a problem, and working without normalization allows for making this construction where resulting groups will have a pertinent size.

*Remark Notation.* The difference between two sets noted as  $A - B$  means we remove  $A \cap B$  from  $A$ .

**3.1. Group Degree Centrality.** In graph theory, we simply call this measure the degree. In a nonoriented graph, the degree of a node is the number of edges incident upon a node. In an oriented graph, we distinguish for a vertex  $v$  its indegree and

outdegree. The indegree corresponds to the number of arcs with  $v$  as the target of the arc. The outdegree corresponds to the number of arcs with  $v$  as the source of the arc. The degree of  $v$  is equal to the sum of indegree and outdegree of  $v$  (see [34]).

In sociology, degree centrality measures locally the individual capacity to communicate, independent of its neighbor centralities [35, 36]. Degree centrality is the simplest and most intuitive measure that can be used in network analysis.

**Definition 8** (group degree centrality). We define the group degree centrality in a pretopological space  $(E, a)$  denoted by  $C_d^G(A)$ , where  $A \in \mathcal{P}(E)$  such as

$$C_d^G(A) = |a(A) - A|. \quad (8)$$

In pretopology, this process already exists and is called orle or external semifrontier defined by  $o(A) = a(A) - A$ . The group degree centrality returns the cardinality of this process equal to the amount of direct neighbors of a group.

**Example 9.** Considering Figure 9,  $C_d^G(A) = 1$ .

**3.2. Group Betweenness Centrality.** Betweenness centrality has been proposed by [37]. The concept of this measure is based on the idea that an individual can be weakly connected to others and be distant from them, but this individual can nevertheless be an important intermediary, for example, concerning information exchange between other individuals of the network. The stronger the intermediary role of a person is, the more he or she can take control over the communications or be independent of other members. Such a person can easily influence the network by filtering or distorting information circulating on it. This special position allows an assurance of the network coordination. Here is the classical betweenness centrality definition.

**Definition 10** (betweenness centrality). Let  $n$  be the number of vertices of a graph, let  $g_{jk}$  be the number of geodesic (shortest) paths from vertex  $j$  to vertex  $k$ , and let  $g_{jk}(i)$  be the number of these paths passing through vertex  $i$ , and we note  $C_b(i)$  the betweenness centrality of the vertex  $i$  such as

$$C_b(i) = \sum_j \sum_k \frac{g_{jk}(i)}{g_{jk}} \quad (9)$$

with  $j \neq k \neq i$  and  $j < k$ .

The original measure has been extended to measure the amount of geodesic paths passing through a group [9]. It gives more general results, but the problems resulting from this method remain the same: individuals in a group are treated separately. Reference [38] proposed a more precise measure, but it can only be applied to simple graphs with only one type of relation, and their framework cannot be reused for other centralities measures.

Our approach for processing a general betweenness centrality is slightly different from Freeman's approach: we do not consider paths as in graph theory: there is a path from  $A$  to  $B$  if  $B$  is entirely reachable from  $A$  [15].

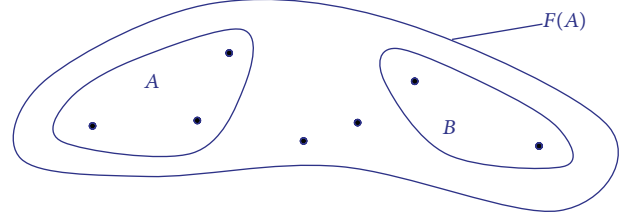


FIGURE 12: Pretopological path from  $A$  to  $B$ .

**Definition 11** (path). Let  $(E, a)$  be a pretopological space,  $\forall A, B \in \mathcal{P}(E)$ , and there is a path from  $A$  to  $B$  only and only if  $B \subseteq F(A)$  (Figure 12).

One can deduce from Definition 11 that there is a path from  $A$  to  $B$  also if  $B \subseteq a^k(A)$  as  $a^k(A) \subseteq F(A)$ .

**Definition 12** (group betweenness centrality). Let  $(E, a)$  be a pretopological space, and  $k, r \in \mathbb{N}$ . Let us define  $p_{xy}$  as a pretopological path between singletons  $\{x\}$  and  $\{y\}$  of  $E$  meaning  $\{y\} \subseteq a^k(\{x\})$ . We note  $p_{xy}(A)$  as a pretopological path (Figure 13) where

$$\exists A \in \mathcal{P}(E) \mid A \subseteq a^k(\{x\}) \wedge \{y\} \subseteq a^r(\{x\}) \wedge k \leq r. \quad (10)$$

The group betweenness centrality of a set  $A$  is defined as

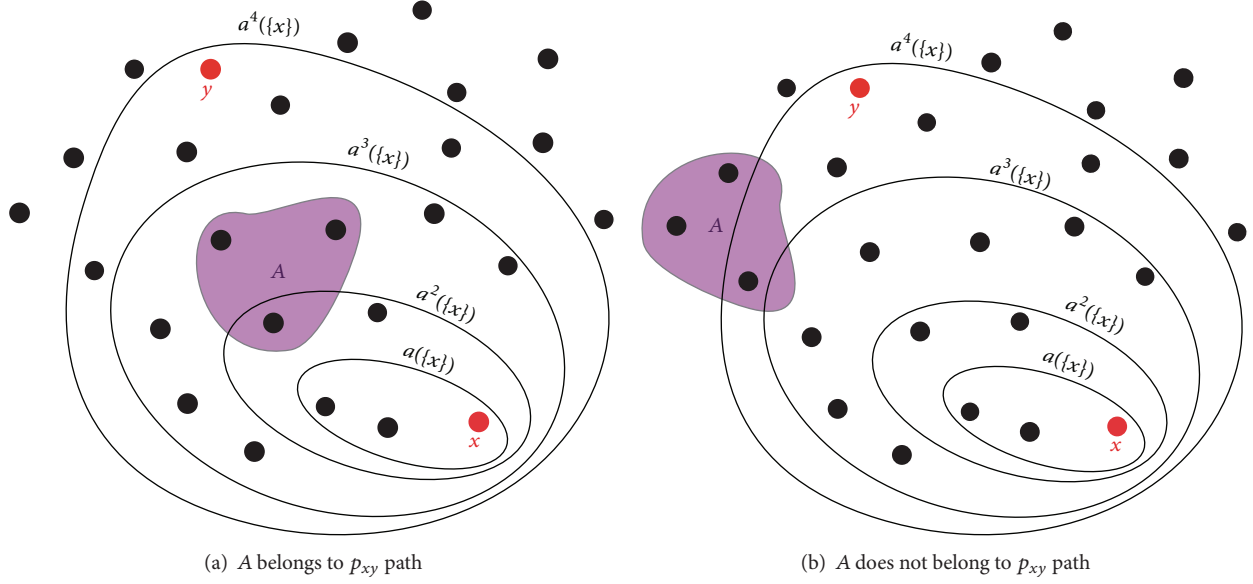
$$C_b^G(A) = \sum_x \sum_y \phi(A) \quad (11)$$

with  $x \neq y$  and  $\phi(A) = \begin{cases} 1 & \text{if } \exists p_{xy}(A) \\ 0 & \text{else} \end{cases}$ .

**Experiments.** To illustrate the interest of the generalized group betweenness centrality measure, we applied it to the network model used in Section 2.3 part using only one relation (esteem) with the same threshold value ( $\alpha = 3$ ). The problem to resolve becomes slightly different: we want to find the groups which give the best score. As all parts of a set cannot all be tested, we rely on the construction of the family of elementary closed subsets. For each singleton  $\{x\}$ , while processing its closure  $F_x$ , we measure the group betweenness centrality of the intermediate sets obtained. Having the configuration  $\{x\} \subseteq a^2(\{x\}) \subseteq a^3(\{x\}) \subseteq \dots \subseteq a^k(\{x\}) \subseteq F_x$ , we measure the group betweenness centrality on each  $a^k(\{x\})$ . The use of a single relationship is justified by the fact that we want to compare our measure to the classic version [2] which cannot handle multirelational networks.

To be able to compare the results, we changed the classic group betweenness centrality (GBC) definition: for a couple  $\{u, v\}$  in the network, considering the GBC of a set  $C$ , we increment its value of 1 if at least one shortest path passing through  $C$  exists. It follows the idea developed in Definition 12. The results (Remark: as classic GBC is based on the shortest paths in a graph, we inversed values, meaning that if a link has its esteem value set to 3 in the pretopological model (max value), we change the value in the network (Section 2.3) to 1 (min value) which corresponds here to a



FIGURE 13: Pretopological path  $p_{xy}$  illustration.

“cost” for a shortest path algorithm) of the group betweenness centralities are given in Table 1. In this table, we highlighted the three groups with the highest GBC values: results between classical and generalized GBC show that our measure is able to detect emergent persons who have not been considered with the classical measure.

It is interesting to find matches in the results from both methods but also to see that a pretopological model highlights some groups that should not have been considered otherwise.

**3.3. Generalized Closeness Centrality.** The closeness of an element  $x$  describes the total distance of  $x$  to all other elements in the network [39, 40]. For calculation, sum the shortest paths  $d(x, x_j)$  of a node  $x$  to all other nodes in the network:

$$\sum_j^n d(x, x_j), \quad (12)$$

where  $x \neq x_j$ ,  $d(x, x_j)$  the shortest path between nodes  $x$  and  $x_j$ , and  $n$  the total number of nodes in the network.

Closeness can be regarded as a measure of how long information will take to spread from a given vertex to other reachable vertices in the network [41].

**Definition 13** (closeness centrality). The closeness centrality  $C_c(x)$  for a vertex  $x$  is standardized on the size of the network:

$$C_c(x) = \frac{n-1}{\sum_j^n d(x, x_j)}. \quad (13)$$

With pretopology, we can generalize this concept, as the pseudoclosure is a function that represents direct proximity and the closure which represents the largest proximity of a set. Our measure is built on the eigenvector centrality and power

TABLE 1: Group betweenness centrality scores of a sample network obtained by classic and generalized methods.

Group	Classic GBC
<b>JOHN_1, VICTOR_8</b>	75
JOHN_1	64
VICTOR_8	48
PETER_4	47
JOHN_1, HUGH_14, WINF_12, BASIL_3, GREG_2	43
LOUIS_11, PETER_4	42
JOHN_1, HUGH_14, WINF_12, AMBROSE_9, VICTOR_8, BASIL_3, GREG_2	34
<b>AMBROSE_9</b>	32
<b>LOUIS_11</b>	32
GREG_2	27
Group	Generalized GBC
<b>JOHN_1, VICTOR_8</b>	39
<b>AMBROSE_9, LOUIS_11</b>	38
LOUIS_11, PETER_4	37
BERTH_6	33
BONAVENT_5, AMBROSE_9, LOUIS_11, PETER_4	33
BONAVENT_5, VICTOR_8, AMBROSE_9, BERTH_6, PETER_4	33
BONAVENT_5, VICTOR_8, AMBROSE_9, LOUIS_11, BERTH_6, PETER_4	29
BONAVENT_5, AMAND_13	27
AMAND_13	27
HUGH_14	22

centrality ideas [42, 43]. Regarding a set, the further away an element is from this set, the less it is influenced by this set.

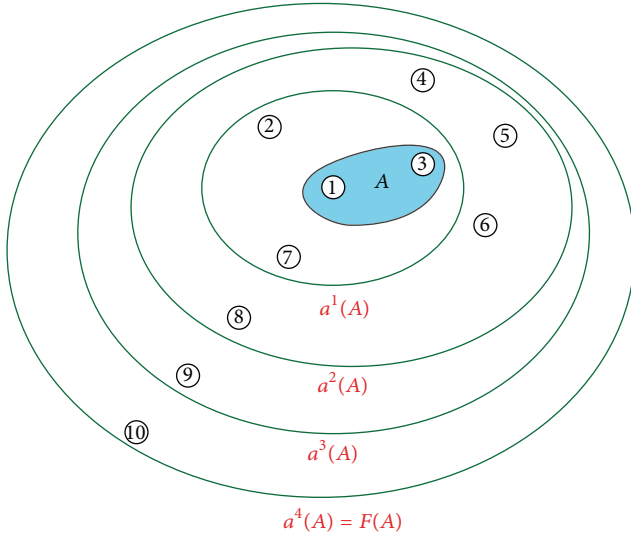


FIGURE 14: Successive pseudoclosures of a set  $A$  for generalized closeness centrality calculation.

**Definition 14** (generalized closeness centrality). We define generalized closeness centrality in a pretopological space  $(E, a)$  denoted by  $C_c^G(A)$ , where  $A \in \mathcal{P}(E)$  such as

$$C_c^G(A) = \sum_k \frac{|a^k(A) - a^{k-1}(A)|}{k} \quad (14)$$

until we reach closure of  $A$  ( $a^{k+1}(A) = a^k(A)$ ) with  $k$  a positive integer.

Closure is defined in Definition 3 and illustrated in Figure 4. The more pseudoclosure steps are needed to reach closure, the less  $A$  is close to the rest of the network. It depends on the number of elements that can be reached too. The distances used in classic closeness centrality (Definition 13) are represented in our generalized definition by the number of pseudoclosures needed to reach an element. The case  $F(A) = A$  means  $A$  is a closed subset, and it cannot reach other elements but itself. In that case,  $C_c^G(A) = 0$ . Here, one can realize the meaning of closure: all elements in the closure of  $A$  can be reached from  $A$ .

**Example 15.** Figure 14 gives an example of the calculation of the generalized closeness centrality. After making the closure of  $A$ , we obtain its “influence” domain, the elements absorbed for each new pseudoclosure being less and less close to  $A$ . We get  $C_c^G(A) = 2 + (4/2) + (1/3) + (1/4) \approx 4.58$ .

By this measure, we can find which elements or groups are the “closest” to others in a complex network.

## 4. Conclusion

In this paper, we presented ideas and concepts which lead to two main contributions.

First, we proposed using pretopology, a general mathematical framework, in order to model groups and relations in social networks. With definitions and examples, we have highlighted the interest to model complex interactions using the pretopology formalism, generalizing graph theory. Furthermore, the introduction of various pseudoclosures functions through examples illustrates our approach.

The second contribution is the generalization of various social network analysis measures, which are degree, betweenness, and closeness centralities. These measures are real group-based measures as we consider a group as a whole entity, not as a combination of individuals.

This paper provides tools to model group interactions in multirelational complex networks. A point not discussed in this paper which can be seen as a perspective is that we can also model dynamic relationships as the neighbourhood functions of a pseudoclosure can evolve through time. Future works can also be axed on some new methods using generalized measures to find groups which maximize a given centrality, to optimize the “efficiency” of a group or to identify emergent groups in a network. Moreover, other centralities and measures could be generalized and based on the same mathematical framework.

## Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

## References

- [1] C. Kadushin, *Understanding Social Networks: Theories, Concepts, and Findings*, Oxford University Press, 2012.
- [2] M. Everett and S. P. Borgatti, “Extending centrality,” in *Models and Methods in Social Network Analysis*, pp. 57–76, Cambridge University Press, Cambridge, UK, 2005.
- [3] S. P. Borgatti and M. G. Everett, “A graph-theoretic perspective on centrality,” *Social Networks*, vol. 28, no. 4, pp. 466–484, 2006.
- [4] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, UK, 2008.
- [5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, “Complex networks: structure and dynamics,” *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [6] M. Newman, A. L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks*, Princeton University Press, 2006.
- [7] A. E. Motter and R. Albert, “Networks in motion,” *Physics Today*, vol. 65, no. 4, pp. 43–48, 2012.
- [8] R. Lambiotte, “Multi-scale modularity and dynamics in complex networks,” in *Dynamics On and Of Complex Networks, Modeling and Simulation in Science, Engineering and Technology*, pp. 125–141, Springer, New York, NY, USA, 2013.
- [9] M. G. Everett and S. P. Borgatti, “The centrality of groups and classes,” *Journal of Mathematical Sociology*, vol. 23, no. 3, pp. 181–201, 1999.
- [10] B. Saha and L. Getoor, “Group proximity measure for recommending groups in online social networks,” in *Proceedings of the 2nd International Workshop on Social Network Mining and Analysis Held in Conjunction with the 14th ACM SIGKDD*

- International Conference on Knowledge Discovery and Data Mining (SNA-KDD '08)*, ACM Press, Las Vegas, Nev, USA, August 2008.
- [11] B. T. Dai, F. C. T. Chua, and E.-P. Lim, "Structural analysis in multi-relational social networks," in *Proceedings of the SIAM International Conference on Data Mining (SDM '12)*, pp. 451–462, April 2012.
  - [12] M. de Domenico, A. Solé-Ribalta, E. Cozzo et al., "Mathematical formulation of multilayer networks," *Physical Review X*, vol. 3, Article ID 041022, 2013.
  - [13] M. Szell, R. Lambiotte, and S. Thurner, "Multirelational organization of large-scale social networks in an online world," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 31, pp. 13636–13641, 2010.
  - [14] V. Levorato, "Modeling groups in social networks," in *Proceedings of the 25th European Conference on Modelling and Simulation (ECMS '11)*, pp. 129–134, Krakow, Poland, June 2011.
  - [15] Z. Belmandt, *Manuel de Prétopologie et Ses Applications*, Hermes Sciences Publications, 1993.
  - [16] Z. T. Belmandt, *Basics of Pretopology*, Hermann, 2011.
  - [17] T. V. Le, N. Kabachi, and M. Lamure, "Pretopology and a homogeneous method for data clustering," in *Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future in Computing & Communication Technologies (RIVF '08)*, IEEE Computer Press, 2008.
  - [18] M. Brissaud, "Les espaces prétopologiques," *Comptendu de l'Académie des Sciences*, vol. 280, pp. 705–708, 1975.
  - [19] M. Fréchet, *Espaces Abstraits*, Hermann, 1928.
  - [20] C. Kuratowski, *Topologie*, Polskie Towarzystwo Matematyczne, Warszawa, Poland, 1952.
  - [21] E. Čech, *Topological Spaces*, John Wiley and Sons, New York, NY, USA, 1966.
  - [22] M. Dalud-Vincent, M. Brissaud, and M. Lamure, "Closed sets and closures in pretopology," *International Journal of Pure and Applied Mathematics*, vol. 50, no. 3, pp. 391–402, 2009.
  - [23] A. Degenne, "Between tool and theory, social networks," *Social Networks of the Internet*, 2004.
  - [24] M. Dalud-Vincent, *Modèle prétopologique pour une méthodologie d'analyse de réseaux: concepts et algorithmes* [Ph.D. thesis], University Claude Bernard Lyon 1, Lyon, France, 1994.
  - [25] S. Sampson, *A novitiate in a period of change. An experimental and case study of social relationships* [Ph.D. thesis], Cornell University, 1968.
  - [26] C. Langeron and S. Bonnevey, "A pretopological approach for structural analysis," *Information Sciences*, vol. 144, no. 1–4, pp. 169–185, 2002.
  - [27] S. BenAmor, S. Bonnevey, M. Bui, and M. Lamure, "A stochastic and pretopological modeling aerial pollution of an urban area," *Revue Santé et Systémique*, vol. 10, no. 3–4, pp. 73–85, 2008.
  - [28] C. Basileu, *Modelling structural of social networks: towards a decision support tool in case of a Public Health Crisis* [Ph.D. thesis], Université Lumière Lyon 2, 2011.
  - [29] V. Levorato and M. Bui, "Modeling the complex dynamics of distributed communities of the web with pretopology," in *Proceedings of the 7th International Workshop on Innovative Internet Community Systems (I2CS '07)*, Munich, Germany, June 2007.
  - [30] K. Musiał, P. Kazienko, and P. Bródka, "User position measures in social networks," in *Proceedings of the 3rd SNA-KDD Workshop on Social Network Mining and Analysis*, ACM Press, June 2009.
  - [31] P. Bonacich, A. C. Holdren, and M. Johnston, "Hyper-edges and multidimensional centrality," *Social Networks*, vol. 26, no. 3, pp. 189–203, 2004.
  - [32] P. Floriano, A. Goldman, and L. Arantes, "Formalization of the necessary and sufficient connectivity conditions to the distributed mutual exclusion problem in dynamic networks," in *Proceedings of the 10th IEEE International Symposium on Network Computing and Applications (NCA '11)*, pp. 203–210, August 2011.
  - [33] M. A. Rodriguez and J. Shinavier, "Exposing multi-relational networks to single-relational network analysis algorithms," *Journal of Informetrics*, vol. 4, no. 1, pp. 29–41, 2010.
  - [34] B. Bollobás, *Modern Graph Theory*, Springer, New York, NY, USA, 1998.
  - [35] U. Brandes and T. Erlebach, *Network Analysis: Methodological Foundations*, Springer, New York, NY, USA, 2005.
  - [36] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978–1979.
  - [37] L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, pp. 35–41, 1977.
  - [38] E. D. Kolaczyk, D. B. Chua, and M. Barthélemy, "Group betweenness and co-betweenness: inter-related notions of coalition centrality," *Social Networks*, vol. 31, no. 3, pp. 190–203, 2009.
  - [39] C. Dangalchev, "Residual closeness in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 365, no. 2, pp. 556–564, 2006.
  - [40] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, pp. 581–603, 1966.
  - [41] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
  - [42] P. Bonacich, "Power and centrality: a family of measures," *American Journal of Sociology*, vol. 92, pp. 1170–1182, 1987.
  - [43] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, vol. 23, no. 3, pp. 191–201, 2001.



