A Proposal for the EI Index for Fuzzy Groups

Ricardo Andrade (ricardo.andrade@jaboatao.ifpe.edu.br)
Universidade Federal de Pernambuco https://orcid.org/0000-0001-9639-9797
Leandro Rêgo
Universidade Federal do Ceara - Campus do Pici

Research Article

Keywords: Social Networks, Homophily, Fuzzy Groups, EI Index

Posted Date: November 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-827211/v1

A Proposal for the *EI* Index for Fuzzy Groups

Ricardo Lopes de Andrade[a,*], Leandro Chaves Rêgo[a,b,c]

[a] *Management Engineering Graduate Program, Federal University of Pernambuco*
[b] *Statistics and Applied Math Department, Universidade Federal do Ceará*
[c] *Statistics Graduate Program, Universidade Federal de Pernambuco*

Abstract

In this paper, it is proposed a measure that quantifies the relational structure within and between groups that comprehend not only the analysis of disjoint groups or non-disjoint groups but also in fuzzy groups. This measure is based on the existing measure known as the *EI* index. The current *EI* index is a measure of homophily applied to networks with the presence of disjoint groups, although disjoint groups on a large scale rarely exist in many empirical networks. The new measure permits the expansion of the analysis of social networks, for several types of attributes, and thus generating previously untapped knowledge. Moreover, it is also proposed combining edges' and nodes' weights in the evaluation of the *EI* index. The new measure is tested in two networks in different contexts. The first one is a co-authorship network, where researchers, actors in the network, are divided according to the time of completion of the doctorate. The second network is formed by trade relations between countries of the American continent, where countries are grouped according to the Human Development Index.

Keywords: Social Networks, Homophily, Fuzzy Groups, *EI* Index

1. Introduction

In general, a social network is a structure formed by nodes (actors) and edges (interactions) used by social and mathematical sciences alike in studies of the relationships between individuals, groups or organizations. Essentially focused on the topological structure, social networks studies apply a set of methods and measures to identify, visualize and analyze social networks looking for patterns of interactions and its implications. (Newman, 2001b,a).

In several networks, it is common to observe that actors tend to have affinities or similarities (attributes) with their peers. According to Crandall et al. (2008) there are two mechanisms reasons for this: actors can modify their behavior to make them more aligned with the behavior of their peers, a process known as social influence (Friedkin, 2006); another distinct reason, an effect termed homophily, is that actors tend to form relationships with others who are already like them. In other words, in the homophily, individual characteristics drive the formation of links, while in social influence, the links existing in the network serve to engage actors' characteristics. Kim and Altmann (2017) mention that the homophily nature is shown in many empirical and theoretical studies.The study of these authors also concluded that homophily affects the network formation. Homophily is the term used for the preference of actors to connect with other actors who share common attributes (McPherson et al., 2001). In studies on homophily, we seek to know if the nodes of a network disproportionately establish links with others that are similar to them in some respect, that is, we want to verify the occurrence of a higher incidence of relations between actors that have similar attributes.

*Corresponding author
Email address: ricardolopesa@gmail.com (Ricardo Lopes de Andrade)

However, actors can belong to many associative groups simultaneously, with various levels of affiliation, and distinct disjoint groups rarely exist on a large scale in many empirical networks (Leskovec et al., 2008). Currently publications that use the *EI* index as a measure of homophily are concentrated in disjoint or mutually exclusive groups. Situations where network actors are present in more than one group are not commonly explored. One of the barriers encountered in the analysis of non-disjoint groups is the absence of a measure, since the *EI* index is defined for disjoint groups (Andrade and Rêgo, 2019). However, Andrade and Rêgo (2019) suggest a method that generalized the *EI* index developed by Krackhardt and Stern (1988) and quantifies the relational structure within and between groups that encompass not only the analysis of disjoint groups but also of non-disjoint groups. Furthermore, we observe that the process of social influence has already been studied in the context of fuzzy groups (Li and Wei, 2019; Khalid and Beg, 2019)

In this context, the objective of this work is to expand the generalized metric suggested by Andrade and Rêgo (2019), adapting it to also encompass groups where the nodes present various levels of affiliations, fuzzy groups. Thereafter, we can analyze, for example, networks that analyze political behavior, studying relationships between voters with different positions in the political spectrum and networks of friendships with bilingual speakers, analyzing the relationships between speakers with different levels of language fluency. In our work, we analyzed two networks. A co-authorship network formed by researchers with a PhD in production engineering, where the time of completion of the doctorate defined the fuzzy groups. The other network is formed by trade relations between American countries, we use the Human Development Index (HDI) to form fuzzy groups.

This paper is organized as follows. In Section 2, we briefly present the *EI* index proposed by Krackhardt and Stern (1988), which measures homophily in networks with disjoint groups. Then, in Section 3, we present our measure, which is a generalization of the current *EI* index, encompassing fuzzy groups. Two applications of the proposed measure are made in Section 4. Finally, we discuss the results of the applications in Section 5 and present conclusions.

2. *EI* Index

The *EI* index, proposed by Krackhardt and Stern (1988), essentially quantifies the relational structure within and between groups (Everett and Borgatti, 2012; Krackhardt, 1994). The *EI* index was implemented in the popular social network analysis package UCINET (1999) as a measure for homophily, which analyzes the tendency of people to connect with others similar to themselves, as well as insertion, i.e., how a node or group of nodes decides to connect to other nodes in a network Hanneman and Riddle (2005).

Homophily is one of the most pervasive and robust trends in human interaction, describing how people tend to look for and interact with others that are more like them - often characterized as “birds of a feather” named by McPherson et al. (2001). As a mechanism of social relations, it can explain the group composition in terms of social identities that range from ethnicity to age (Lazarsfeld et al., 1954). In fact, ethnicity, along with geography and kinship (McPherson et al., 2001), has been demonstrated as one of the main motivating factors behind homophilic practices. Everett and Borgatti (2012) are among the researchers who treat the *EI* index as a measure of homophily and heterophily, where smaller values (internal connections) indicate greater homophily and higher values (external connections) indicate lower homophily or greater heterophily. The *EI* index as a measure of homophily is essentially quantifying individuals’ propensity to interact with similar actors (Burt, 1991; McPherson et al., 2001). Furthermore, the *EI* can be used as a segregation measures (Sweet and Zheng, 2017). Segregation is defined as the “unequal” distribution of two or more groups of people in different units or social positions (Bojanowski and Corten, 2014).

Defined as the difference between the intergroup and intragroup ties, divided by the total number of ties for normalization. The *EI* index is a simply and attractive measure of homophily because it does not depend on the density of a network (Everett and Borgatti, 2012). Formaly, the EI index is given by

$$EI \text{ index} = \frac{EL - IL}{EL + IL}, \tag{1}$$

where EL is the number of external links (links between nodes belonging to different groups); IL is the number of internal links (links between nodes belonging to the same group). The EI index ranges from -1 (all bonds are internal) to +1 (all bonds are external). The index can be calculated for the entire network, for each group or for each individual actor.

Although commonly used in an unweighted network, some authors like De Andrade and Rêgo (2018) and Danchev and Porter (2016) also used the EI index in weighted networks. In weighted networks, the EI index is calculated using the weight of the edges, this way EL is the sum of the edge weights that connect different cells of the partition and IL is the sum of the edge weights that connect actors of the same cell of the partition. As with the unweighted network, the EI index for weighted networks assumes values between -1 and +1. Generally the edges' weight represents the frequency or strength of the relationship. Therefore, when the EI index value approaches -1, it means that the internal relations are stronger or more intense. As the index approaches +1, it shows that external relations are stronger or more intense.

In recent years, the inclusion of numerical attributes has been observed in the analysis of social networks. The attributes are resources of the nodes and are used to give weight to them, representing their importance or contribution in the network (De Andrade and Rêgo, 2018; Liu et al., 2015; Benyahia and Largeron, 2015). In this work, we will also consider the nodes' weights and insert it in the topological structure of the network. For this, we use the method proposed by De Andrade and Rêgo (2018). By this method, the weight of the edge is equal to the frequency or strength of the relationship between two nodes multiplied by the average nodes' weights. The intuition is that in cases where information about quantitative features of nodes are available, a link's weight must not only depend on the strength of the connection (original edge weight) but also on the average importance of the nodes being connected. Formally, if v_i is the weight of node i and w_{ij} is the original weight of the link between nodes i and j, then by including the nodes' weights, the new edges' weights become

$$z_{ij} = w_{ij}\frac{v_i + v_j}{2}.$$

The inclusion of the nodes' weights contributes to a more efficient analysis of the network by combining factors inherent to the network with external factors (De Andrade and Rêgo, 2018). As external factors attribute a certain "status" to individuals in the network and through the EI index it is possible to verify whether this status also influences the formation of relationships. However, this conclusion is only obtained by comparing it with the EI index without considering external factors.

3. EI Index: Fuzzy case

Every day, when describing certain phenomena (characteristics), we use degrees that represent qualities or partial truths. As an example, we will define the group of elderly people. A suggestion to formalize this set mathematically could have at least two approaches. The first, distinguishing from which age the individual is considered elderly. In this case, the set is well-defined. The second, less conventional, is given in such a way that individuals are considered elderly with more or less intensity, that is, there are elements that would belong more to the elderly class than others. This means that the younger the individual, the lower his or her degree of belonging to this class. Thus, we can say that individuals belong to the class of elderly people, with more or less intensity. Mathematically, we call fuzzy sets, the sets to which the elements have degrees of pertinence. The formalization of fuzzy sets was presented by Zadeh (1996) as an extension of the classic notion of sets.

In order to explore cases of fuzzy groups, we have developed a new metric to obtain the EI index, which is an adaptation of the metric proposed by Andrade and Rêgo (2019) to generalize the original EI index measure for use with overlapping groups.

Let $\mathscr{A}$ be the set of all attributes for nodes in a social network with n nodes. For $X \in \mathscr{A}$, let $\mu_X(v_i)$ be the level of membership of node v_i to a given group, $0 \leq \mu_X(v_i) \leq 1$. Moreover, for a generic set of nodes, S, consider the following sets of indices $\mathscr{I}(S) = \{i : v_i \in S\}$ and $\mathscr{J}^i(S) = \{j : (v_j \in S \text{ and } j > i) \text{ or } (v_j \notin S \text{ and } j \neq i)\}$. Thus, the number of external and internal links for a generic set of nodes, S, are given, respectively, by:

$$EL(S) = \sum_{i \in \mathscr{I}(S)} \sum_{j \in \mathscr{J}^i(S)} x_{ij}(1 - \max_{X \in \mathscr{A}}\{\mu_X(v_i)\mu_X(v_j)\})$$

and

$$IL(S) = \sum_{i \in \mathscr{I}(S)} \sum_{j \in \mathscr{J}^i(S)} x_{ij} \max_{X \in \mathscr{A}}\{\mu_X(v_i)\mu_X(v_j)\},$$

where in the unweighted case x_{ij} is either 1 or 0 depending on whether there is a link between node v_i and v_j, in the case with only edges' weights $x_{ij} = w_{ij}$ and in the case of edges' and nodes' weights $x_{ij} = z_{ij}$.

Alternatively, for $X \in \mathscr{A}$, we can define the number of external and internal links for the group of nodes, S_X, that has attribute X, respectively, as follows:

$$EL(S_X) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}\mu_X(v_i)(1 - \mu_X(v_j))$$

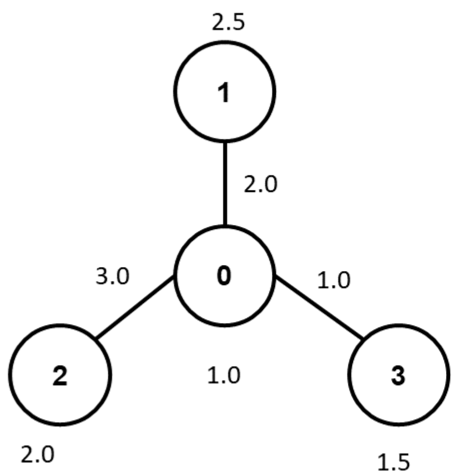
and

$$IL(S_X) = \sum_{i=1}^{n} \sum_{j>i} x_{ij}\mu_X(v_i)\mu_X(v_j),$$

where x_{ij} is defined exactly as before.

To better explain our proposed method, here we present a simple example to expound how the new metric works in a specific network. Suppose there is a network with four nodes that belong with different membership levels to two groups, A and B (show in Figure 1). In the network, $set\{1,2\}$ is considered. Note that node 1 and 2 have no connections and that 0 is connected to both. Disregarding the edges' and nodes' weights, we have $x_{10} = 1$ or $x_{01} = 1$, $\max_{X\in\{A,B\}}\{\mu_X(0)\mu_X(1), \mu_X(0)\mu_X(1)\} = \max_{X\in\{A,B\}}\{\ 0.65 * 0.80, 0.75 * 0.60\} = 0.52$ and $x_{20} = 1$ or $x_{02} = 1$, $\max_{X\in\{A,B\}}\{\mu_X(0)\mu_X(2), \mu_X(0)\mu_X(2)\} = \max_{X\in\{A,B\}}\{\ 0.65*0.65, 0.75*0.15\} = 0.4225$, so $EL(set\{1,2\}) = (1-0.52)+(1-0.4225) = 1.06$ and $IL(set\{1,2\}) = 0.52+0.4225 = 0.94$ and therefore, $EI(set\{1,2\}) = \frac{1.06-0.94}{1.06+0.94} = 0.06$. Now consider group A, we have the following edges x_{01} or x_{10}, x_{02} or x_{20} and x_{03} or x_{30}. $EL(groupA) = x_{01}\mu_X(0)(1-\mu_X(1)) + x_{10}\mu_X(1)(1-\mu_X(0)) + x_{02}\mu_X(0)(1-\mu_X(2)) + x_{20}\mu_X(2)(1-\mu_X(0)) + x_{03}\mu_X(0)(1-\mu_X(3)) + x_{30}\mu_X(3)(1-\mu_X(0)) = 0.65(1-0.80)+0.80(1-0.65)+0.65(1-0.65)+0.65(1-0.65)+0.65(1-0.25)+0.25(1-0.65) = 0.13+0.28+0.2275+0.2275+0.4875+0.0875 = 1.44$ and $IL(groupB) = x_{01}\mu_X(0)\mu_X(1)+x_{02}\mu_X(0)\mu_X(2)+x_{03}\mu_X(0)\mu_X(3) = 0.65*0.80+0.65*0.65+0.65*0.25 = 1.105$, so $EI(groupA) = \frac{1.44-1.104}{1.44+1.104} = 0.13$.

Table 1 displays the results for the graph shown in Figure 1. It is easy to verify that the proposed metric is a generalization of the EI index proposed in Krackhardt and Stern (1988) in the sense that if groups are disjoint and membership functions are either 0 or 1, then it coincides with (1).



	A	B
0	0.65	0.75
1	0.80	0.60
2	0.65	0.15
3	0.25	0.45

Figure 1: Social network with fuzzy groups of nodes.

Table 1: *EI* Index fuzzy groups example

	Unweighted			Weighted		
	EL	*IL*	*EI* Index	*EL*	*IL*	*EI* Index
set {0}	1.72	1.28	0.15	3.36	2.65	0.12
set {1}	0.48	0.52	-0.04	0.96	1.04	-0.04
set {2}	0.58	0.42	0.16	1.73	1.27	0.16
set {3}	0.66	0.34	0.32	0.66	0.34	0.33
set {C}[a]	1.72	1.28	0.15	3.36	2.65	0.12
set {1,2}	1.06	0.94	0.06	2.69	2.31	0.08
set {1,3}	1.14	0.86	0.14	1.62	1.38	0.08
set {2,3}	1.24	0.76	0.24	2.40	1.61	0.20
Group A	1.44	1.10	0.13	2.76	2.47	0.06
Group B	1.65	0.90	0.29	3.45	1.58	0.37
	Z_unweighted			Z_weighted		
	EL	*IL*	*EI* Index	*EL*	*IL*	*EI* Index
set {0}	2.53	1.97	0.13	5.11	4.14	0.10
set {1}	0.84	091	-0.04	1.68	1.82	-0.04
set {2}	0.87	0.63	0.16	2.60	1.90	0.16
set {3}	0.83	0.42	0.33	0.83	0.42	0.33
set {C}	2.53	1.97	0.13	5.11	4.14	0.10
set {1,2}	1.71	1.54	0.05	4.28	3.72	0.07
set {1,3}	1.67	1.33	0.11	2.51	2.24	0.06
set {2,3}	1.69	1.06	0.23	3.43	2.32	0.19
Group A	2.12	1.75	0.10	4.20	3.92	0.03
Group B	2.33	1.38	0.26	5.14	2.50	0.34

[a] C is any set containing node 0 or the set {1,2,3}.

4. Application

In this section, we apply the method proposed in two networks object of studies in previous publications. These networks present the fundamental element for our approach, which is the presence of fuzzy groups, in addition to information about the nodes' weights. As a means of comparison, we also analyze the cases of disjoint (Everett and Borgatti, 2012) and non-disjoint (Andrade and Rêgo, 2019) groups. In this way, the *EI* Index will be obtained for 4 situations: without considering the weight of the edges and the nodes, unweighted (UW); regarding only the nodes' weight, Z_unweighted (ZU); considering only the edges' weight, weighted (W); taking into account both weights, Z_weighted (ZW).

To evaluate whether the *EI* Index for a given group is compatible with the expected when connections occur randomly, i.e., without preference of members for external or internal relations, for the unweighted and the Z_unweighted cases, we calculate the expected *EI* index for each one of the cases analyzed considering the average of 5000 randomly generated binomial graphs with the same density and size as that of the original graphs. We also added a probability, p-value, that expresses how unlikely it is to obtain an *EI* index at least as extreme as the one observed in the randomly generated binomial graphs. We considered one sided p-values so that it is given by the relative frequency of times that the simulated *EI* has a value greater (resp., less) than or equal to the observed *EI*, when the expected *EI* is less (resp., greater) than the observed one.

4.1. Data

In order to implement the proposed *EI* index, we use data from two real networks. Next, we give some details about these networks.

(i) Co-authorship PQ: The PQ network is a co-authorship network among researchers in the area of Production Engineering in Brazil that held a Productivity Research scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) in 2015. It has 124 nodes and 131 edges. The network is undirected and the edges' weights represent the number of publications made in co-authorship by a given pair of researchers in the period of 2005 to 2014. (Andrade and Rêgo, 2017).

(ii) Trade of American Countries: The network of commerce between American countries is formed by 30 countries and 356 edges. This network was created from the network of international trade developed by (De Andrade and Rêgo, 2018) which includes 178 countries from all continents, forming a unique main component with 10,419 edges. The network is undirected and the edges' weights represent the mean of export and import commercial transactions between a pair of countries during 2015.

4.2. PQ Network

We first show how the arbitrary choice of disjoint groups, according to the doctoral completion time, affects the *EI* index of these groups. We delimit three cases of the disjoint groups (T1, T2 and T3) by varying the limits of the groups, Table 2, in the fuzzy regions, Table 3. Figure 2 shows the *EI* index for the entire network, for each of the arbitrary limits. It is clear that the result is strongly dependent on these limits.

Table 2: Criteria for defining disjoint groups in the PQ Network

Case	PhD time	criterion	group size
T1	Young	≤ 5	2
	experient	≥ 6 and ≤ 24	100
	Senior	≥ 25	22
T2	Young	≤ 6	9
	experient	≥ 7 and ≤ 23	89
	Senior	≥ 24	26
T3	Young	≤ 7	12
	experient	≥ 8 and ≤ 20	73
	Senior	≥ 21	39

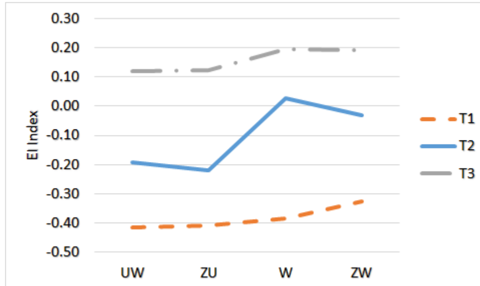


Figure 2: *EI* indexes for the whole PQ Network

The definitions of the groups formed according to the doctoral completion time for the disjoint, non-disjoint and fuzzy case, followed the criteria in Table 3. For the disjoint case, we considered the intermediary case T2.

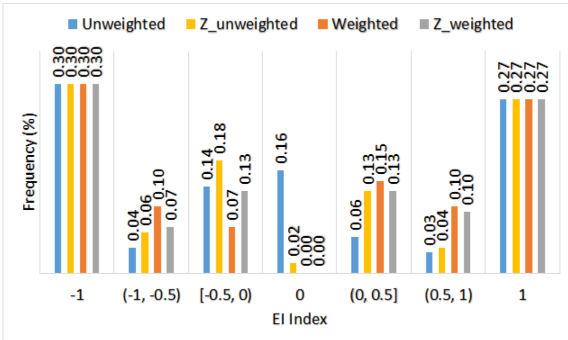
Table 3: Criteria for defining groups in the PQ Network

Case	PhD time	criterion	group size
Disjoint	Young	≤ 6	9
	experient	≥ 7 and ≤ 23	89
	Senior	≥ 24	26
Non-disjoint	Young	≤ 7	12
	experient	≥ 5 and ≤ 25	104
	Senior	≥ 20	45
Fuzzy	Young	$\mu(x) = \begin{cases} 1, & x < 5 \\ \frac{7-x}{2}, & 5 \leq x < 7 \end{cases}$	7 2
	experient	$\mu(x) = \begin{cases} \frac{x-5}{2}, & 5 < x \leq 7 \\ 1, & 7 < x < 20 \\ \frac{25-x}{5}, & 20 \leq x < 25 \end{cases}$	10 67 23
	Senior	$\mu(x) = \begin{cases} \frac{x-20}{5}, & 20 < x \leq 25 \\ 1, & x > 25 \end{cases}$	21 18

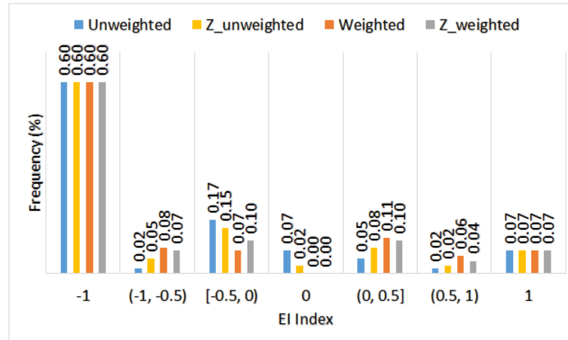
We use the researchers' h-index as the nodes' weights. The h-index is a measure that combines, in a simple way, the quantity of publications and the impact of publications and is given by the maximum value of h such that a researcher has published h works and each one of these works have been cited h or more times Hirsch (2010).

Figure 3 shows how the relationships between researchers occur. In general, most nodes in the non-disjoint case have an *EI* index of -1 (60%). In the fuzzy case and the disjoint case, the nodes have a similarity in relation to the proportion of *EI* index higher and lower than zero, however, in the fuzzy case the distribution of the *EI* index is more uniform.

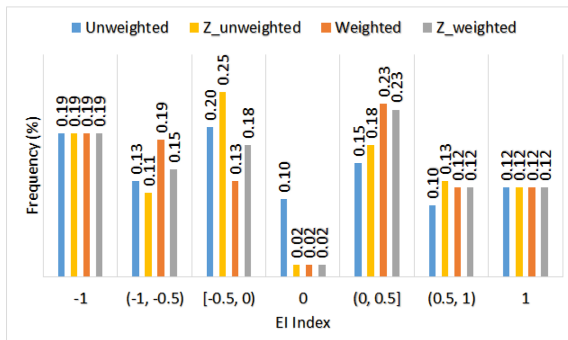
Figure 4 shows the *EI* index for the entire network. In general, when the nodes belong to non-disjoint groups, it is observed that the *EI* indexes are smaller, with a predominance of in-group relationships. On the other hand, when the groups are disjoint, the network has higher, but still negative, *EI* indexes. This means that, on a global level, there is a high level of cooperation between researchers from the same time group. As for the strength of the connections, it is observed that in the W network there is the lowest *EI* indexes and the ZW network the highest *EI* indexes. The first result indicates that the relationships are stronger between researchers in the same group and the second indicates that researchers who connect to researchers in other groups tend to link to researchers with higher h-indexes. It is worth mentioning that the negative *EI* indexes of the network, revealing a predominant in-group relationship, does not differ significantly from the result obtained in a random simulated network since the p-values are all greater than 0.05.



(a)



(b)



(c)

Figure 3: Distribution of the *EI* index values in the PQ Network. (a) disjoint, (b) non-disjoint and (c) fuzzy.

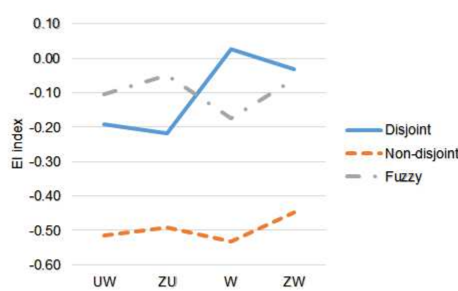


Figure 4: *EI* indexes for the whole PQ Network

The analysis of the *EI* index of the experience level groups is shown in Figure 5. In general, when nodes belong to non-disjoint groups, it is observed that the *EI* indexes are smaller. In the case of disjoint and fuzzy groups, the *EI* indexes are close, with the *EI* indexes of the disjoint case slightly higher. The *EI* indexes of the experient group are negative, specially in the non-disjoint case. This shows that the internal connections of this group are higher than the external ones. The youth and senior groups have a positive EI index, with the youth being superior to the senior. This shows that the external relations of these outweigh the internal ones. Therefore, we can conclude that the experts cooperate with each other while young and senior PhDs are more open to cooperating with other groups. It is worth mentioning that the *EI* indexes obtained do not reveal a tendency towards homophily or heterophily, as they do not differ significantly from the results obtained by the random simulated network since the p-values are all greater than 0.05. Note that the weighting of the edges affected the *EI* index of the disjoint case more, making the relationships more heterogeneous. This is more noticeable in the case of experient groups.

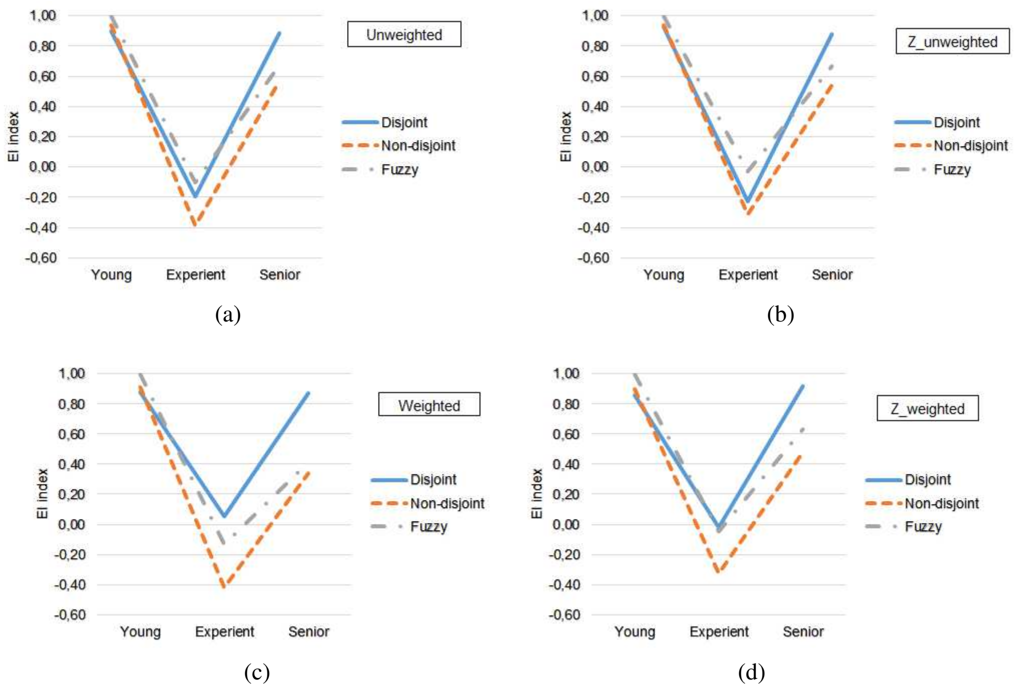


Figure 5: *EI* indexes for experience level areas in the PQ Network.

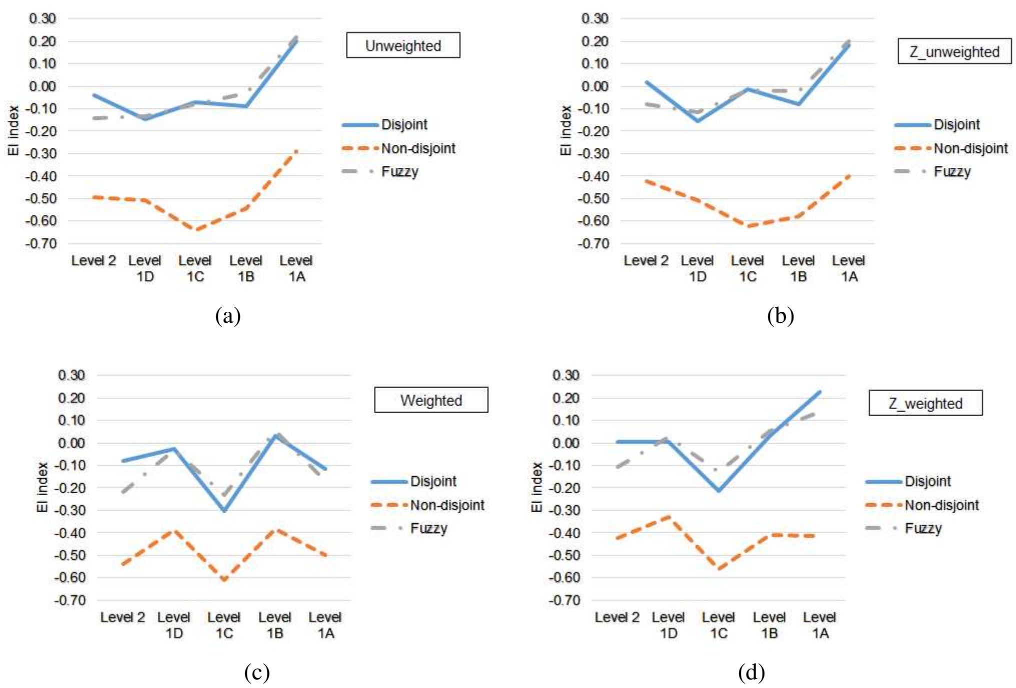


Figure 6: *EI* indexes for scholarship level groups considering the experience level as class attributes in the PQ Network.

We also analyzed the behavior of groups of researchers with the same level of scholarship in respect to the experience level group attributes. The scholarship level in order of importance and the total number of researchers are: 1A (8 %), 1B (5 %), 1C (8 %), 1D (19 %) and 2 (59 %). The analyses of the EI index of these groups are shown in Figure 6 for the cases of disjoint, non-disjoint and fuzzy groups, and studying the UW, ZU, W and ZW networks. In general, when the nodes belong to non-disjoint groups, it is observed that the *EI* indexes are smaller, predominance of in-group relationships. On the other hand, when the groups are disjoint or fuzzy, the network has higher *EI* indexes.

As for scholarship levels, there is a different behavior of the *EI* indexes for the different connection types, weighted or unweighted. Level 1A presents the highest *EI* indexes in the unweighted network, without or with the inclusion of the nodes' weights and in the weighted network considering the nodes' weights. Level 1A, the highest level of the scholarship, concentrates the most productive and influential researchers in the research area, being composed of 10 exclusively senior researchers and 2 exclusively experient. Although most are seniors, the in-group relationship is predominant in the non-disjointed case and external relationships are more common when the group is fuzzy or disjoint. Level 1A *EI* indexes are all negative in the weighted network. Level 1C, an intermediate scholarship level, also does not comprise young researchers. In the weighted network, without and with nodes' weights, as well as in the unweighted network (only in the non-disjoint case), the *EI* index of the level 1C is the smallest and negative. Therefore, for researchers at this level, most connections occur between researchers of the same experience level group. It is noteworthy that the *EI* indexes obtained do not reveal a tendency towards homophily or heterophily, as they do not differ significantly from the results obtained by random simultaed networks since the p-values are all greater than 0.05.

4.3. Trade of American Countries Network

We use the Human Development Index (HDI) to form groups and first show how the arbitrary choice of disjoint groups, according to the HDI, affects the *EI* index of these groups. We delimit three cases of the disjoint groups (T1, T2 and T3) by varying the limits of the groups, Table 4, in the fuzzy regions, Table 5. Figure 7 shows the *EI* index for the entire network, for each of the arbitrary limits. It is clear that the result is strongly dependent on these limits.

Table 4: Criteria for defining groups in the Trade network

Case	PhD time	criterion	group size
T1	Low	≤ 0.5	2
	Medium	> 0.5 and ≥ 0.675	5
	High	> 0.675 and ≥ 0.775	15
	Very high	> 0.775	8
T2	Low	≤ 0.5625	2
	Medium	> 0.5625 and ≥ 0.7	6
	High	> 0.7 and ≥ 0.8	18
	Very high	> 0.5	4
T3	Low	≤ 0.6	2
	Medium	> 0.6 and ≥ 0.725	12
	High	> 0.725 and ≥ 0.825	12
	Very high	> 0.825	4

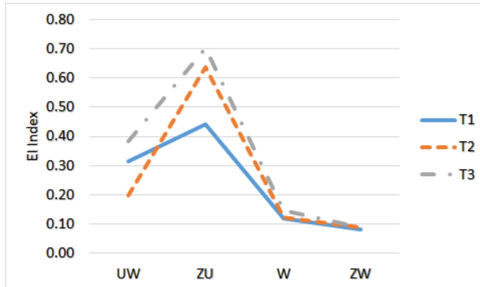


Figure 7: *EI* indexes for the whole Trade Network

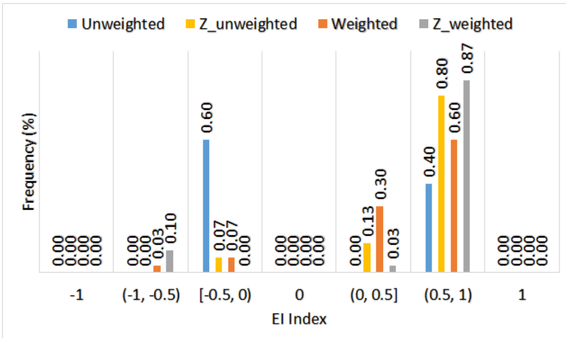
The definitions of the groups formed according to the HDI for the disjoint, non-disjoint and fuzzy case, followed the criteria in the Table 5, where the intermediary case T2 was considered for the disjoint case.

Table 5: Criteria for defining groups in the Trade Network

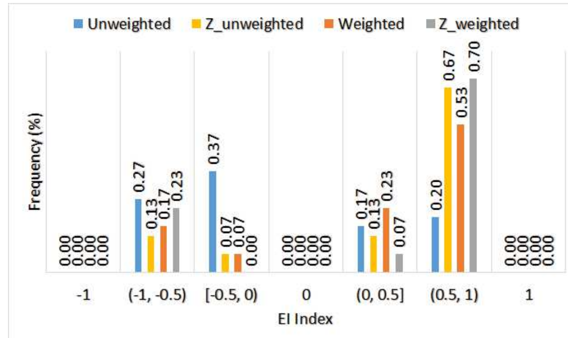
Case	HDI	criterion	group size
Disjoint	Low	≤ 0.5625	2
	Medium	> 0.5625 and ≤ 0.7	6
	High	> 0.7 and ≤ 0.8	18
	Very high	> 0.8	4
Non-disjoint	Low	≤ 0.6	2
	Medium	≥ 0.525 and ≤ 0.725	12
	High	≥ 0.675 and ≤ 0.825	19
	Very high	≥ 0.775	8
Fuzzy	Low	$\mu(x) = \begin{cases} 1, & x < 0.4 \\ \frac{0.6-x}{0.6-0.4}, & 0.4 \leq x < 0.6 \end{cases}$	0 2
	Medium	$\mu(x) = \begin{cases} \frac{x-0.525}{0.6-0.525}, & 0.525 < x \leq 0.6 \\ 1, & 0.6 < x < 0.675 \\ \frac{0.675-x}{0.675-0.725}, & 0.675 < x \leq 0.725 \end{cases}$	0 5 7
	High	$\mu(x) = \begin{cases} \frac{x-0.675}{0.6-0.725}, & 0.625 < x \leq 0.725 \\ 1, & 0.725 < x < 0.775 \\ \frac{0.825-x}{0.825-0.775}, & 0.775 \leq x < 0.825 \end{cases}$	11 8 4
	Very high	$\mu(x) = \begin{cases} \frac{x-0.775}{0.9-0.775}, & 0.775 < x \leq 0.9 \\ 1, & x > 0.9 \end{cases}$	6 2

Figure 8 shows the *EI* index at the individual level of the 30 countries. In general, countries have positive *EI* indexes, that is, intergroup relations superior to in-groups ones. In the non-disjoint case, it is possible to notice that some countries predominate in-group relations. The in-group relationship is also more visible when the network is unweighted.

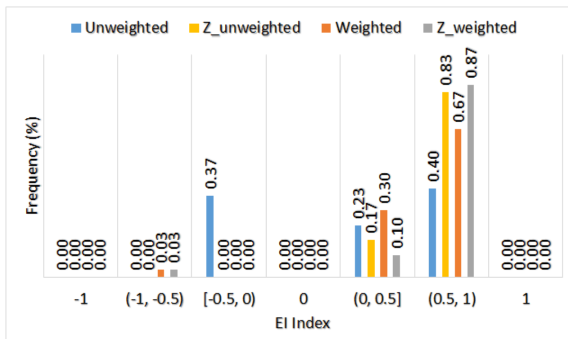
Figure 9 shows the *EI* index for the entire network. In general, when nodes belong to non-disjoint groups, it is observed that the *EI* indexes are smaller. On the other hand, when the groups are fuzzy, the network has higher *EI* indexes. The *EI* indexes are positive, except for the *EI* index in the case of non-disjoint groups in the weighted network. This indicates that, on a global level, trade occurs between countries of different HDI groups. The predominant intergroup relationships do not differ significantly from the result obtained by random simulated networks since the p-values are greater than 0.05.



(a)



(b)



(c)

Figure 8: Distribution of *EI* index values in Trade Network. (a) disjoint, (b) non-disjoint and (c) fuzzy.

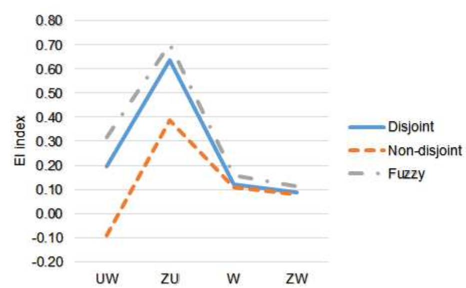


Figure 9: *EI* indexes for the whole Trade Network.

The analysis of the *EI* index of the HDI groups is shown in Figure 10. In general, the low and medium groups have the highest *EI* indexes, close to 1, countries in these groups have intergroup relations higher than in-groups, the *EI* indexes are statistically significant, that is, these groups are prone to heterophily. The group with high HDI has the lowest *EI* indexes in the unweighted network, being the one with the highest in-group relationship, but the *EI* indexes increases significantly in the Z_Unweighted, weighted and Z_Weighted network, that is, the relationships are stronger with other groups. The group of countries with a very high HDI has the lowest *EI* indexes in the weighted network, with and without the nodes' weights, revealing a stricter relationship between countries in the group. The *EI* indexes of the groups with high and very high HDI do not differ statistically from that presented by the random network.

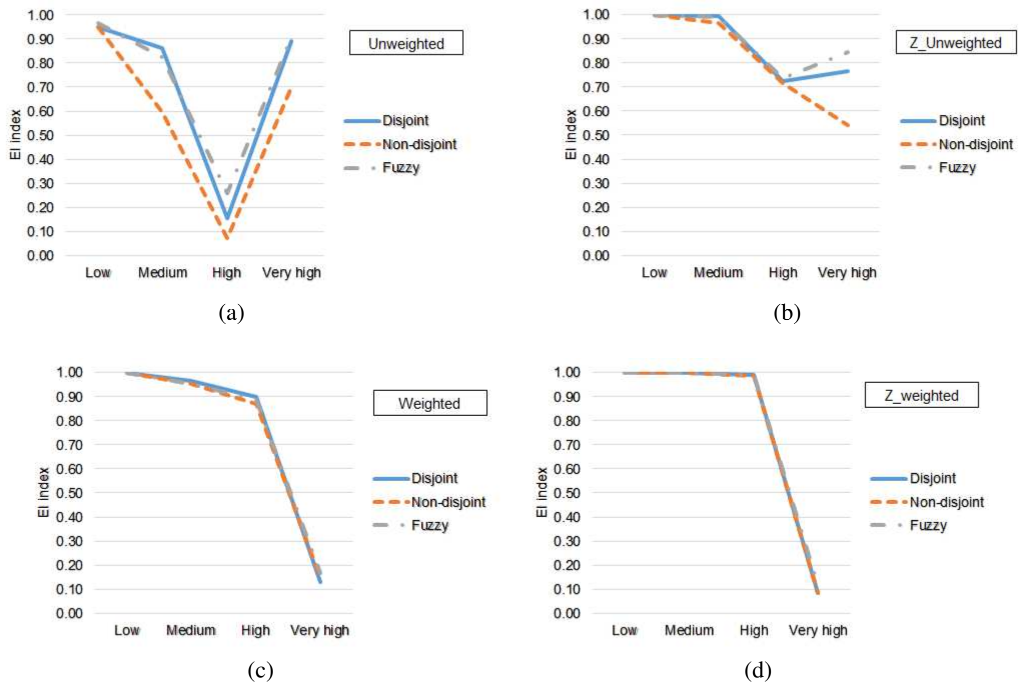


Figure 10: *EI* indexes for HDI groups in the Trade Network

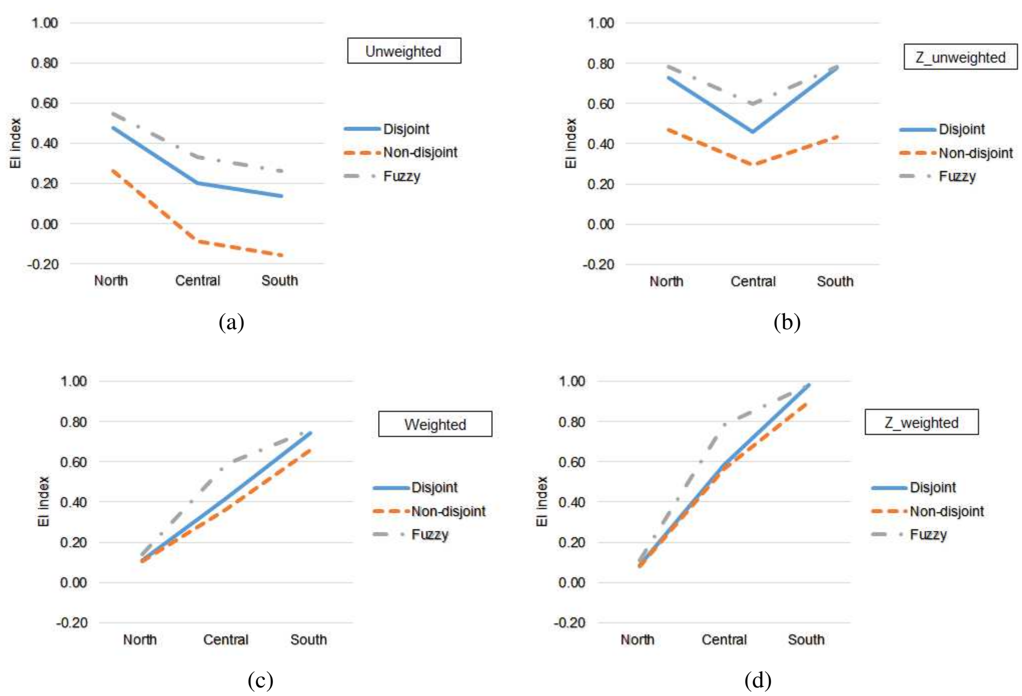


Figure 11: *EI* indexes for regional groups considering the HDI groups as class attributes in the Trade Network

We also analyzed the behavior of groups of countries by region in respect to the HDI group attributes. The regional divisions are north, south and central, with 3, 12 and 15 countries, respectively. The analyses of the *EI* indexes of these groups are shown in Figure 11 for the cases of disjoint, non-disjoint and fuzzy groups, and studying the UW, ZU, W and ZW networks. In general, when the nodes belong to non-disjoint groups, it is observed that the *EI* indexes are smaller. On the other hand, when the groups are disjoint or fuzzy, the regions have higher *EI* indexes. As for the regions, there is a behavior different from the *EI* index depending on the connection type, weighted or unweighted. The northern region has the highest *EI* indexes on the UW and ZU networks. The northern region's *EI* indexes fall in the weighted network, indicating that countries in the northern region have stronger relations with countries in the same HDI group. The southern region in the UW network has the lowest *EI* indexes, positive in the disjoint and fuzzy case, and negative in the non-disjoint case. In weighted networks, with and without nodes' weights, the *EI* indexes are positive and higher in the southern region, indicating that the forces of relations are more intense between countries of different HDI groups. The *EI* indexes of the regions do not reveal a tendency towards homophily or heterophily, as they do not differ significantly from the *EI* presented in the simulated network with random relationships.

5. Conclusions

In this work, we have proposed a new network measure which is a generalization of the *EI* index to measure homophily in fuzzy group cases. Given that actors can belong to many associative groups simultaneously, with various levels of affiliation, therefore, to better understand the structure of the networks, the measure developed allows the analysis of multiple associations and varying levels of association. We also showed that incorporating nodes' weights in the analysis can give us more insights about the homophily of relations.

We explored two networks with the new measure. In a co-authorship network, the doctoral completion time was used to form groups. In a commercial network among countries, we use the Human Development Index (HDI) to form groups. We obtain the *EI* index for the networks considering the cases of disjoint, non-disjoint and fuzzy groups, and analyzing different relational forces, unweighted, weighted, without and with the nodes' weights. As we have seen in these networks, the proposed measure allows for an expansion of the analysis of social networks. Through a homophily analysis, it is possible to identify whether a certain group of nodes has a tendency to work together or not.

In general, it is clear that fuzzy groups generate more homogeneous cooperation or commercial relations. This was already expected due to the fact that the actors present multiple associations with the same degree of association, equal to 1. In the co-authorship network, we noticed that the researchers allocated as experient are the ones that most cooperate with each other. These relationships are favored because there are more experient researchers. The smaller number of young and senior researchers also justifies the predominance of external relations by these researchers. In the trade network, we realize that relations between countries with different levels of development are more common. In the case of the groups with low and medium HDI, we noticed that the *EI* index close to 1 is statistically significant, revealing the tendency towards heterophily in these two groups, revealing their dependency on more developed countries.

Acknowledgments

The work of R. L. de Andrade was also supported by the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), IBPG-1182-3.08/18. The work of L. C. Rêgo was also supported by the Conselho Nacional de Desenvolvimento Científico e Tecnolóogico (CNPq) under grants 307556/2017-4 and 428325/2018-1. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

References

Andrade, R.L.d., Rêgo, L.C., 2017. Exploring the co-authorship network among cnpq's productivity fellows in the area of industrial engineering. Pesquisa Operacional 37, 277–310.

Andrade, R.L.d., Rêgo, L.C., 2019. Proposal for the E-I index for non-disjoint groups. The 8th International Conference on Complex Networks and Their Applications. Complex Networks .

Benyahia, O., Largeron, C., 2015. Centrality for graphs with numerical attributes, in: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, pp. 1348–1353.

Bojanowski, M., Corten, R., 2014. Measuring segregation in social networks. Social Networks 39, 14–32.

Burt, R.S., 1991. Measuring age as a structural concept. Social Networks 13, 1–34.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S., 2008. Feedback effects between similarity and social influence in online communities, in: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 160–168.

Danchev, V., Porter, M.A., 2016. Heterogeneity of global and local connectivity in spatial network structures of world migration. Available at SSRN 2755271 .

De Andrade, R.L., Rêgo, L.C., 2018. The use of nodes attributes in social network analysis with an application to an international trade network. Physica A: Statistical Mechanics and its Applications 491, 249–270.

Everett, M.G., Borgatti, S.P., 2012. Categorical attribute based centrality: E-I and G-F centrality. Social Networks 34, 562–569.

Friedkin, N.E., 2006. A structural theory of social influence. volume 13. Cambridge University Press.

Hanneman, R.A., Riddle, M., 2005. Introduction to social network methods.

Hirsch, J.E., 2010. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. Scientometrics 85, 741–754.

Khalid, A., Beg, I., 2019. Soft pedal and influence-based decision modelling. International Journal of Fuzzy Systems 21, 620–629.

Kim, K., Altmann, J., 2017. Effect of homophily on network formation. Communications in Nonlinear Science and Numerical Simulation 44, 482–494.

Krackhardt, D., 1994. Graph theoretical dimensions of informal organizations. Computational organization theory 89, 123–140.

Krackhardt, D., Stern, R.N., 1988. Informal networks and organizational crises: An experimental simulation. Social psychology quarterly , 123–140.

Lazarsfeld, P.F., Merton, R.K., et al., 1954. Friendship as a social process: A substantive and methodological analysis. Freedom and control in modern society 18, 18–66.

Leskovec, J., Lang, K.J., Dasgupta, A., Mahoney, M.W., 2008. Statistical properties of community structure in large social and information networks, in: Proceedings of the 17th international conference on World Wide Web, pp. 695–704.

Li, S., Wei, C., 2019. Modeling the social influence in consensus reaching process with interval fuzzy preference relations. International Journal of Fuzzy Systems 21, 1755–1770.

Liu, J., Li, Y., Ruan, Z., Fu, G., Chen, X., Sadiq, R., Deng, Y., 2015. A new method to construct co-author networks. Physica A: Statistical Mechanics and its Applications 419, 29–39.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 27, 415–444.

Newman, M.E., 2001a. Scientific collaboration networks. i. network construction and fundamental results. Physical review E 64, 016131.

Newman, M.E., 2001b. The structure of scientific collaboration networks. Proceedings of the national academy of sciences 98, 404–409.

Sweet, T.M., Zheng, Q., 2017. A mixed membership model-based measure for subgroup integration in social networks. Social Networks 48, 169–180.

UCINET, U., 1999. Ucinet 6 for windows software for social network analysis .

Zadeh, L.A., 1996. Fuzzy sets, in: Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh. World Scientific, pp. 394–432.