Social Network Measures for "Nosduocentered" Networks, their Predictive Power on Performance*.

Lluís Coromina, Jaume Guia and Germà Coenders†

Departament d'Economia,
Universitat de Girona

Girona, February 2005

Abstract

Our purpose in this article is to define a network structure which is based on two egos instead of the egocentered (one ego) or the complete network (n egos). We describe the characteristics and properties for this kind of network which we call "nosduocentered network", comparing it with complete and egocentered networks. The key point for this kind of network is that relations exist between the two main egos and all alters, but relations among others are not observed. After that, we use new social network measures adapted to the nosduocentered network, some of which are based on measures for complete networks such as degree, betweenness, closeness centrality or density, while some others are tailor-made for nosduocentered networks.

We specify three regression models to predict research performance of PhD students based on these social network measures for different networks such as advice, collaboration, emotional support and trust. Data used are from Slovenian PhD students and their supervisors.

The results show that performance for PhD students depends mostly of the emotional network, because it is significant for all three models. Trust and collaboration networks are significant for two models and advice is not significant for any model.

As regards network measures, classic and tailor-made measures are about equally good. Measures related to the total intensity of contacts (e.g., density, degree centralization and size) seem to work best to predict performance.

Keywords: nosduocentered network, academic achievement, performance, network measures

JEL classification: M12, A23, C69, I29

---

* This work was partly supported by the University of Girona grants GRHCS66, BR00/UdG and 4E200304 and is a partial result of a wider project carried out at the INSOC (International Network on Social Capital and Performance http://srcvserv.ugent.be/insoc/insoc.htm). Acknowledgements are due to all INSOC members who contributed to the proposal, the questionnaire design and the data collection: Anuška Ferligoj, Tina Kogovšek, Hans Waege, Daniëlle de Lange, Filip Agneessens, Uroš Matelič, Dagmar Krebs, Jurgen Hoffmeyer-Zlotnik, Brendan Bunting, Valentina Hlebec, Bettina Langfeldt and Joanne Innes. Many of them read earlier version of this article and made crucial suggestions.

† Address: Departament d'Economia. Universitat de Girona, Campus de Montilivi, 17071 Girona, Spain. E-mail: lluis.coromina@udg.es

1. Introduction

The aim of this paper is threefold. Firstly we explain a network whose structure is defined somewhere between the complete and egocentered networks. We call it nosduocentered network. Secondly we define social network measures for this network based on Freeman's (1979) complete networks measures (centrality degree, closeness, etc.) and some tailor-made measures; after that, we apply these measures in different networks such as scientific advice, collaboration, emotional support and trust. Thirdly, we specify a regression model for research performance of PhD students; the measures used are these nosduocentered network measures.

We define this kind of network as a mixture between complete and egocentered networks according to social network theory. These two kinds of networks have been widely explained and studied by Granovetter (1973, 1982), Burt (1992), Coleman (1990), Knoke & Kuklinski (1982), Wasserman & Faust (1994), Scott (2000) and other authors. In this paper, we are not mainly concerned by discussing the adequacy of network theories (for instance structural holes, network closure and so on) but the network structure understood as network measures development. However, we will be able to make our own observations about the theoretical relevance of the measures related to this specific network.

The key of this nosduocentered network is that it is based on two egos which may be similar to a greater or lesser extent and which may be or fail to be linked. Network information is received from these two egos and there is no external information from alters. Alters in the networks are not linked to one another, but this does not mean that they do not have relations among them but only that we do not have this information. Summarizing, the two egos (from hereafter we call these egos as Ego_A and Ego_B) provide us with information between their mutual relationship and their relations with all alters in the network, but not about relationships among alters.

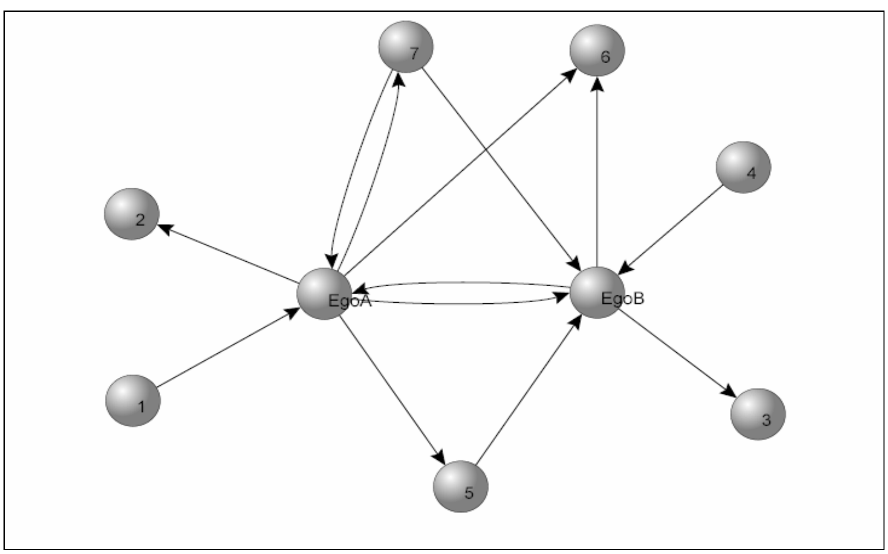

Figure 1: Example of nosduocentered network

An example of nosduocentered network is shown in Figure 1. This structure could be shown as a matrix, where the main characteristic would be that the cells of relations among alters must be zero, but the large number of zeros makes it easier to present this network structure as a graph or sociogram.

As we said, our second goal is to assess new social network measures for this network. We create measures of social networks based on complete network such as degree centrality, closeness, betweenness, density or centralization (Nieminen, 1974; Freeman, 1979; Scott, 2000; Wasserman & Faust, 1994) and some tailor-made network measures for the nosduocentered network, which are our main contribution in this article. For some of these measures, standard software for social network analysis can be used. We also have to take in consideration the new perspective of this network and which are the most relevant measures for research, which we will show in the results.

The third goal is to find a model for research performance of PhD students, based on these nosduocentered network measures. Once we obtain these measures, we specify a model in order to predict performance. PhD performance (in terms of publications and conferences) is the dependent variable and these nosduocentered network measures (Freeman's centrality and tailor-made measures) are the independent variables. We want to analyze the influence of these network measures over research performance of PhD students.

As we can see in Figure 1, we are able to find different relations in the network. In Table 1 the relations are shown and named. We have to differentiate between directed and undirected relationships. Figure 1 and Table 1 are made for directed relations. In the undirected case, we would have no arrows at the end of the lines and the relation would be symmetrical.

		Relations in Figure 1
a_I	Incoming to Ego_A, except contact from Ego_B	from alter1 to Ego_A
a_O	Outgoing from Ego_A, except contact to Ego_B	from Ego_A to alters 2, 5, 6. from Ego_A to 7.
b_I	Incoming to Ego_B, except contact from Ego_A	From alters 4, 5 to Ego_B
b_O	Outgoing from Ego_B, except contact to Ego_A	from Ego_B to alter 3
c_I	Shared incoming to Ego_A and Ego_B	from alter 7 to Ego_B and Ego_A
c_O	Shared outgoing from Ego_A and Ego_B	from Ego_B and Ego_A to alter 6
d_I	Ego_A incoming from Ego_B	from Ego_B to Ego_A
d_O	Ego_A outgoing to Ego_B	from Ego_A to Ego_B
e_I	Ego_B incoming from Ego_A	from Ego_A to Ego_B
e_O	Ego_B outgoing to Ego_A	from Ego_B to Ego_A

Table 1: Types of relation for the nosduocentered network from Figure 1.

Sub index "I" means incoming (relation to an ego) and "O" outgoing (relation from an ego) and by definition $d_i=e_O$ and $d_O=e_I$.

In the undirected network, the difference in the table is that we only would have a, b, c and d=e without sub indices. The fact that there is no distinction between incoming and outgoing relationships when the network is undirected and the set of c or common relationships can wider. For instance, in Figure 1 alter 5 would be now be a common relationship to both egos.

2. Definition of "nosduocentered network"

The network structure we propose is called "nosduocentered network". This network is a mixture between egocentered and complete network. Literally, "*ego*" means *"I"* and *"nosduo"* means *"the two of us"*. A nosduocentered network is formed by the relationships of two egos with a set of alters and the mutual relationship between both egos (if it exists). No relationships among alters are observed.

In some cases, it may not be enough to analyze some network structures according egocentric network theory, focused in only one ego. These egocentered networks are also called personal networks, they consist of a single individual (usually called ego) with one or more relations defined between him/her and a number of other individuals —the members of his/her personal network— called alters.

There are several networks that are difficult to study as one unique ego, since one ego has an especially relevant connection to another. Our case of study is to analyze research performance of PhD students in their doctorate. If we consider this kind of network structure as egocentric, we possibly miss relevant information. The main point for using nosduocentered network in our study is that PhD student's performance can not be understood without supervisor's influence.

For this reason, we consider that not only the students' network should be studied, but the supervisor' network is also necessary. If we omit supervisor from student's network or if we consider the supervisor as a simple other alter (without differentiation between supervisor and the rest of alters) in the students' network, we would be analyzing a biased students' network. In our experiment the main egos are PhD students and their supervisors. Others examples where the network is centered on a pair and not on individual could also be husband and wife or president and primer minister where it exists.

We know that an ideal situation would be when a researcher gets the complete network information (all relationships among all actors) but, it can happen that, it is really difficult to obtain the complete network. In several cases, alters are not central in the network and therefore more difficult is to reach than the two main egos, especially when we suppose that the network may be very large. This is another argument for using nosduocentered networks. In our case of study, these alters could be people from other departments or universities and to interview them would make to increase the cost and time and reduce the possibility of collaboration in the survey.

Even if you get the complete information, you have to analyze very carefully the raw data. The reason is that you could suffer from non-response and/or data quality problems because respondents have to answer about too many people and they simply put nonsenses or very homogeneous responses for everybody.

Thus, when researchers are involved in problems similar to these ones, they usually change the network from complete to egocentered in order to get at least some information about the topic they are analyzing, but we propose to move to a nosduocentered network instead of egocentered whenever two egos are clearly central in the network and alters are not easy to reach or the complete network is too large.

If researchers finally choose to use nosdoucentered network structure, they should take in consideration the characteristics described below:

- Two main actors (Ego_A and Ego_B) have to be clearly central and both have to be used as egos instead of one ego, as opposed to the egocentered network.
- Actors who are not defined as Ego_A or Ego_B are called alters.
- One major characteristic related to contacts of nosduocentered networks is that there is no relation observed among alters network members, because no information about it is available.
- Actors who do not have any contact are considered as isolates. These isolate members are not considered as a part of the nosduocentered network, so they do not appear in a nosduocentered network graph. This is one similarity with egocentered networks.
- Relationships or lines can be of different types such as directed or undirected and valued or binary.

3. Network measures for nosduocentered networks

Social network measures defined by Nieminen (1974), Freeman (1979), Freeman et al. (1980, 1991), Marsden & Lin (1982), Faust & Wasserman (1992) and Everett & Borgatti (1999) are used. The measures are basically centrality measures (Bonacich, 1987), which is restricted to the idea of point centrality, centralization (the extent to which the cohesion is organized around particular focal points), density (general level of cohesion in a network), which are used to refer to particular properties of the graph structure as a whole (Scott, 2000).

There are three major types of centrality measures; degree centrality (how well connected an actor is within the network), closeness (how close an actor is to the alters in the network) and betweenness (measures the extent to which a particular actor lies between the various other actors in the network).

We first adapt these social network measures to the nosduocentered network. This network is more focused in analyzing actor's centrality than the density for the whole network because only two egos can be analyzed, although, we can compute some measures for the whole.

Tailor-made measures, which are a second group of measures, will also be created in this section and they are specific measures for nosduocentered networks.

3.1. Degree centrality

The first type of centrality, which is useful for nosduocentered network, is called degree centrality. Degree centrality is a measure which indicates how well connected an actor is within the network. This type of centrality focuses only on direct or adjacent contacts (Wasserman & Faust, 1994) and is assessed by the number of geodesic (shortest path between two actors) contacts that an ego posses. The more contacts an ego has, the more central in terms of degree this ego is.

Nieminen's (1974) measurement counts the degree or number of adjacencies, for an actor p_k:

$$C_D(P_k) = \sum_{i=1}^{n} a(p_i, p_k) \qquad (1)$$

where:

- $C_D(P_k)$ = number of contacts connected to Ego k.
- $a(p_i,p_k)$ = contact for p_i to p_k. 0 or 1 in binary networks or any non-negative real number for valued networks.
- n = network size.

For undirected networks a general measure of degree centrality is obtained for Ego_A and Ego_B. Nosduocentered data make it impossible to compute centrality for alters. We have to differentiate between valued and binary networks. For binary data (Freeman et al., 1991) the centrality degree is the count of contacts for the ego analyzed and can also be computed as the sum of the 0 and 1 contacts. For valued data, the degree centrality measure is the sum of egos' contacts with alters in the network.

For directed networks a general degree centrality measure cannot be obtained. An outdegree centrality $C_{DO}(P_k)$ and an indegree centrality $C_{DI}(P_k)$ are obtained instead for both Ego_A and Ego_B. In directed networks, depending on the information we have (contacts from the egos, to the egos or both), outdegree, indegree or both centralities can be computed. We also must differentiate between binary or valued network data. For binary data (Freeman et al., 1991) outdegree centrality is the count of actors in the network to whom the ego gives its relation. Indegree centrality for an ego is the count of alters who give their relationship to the ego. For valued data, outdegree centrality is the sum of contacts that Ego_A or Ego_B have towards alters. Indegree for Ego_A or Ego_B is the sum of relationships that alters have towards Ego_A or Ego_B.

Freeman (1979) proposed a relative measure of degree centrality, $C'_D(P_k)$, in which the actual number of connections is related to the maxim number that it could sustain (Scott, 2000). We obtain the relative degree centrality for p_k:

$$C'_D(P_k) = \frac{\sum_{i=1}^{n} a(p_i, p_k)}{n-1} \qquad (2)$$

For binary data, this relative degree is described as a percentage of people in network related with ego analyzed. For valued data, the interpretation of relative outdegree is a mean of contacts for Ego_A and Ego_B.

Formulae (1) and (2) can be computed using standard software. As an alternative, computation by hand is very simple if we realize that in an undirected nosduocentered network there are only 4 possible relations (a, b, c and d=e) as shown in Table 1, which must only be summated, which will yield a proper sum (valued networks) or a count (binary networks).

In undirected nosduocentered networks, we can assess the centrality degree for Ego_A and Ego_B respectively, as follows:

$$C_D(P_a) = a + c + d \qquad C_D(P_b) = b + c + e \qquad (3)$$

Where a, b, c and d=e are defined in Table 1 and P_a and P_b refers to Ego_A and Ego_B.

If the network is directed, outdegree and indegree centralities are obtained separately. Sub indices will be necessary in order to be able to asses these centralities for asymmetric data:

$$C_{DO}(P_a) = a_O + c_O + d_O \qquad C_{DO}(P_b) = b_O + c_O + e_O \qquad (4)$$

$$C_{DI}(P_a) = a_I + c_I + d_I \qquad C_{DI}(P_b) = b_I + c_I + e_I \qquad (5)$$

Outdegree is indicated by the subindex "O" and indegree by the subindex "I".

Any of these expressions can be converted into relative centralities by dividing by n-1.

Some degree centrality properties for nosduocentered networks are:

- It is easily applicable to directed (asymmetric) and undirected (symmetric) networks.
- It can be used with binary and valued network data.
- It can be applicable to nosduocentered, egocentered and complete networks. Even, it can be computed with standard software for network analysis though for nosduocentered networks it can only be computed for Ego_A and Ego_B.
- It is a simple function of the components of network defined in Table 1.

3.2. Closeness centrality

Closeness centrality (Freeman, 1979; Harary, 1959) measures how close an actor is to the alters in the network. This centrality is obtained using the geodesic paths to reach all actors in a network (Sabidussi, 1966; Freeman, 1979). An actor is close to a large number of points, so closeness can be computed as the reciprocal of the sum of distances.

The general equation used come from Nieminen (1974) and it is the following:

$$C_C(P_k) = \left[\sum_{i=1}^{n} d(p_i, p_k) \right]^{-1} \qquad (6)$$

where:

- $Cc(P_k)$ = Closeness centrality
- $d(p_i,p_k)$ = distances: number of paths that ego k has to follow to reach each actor in the network.

According to this general formula, we can rapidly adjust this measure from complete to nosduocentered networks. Using the following formulae we will be able to obtain closeness centrality for undirected binary networks for Ego_A and Ego_B, respectively.

$$Cc(P_a)^{-1} = \sum d(p_i, p_a) = 1(a+c) + d + 2b(d) + (3b(1- d) + 2(1- d))(c>0) \qquad (7)$$

$$Cc(P_b)^{-1} = \sum d(p_i, p_b) = 1(b+c) + e + 2a(e) + (3a(1- e) + 2(1- e))\ (c>0) \qquad (8)$$

Where c>0 is a logical expression which equal 1 if true and 0 if false and a to e are defined in Table 1.

If both d=e and c are equal to zero, the network is not connected and closeness centrality can not be computed. This is unlikely to happen as it would mean that Ego_A and Ego_B have no direct relationship and no shared alter, so that they define two separate egocentered networks.

Using these formulae we obtain farness centrality, which definition is how far is an ego to all other members of the network (Wasserman & Faust, 1994). This measure is measuring distance and we would need a measure of proximity to compute centrality, which is called closeness and considering the network size. For this reason reverse of farness centrality is used, which is called closeness centrality, $Cc(P_k)$.

Comparisons of $Cc(P_k)^{-1}$ must be done in graphs of the same size. To solve that complexity, Beauchamp (1965) worked with a definition of relative centrality $C'c(P_k)$ for closeness centrality in p_k. This formula is the “inverse of the mean distance among p_k and alters”.

$$C'_C(P_k) = \left[\frac{\sum_{i=1}^{n} d(p_i, p_k)}{n-1} \right]^{-1} = \frac{n-1}{\sum_{i=1}^{n} d(p_i, p_k)} \qquad (9)$$

These equations and formulae are done by undirected binary networks. For directed networks, paths must be measured through lines that run in the same direction.

For directed networks “in-closeness” and “out-closeness” can be obtained. It is more likely that a number of actors can be at an infinite distance because a directed nosduocentered network may fail to be connected.

In Figure 2 we can see an example of not connected network with infinite distances, in which neither Ego_A nor Ego_B can reach alter 1.

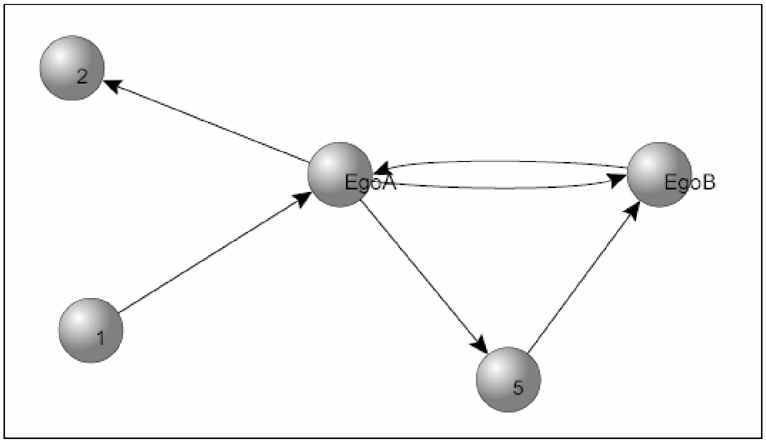

Figure 2: Not connected directed nosduocentered network

Some specific properties of closeness centrality for nosduocentered networks are:

- This centrality measure can be used only for binary networks. In fact, if we have valued data, we should to dichotomize the matrix to 0 and 1.
- This centrality measure can often lead to infinite distances for directed networks.
- This centrality measure can be applicable to nosduocentered and complete networks. Even, it can be computed with standard software for network analysis though for nosduocentered network it can only be computed for Ego_A and Ego_B.

3.3. Betweenness centrality

Betweenness centrality measures the extent to which a particular actor lies "between" the various other actors in the graph: an actor of relatively low degree may play an important "intermediary" role and so be very central to the network (Freeman, 1979; Freeman et. al, 1991; Scott, 2000). The advantage of this intermediary actor is described by Burt (1992) as structural holes. For instance, the existence of structural holes allows the third actor to act as a broker, tertius gaudens (Burt, 1992), rush (Anthonisse, 1971) and so on.

Betweenness centrality can be defined in terms of probability ($1/g_{ij}$),

$$i_{ij}(p_k) = \sum_{i<j} \frac{1}{g_{ij}} * g_{ij}(p_k) = \sum_{i<j} \frac{g_{ij}(p_k)}{g_{ij}} \qquad (10)$$

where:

- $i_{ij}(p_k)$ = probability that actor p_k is in a geodesic randomly chosen among the ones which join p_i and p_j.
- g_{ij} = number of geodesics that bond actors p_i and p_j.
- $g_{ij}(p_k)$= number of geodesics which bond p_i and p_j and contain the p_k.

Betweenness centrality is the sum of these probabilities.

For nosduocentered networks it is not possible to calculate this centrality measure. The reason is that we do not have the relations among alters needed to compute g_{ij}. Therefore, betweenness centrality, or any other measure which depends on relationships between third parties, can not be computed for nosduocentered network data.

3.4. Centralization Indicator

Centralization measures the extent to which cohesion is organized around particular focal points. Centralization is an expression of how tightly the network is organized around its most central actor (Freeman, 1979; Scott, 2000).

The general procedure is to look for differences between centrality scores of the most central point and those of all other points. We only have two egos; therefore we compare one centrality with the other.

Actor's centrality is standardized, taking in account the network size. The formula of centralization for the degree centrality is as follows:

$$C_D = \frac{C_D(P_a) - C_D(P_b)}{(n-1)} = C'_D(P_a) - C'_D(P_b) \quad (11)$$

where:

- C_D = Centralization indicator for degree centrality
- $C_D(P_a)$ = degree centrality for Ego_A
- $C_D(P_b)$ = degree centrality for Ego_B
- $C'_D(P_a)$ = relative degree centrality for Ego_A
- $C'_D(P_b)$ = relative degree centrality for Ego_B
- n = network size.

The interpretation for this formula is that we asses the relative degree for Ego_A minus the relative degree for Ego_B. If the result is positive, it means that Ego_A is more central than Ego_B. Therefore in this case Ego_A would have a larger non shared network. Since we only have two egos, this centralization indicator provides all needed information about centrality. Depending on the circumstances, in or out centralization or both can be computed adding the suitable sub indices for outdegree and indegree.

A centralization indicator can also be computed for closeness centrality using a very similar called C_C, (Wasserman & Faust, 1994) which is the difference between the two egos according to the closeness centrality measure. If the result is positive, it means that Ego_A is closer to the rest of actors in the network than Ego_B, therefore in this case Ego_A would be more central in terms of closeness. Depending on the circumstances, in or out centralization or both can be computed adding the suitable sub indices for outdegree and indegree. Standard software may be needed to compute centrality. Centralization must be worked by hand.

3.5. Density

Density (Burt, 1983) is also a measure for the whole network and the simplest idea is that the more actors are connected to one another, the more dense the network will be. According to Wasserman & Faust (1994), density of a graph is the proportion of possible lines that are actually present in the graph over the maxim possible number of lines that would be present if the network were complete. This maxim possible number is determined by the number of actors. Since there are n actors in the network in a complete undirected binary network there are n(n-1)/2 possible unordered pairs of actors, and thus n(n-1)/2 possible lines that could be present in the network or graph. Density is the ratio of number of lines present, *L*, to the maximum possible. The density of an undirected network, denoted by Δ, is calculated as:

$$\Delta = \frac{L}{n(n-1)/2} \tag{12}$$

where:

- L = number of present lines.
- n = network size.

The density of a graph goes from 0, if no lines are present, to a maximum of 1, if all lines are present.

We are able to adapt this density measure to a binary undirected nosduocentered network. There are n actors in the network and relationships among alters are excluded. Each of the network alters (n-2) can be connected to both egos and both egos can also be mutually connected, and thus (n-2)2+1 = 2n-3 possible lines could be present in the network. We can easily see that this measure is different from density for complete networks. We denote that the density for this type of network by ΔN, named "nosduocentered density" and computed as:

$$\Delta\text{N} = \frac{C_D(P_a) + C_D(P_B) - 1*(d>0)}{(2n-3)} = \frac{a+b+2c+d}{(2n-3)} \tag{13}$$

This is the density for binary undirected nosduocentered network and it counts d=e only once. The results interpretation can be made in the same way as density for complete networks, it means from 0 (no lines are present) to 1 (all possible contacts are present).

A simple measure which is not bounded between 0 and 1 could be:

$$C'_D(P_a) + C'_D(P_b) \tag{14}$$

This measure is the sum of relative degree centralities. Implicitly it gives a double weight to the relationship between both egos, which is not unreasonable given the importance of this key relationship in a nosduocentered network.

Several modifications should be done to compute density for binary directed nosduocentered networks. It is possible to work out the density of the network using indegree and outdegree together. Density measures for binary directed nosduocentered network are the following, sum of outdegree for Ego_A, $C_{DO}(P_a)$, and Ego_B, $C_{DO}(P_b)$, and also indegree for both egos, $C_{DI}(P_a)$ and $C_{DI}(P_b)$. All alters (n-2) can be connected to and from both egos and both egos can also be mutually connected, thus (n-2)4+2=4n-6 relationships are possible. With all these combinations, density for binary directed nosduocentered networks, ΔN_D[1], is:

$$\Delta N_D = \frac{C_{DO}(P_a) + C_{DO}(P_b) + C_{DI}(P_b) + C_{DI}(P_b) - 1*(d_O > 0) - 1*(d_I > 0)}{4n-6} \quad (15)$$

The reason for the introduction of these logical expressions in the formula is due to the fact that $d_O=e_I$ and $d_I=e_O$ must only be counted one.

We can also calculate this density of only a part of the relationships in a binary nosduocentered directed network, either incoming or outgoing relationships. The maximum number of relationships becomes (n-2)2+2 = 2n-2 and the density measure is computed as follows:

$$\Delta N_{DO} = \frac{C_{DO}(P_a) + C_{DO}(P_b)}{2n-2} \quad (16)$$

This partial density is also bounded between 0 and 1and the interpretation is the same as density for undirected networks. This outdegree measure is appropriate if we only are interested in knowing the relationship which goes from the egos to alters.

Density measures can also be computed using valued data; therefore the calculation will be different. The denominator should be changed; in fact, it should be multiplied by the maxim intensity that a line or relationship can have. For instance, if the intensity is from 0 (never) to 7 (every day of the week), then the denominator will be multiplied by 7 in order to cover the maxim frequency. The interpretation is different for valued and binary data. For valued data the result obtained is the mean of the strength of the contacts in the network as a whole as a proportion of the maximum possible strength. With valued data, researchers can not know if all contacts are present, but they are able to know a mean intensity for the whole network. The same mean intensity can arise from a large number of low intensity contacts or from a low number of high intensity contacts. Standard software may be needed to compute density. Density for nosduocentered networks must be worked by hand. Other alternative approaches using weights to compute measures with valued data are in Bonacich (1972).

3.6. Tailor-made measures for nosduocentered networks

The main idea for these tailor-made measures is go back to the origin and to use several measures that are as closely related as possible to a, b, c and d=e. We use measures which are especially significant for analyzing nosduocentered networks.

[1] Sub index "D" means that density is computed for directed nosduocentered networks

For instance, a can be considered as a single measure, since it measures the part of contacts that are linked with this ego and no one else. This part of the network is only reached by Ego_A. The same interpretation can be done for measure b but referring to Ego_B. It will be the case for one of our models to predict research performance of PhD students.

The centrality measures directly related to centrality of Ego_A could be:

- a = number or sum of direct contacts of Ego_A with alters others than Ego_B and Ego_B's contacts.
- c = number or sum of shared contacts of Ego_A and Ego_B. In a nosduocentered network, the number of shared contacts is closely related to density.
- d = number or sum of direct contact between Ego_A and Ego_B.
- (d/max)b = the influence in Ego_A from Ego_B's contacts through Ego_B, where max is the maximum intensity that a contact can have. It is weaker or stronger depending on the presence or strength of the contact Ego_A and Ego_B. What we mean is that these indirect contacts should be considered as influential; but that they must be given a weight lower than 1 depending on the intensity of the contact with Ego_B.

From the point of view of Ego_B:

- b = number or sum of direct contacts of Ego_B with alters others than Ego_A and Ego_A's contacts.
- c = number or sum of shared contacts of Ego_A and Ego_B.
- e = number or sum of direct contact between Ego_B and Ego_A.
- (e/max)a = the influence in Ego_B from Ego_A's contacts through Ego_A.

The tailor-made measures definitions are used for both binary and valued networks. These measures can also be used for directed and undirected networks. In case we have directed relationships, measures with in and out sub indices should be used.

4. Illustration

4.1. Data, sample and performance

The study is based on the nosduocentered network structure. The network contains the following actors. Ego_A who is a PhD student, Ego_B who is his/her supervisor and alters who are the people who belong to the PhD student's research groups. Somehow, alters are people who work in research close to PhD students and their supervisors.

The population studied are PhD students who began their doctoral studies at the University of Ljubljana in the academic years 1999/2000 and 2000/2001. These PhD students must have a strong tie with their university, in other words, these students must have grants, be assistants or be researchers hired for research projects. This choice has been made because these people have more frequent contact with other researchers, and they can spend more time doing research, which is their main job. Therefore, these students are likely to have more need for advice, cooperation or information than those who are only linked to the university by their doctoral studies and who may even not belong to a research group.

The procedure to figure out the networks, in our case, the research groups of PhD students was the following. Firstly, we defined theoretically the research group in order to know who could belong to the group. Then, PhD students were phoned in order to know who their supervisor was. The next step was to personally interview supervisors and they received name generator questions in order to obtain a list of influential research group members in connection with the topic of the dissertation of the PhD student. The network obtained does not have to correspond to any official or formal research group recognized by the university, and people in the research group or network can belong to other departments or universities or even work outside of the academic world. This is because we are interested in getting the research groups that are relevant to each PhD student.

Once we got the names for each student's research group members, a web questionnaire was designed about PhD students' performance in research. That questionnaire (De Lange et al., 2004) was created within the INSOC (International Network on Social Capital and Performance) research group. In fact, there were two questionnaires, one available for PhD students and other for their supervisors. PhD students and supervisors were asked about some network questions variables from their research group members. Each questionnaire was personalized with the list of their research group member names. Moreover, there was also an open list in case respondents wanted to introduce another influent person for them according to the question. These kind of open lists are very important for nosduocentered networks because they are the major source of a and b contacts. Obviously, a different network is obtained for each PhD student. In our study, a total of 64 pairs were finally analyzed. The response rate was 62% for PhD students and 52% for supervisors, and 30% for nosduocentered networks. We only took in consideration the groups which both answered, because it is the only way to be able to create the nosduocentered network.

After that, we were able to create a nosduocentered network for each four different networks (scientific advice, collaboration, emotional support and trust) for each. The questions are below:

- *Scientific Advice network*: Consider all the work-related problems you've had in the past year (namely since 1 November 2002) and that you were unable to solve yourself. How often did you ask each of your colleagues on the following list for scientific advice?
- *Collaboration network*: Consider all situations in the past year (namely since 1 November 2002) in which you collaborated with your colleagues concerning research, e.g. working on the same project, solving problems together, etc. The occasional piece of advice does not belong to this type of collaboration. How often have you collaborated with each of your colleagues concerning research in the past year?

- *Emotional Support network*: Imagine being confronted with serious problems at work; e.g. lack of motivation, problematic relationship with a colleague. To what extent would you discuss these problems with each of your colleagues?
- *Trust network*: In a working environment it can be important to be able to trust people in work-related matters (e.g. concerning the development of new ideas, your contribution to common goals, the order of co-authorship or the theft of new ideas). Consider the following opposite nouns: distrust and trust. The further to the left you tick off a box, the more you associate your relationship with a particular colleague with "distrust". The further to the right you tick off a box, the more you associate your relationship with that colleague with "trust".

The relations in the network are mostly frequency ranged from 1 (not in the last year) to 8 (daily) and intensity from 1 (certainly not) to 4 (certainly yes). Alters whose link value was 1 both to the supervisor and the PhD student are considered not to belong to the nosduocentered network.

The scientific advice and emotional support networks are directed networks with incoming. The interpretation of this is that the relation of either scientific advice or emotional support come from the alters to the ego (PhD student or supervisor) and we only know the relation from one side to another and we can not know if the ego will give the same support which he/she is receiving. The trust network is also directed but this has outgoing relationships, the relation of trust is from egos to alters and we do not know the trust from alters to egos. Finally, we consider the collaboration network as undirected, because the relation of working together should be mutual.

Using this information, we are able to compute the centrality, density, centralization and tailor-made measures for nosduocentered networks in the way we explain in Section 3. Once we obtain these measures, we will use them as independent variables for the specification of some regression models used to predict research performance of PhD students. Therefore, our goal is to assess the influence of these nosduocentered network measures for the networks of scientific advice, collaboration, emotional support and trust on research performance of PhD students.

To measure research performance, each PhD student was asked about his/her publications, conferences and workshops. We also asked about the type of publication, for instance, if it was an international or national book, chapter of book, paper and so on. The attendance to conferences without presentation is not considered as performance because it is more an input for PhD students than an output.

In all, PhD students were asked about seventeen different kinds of publications and we summarized them into four groups according to the importance of the publications. The first group was called "international articles" (int_art), which was composed of articles in international journals with or without impact factor. One of the reasons why we aggregated articles with or without impact factor was because we realized that a large number of PhD students did not know which journals had impact factor. The second group was called "publications with review" (pub_rev), which was composed of articles in a national journal and books and chapters of books and papers in proceedings, but all of these should had been subject to anonymous reviews. The third group is called "normal publications" (pub_norm), which is composed of articles in a national journal and books and chapter of books and papers in proceedings, without review process. The fourth, and last group, was a group of “conference papers” (pap_conf) that is, international and national conferences or workshops, with oral or poster presentation.

According to these variables, we create an index of performance (Y) for PhD research at university as dependent variable for the regression models. The index of performance is made giving different weights according to the importance of publications as follows:

$$Y = 2(int\_art) + 2(pub\_rev) + (pub\_norm) + (pap\_conf) \qquad (17)$$

We also tried to work with other weights from 4 to 1 or even the same weights for all types of publications, but the index we opt is the least skewed and shows the lowest differences among fields of research.

The next step is to specify a model in order to asses the research performance of PhD students according to the measures for nosduocentered networks.

4.2. Models for nosduocentered network

We specify three different linear regression models for each of the four types of networks (scientific advice, collaboration, emotional support and trust) to analyze the influence of nosduocentered network measures over research performance of PhD students. It means that for each model we obtain a result for each network. These networks have basically four dimensions (a, b, c, d=e) thus using a larger number of measures will lead to perfect collinearity. The three models are presented below:

Model 1: This model uses some of the tailor-made measures we created for the nosduocentered network. It focuses on frequency of direct contacts for Ego_A (PhD student) and moreover the importance of non contacts for Ego_A which are contacts of Ego_B (supervisor) weighted by the frequency of the contact from Ego_A to Ego_B. It is important to note that the qualitative variable field of study is used in all models because the three regression models fit better whether this variable is used. We make four field of study groups; the first group is composed by mathematics, physics and chemical; the second group by biology, genetics, pharmacy and nursing; the third group by electronics, informatics and engineering and the fourth group is composed by alters fields (arts, economics, etc...).

The hypothesis interpretation for this model is that direct contacts have an influence on the performance of PhD students but also supervisor’s contacts are influential if a rather strong relation between PhD and supervisor exists, (d/max)b. According to this interpretation, the model can be specified as follows:

$$Y = f\left(a, c, (d/max)b, d, F \right) \qquad (18)$$

where:

- F = Field of study.

Model 2: This second specified regression model must be differently specified depending on whether the nosduocentered network is directed or undirected. Research performance of PhD students would depend on key characteristics of nosduocentered networks which are the relative measures density, centralization (degree centrality is used) and size. Besides, field of study is also included in the model. As argued before, centrality is not needed because centralization already provides this information. The specification of the second model for undirected nosduocentered networks is:

$$\mathrm{Y} = \mathrm{f}\left(C'_D(P_a) + C'_D(P_b)\ ,\ C'_D(P_a) - C'_D(P_b)\ ,\ \mathrm{n}\ ,\ \mathrm{F} \right) \tag{19}$$

We can interpret this in the following way: using the sum and difference we are testing the variation for this network. When we sum we consider all contacts between egos and alters in the network. While when we use the difference of densities, we consider the difference between $\mathrm{Ego_A}$ and $\mathrm{Ego_B}$. If the difference result is positive means that $\mathrm{Ego_A}$ has a larger network than $\mathrm{Ego_B}$. Shared contacts do not affect this difference, because they are the same for both egos. This model construction has the attractive feature that the sum and difference will tend to have low collinearity.

Model 3: The third model is very similar to model 2, even in interpretation, but using absolute density and centralization measures instead of relative measures and size. The model can be specified as follows:

$$\mathrm{Y} = \left(C_D(P_a) + C_D(P_b),\ C_D(P_a) - C_D(P_b)\ ,\ \mathrm{F} \right) \tag{20}$$

4.3. Results

The regression model results to predict research performance of PhD students are shown in Table 2. It shows the multiple r for the global significance of models (the first row in each model) and standardized regression coefficient for each variable for each model. Field of study is omitted from the table because it is conceptually unrelated to networks. However, it is significant in most of the models and retained in all of them.

The results are in Table 2. The first model has global significance for the emotional support network and the variables which are significant are a (direct $\mathrm{Ego_A}$'s contact with alters others than $\mathrm{Ego_B}$ and EgoB's contacts) and c (shared contacts). The model for the collaboration network is also significant, but at 10% and only a is significant. The other two nosduocentered networks, scientific advice and trust, do not have neither any significant variable nor global significance for model 1. This means that the emotional and collaboration, networks help to predict research performance according to this model. All significant coefficients have a positive sign as expected. The main predictors of performance are the PhD student's direct contacts whether they are exclusive or shared with the supervisor. Indirect contacts through the supervisor lack significance in all models. The contact with the supervisor was also non significant but this may be due to the fact that this contact is present and strong in the 90% of all networks.

	Scientific Advice Network	Collaboration Network	Emotional Support Network	Trust Network
Model 1	.387	**.440****	**.463***	.398
a	.012	**.228****	**.265***	.157
c	.165	.143	**.364***	.220
(d/max)*b	-.151	.207	.056	.074
d	.115	.042	.017	.010
Model 2	.355	.414	**.466***	**.423****
Density	.186	.095	.118	-.001
Size	.054	**.309***	**.375***	**.288***
Centralization	.133	.215	.179	.084
Model 3	.334	**.448***	**.440***	**.452***
Absolute density	.069	**.491***	**.366***	**.369***
Absolute Centralization	.078	**.468***	.154	.289

Table 2: Multiple r and standardized regression coefficients.
* Significant (α=5%), ** Significant (α=10%).

The second model has global significance for the emotional support network and also at 10% for the trust network. Network size is significant for both networks. Since by definition a nosduocentered network contains no isolated alters, size by itself is a good summary of the number of contacts of the network. As expected the sign of the coefficient is consistently positive.

The third model has significance for the collaboration, emotional and trust networks at 5%. Absolute density is a significant variable for all these three networks; this would mean that is an important variable because it fits in most of the networks. Absolute centralization is significant only in the collaboration network.

Until here, the significance of each model variable according to the four different networks has been described. Now, we could make a general view of the Table 2 in order to figure out some global result for nosduocentered network measures and their influences over the research performance of PhD students. We can focus in looking the columns of the table. The network which fits worst with the three models is the scientific advice network, because none of the variables is significant. Then, collaboration and trust networks are significant for two models. Moreover, in the third model for collaboration network all independent variables are significant. The network which fits better is emotional support network because it is globally significant for all models.

The fact that the advice network fails to be significant is at first sight surprising. However, long term collaboration relationships will also include a lot of advice exchange. As the literature suggests (Bondonio, 1998; Bartus, 2000) informal networks of support and trust are also important to work performance, not only work networks.

If we focus on the rows of the table, we realize that no model has a substantially higher multiple r than any other. This is because all models include the main nosduocentered network characteristics, expressed and interpreted in a different manner. Which model to use would thus be rather of taste interpretation, though model 3 retains the advantage of being the most parsimonious.

5. Conclusions

In this article we defined the nosduocentered network structure. The key characteristic is that it is based on two egos and the relations exist between these two main egos and all alters, but relations among these alters are not observed. The next step has been to adapt some social network measures for complete networks to nosduocentered networks. These measures are degree, betweenness and closeness centrality, density and centralization. Furthermore, we design tailor-made measures. The models used in the four nosduocentered networks (scientific advice, collaboration, emotional support and trust) were specified in order to predict research performance of PhD students. This performance index was created according their articles, books, chapters of books and conferences, weighted by order of importance.

The results show that performance for PhD students depends mostly of the network of emotional support because it is significant for all three models. Trust and collaboration networks are significant for two models and scientific advice is not significant for any model.

In fact, the most important part of this article is not the results shown in Table 2. These are only indicators which help us to know if the models and the measures used are significant for predicting research performance of PhD students. It is essential to know that several specific measures for nosduocentered exist and that these measures have an interpretation and fit in several models. Measures related to the total intensity of contacts (e.g., density and degree centralization) seem to work particularly well to predict performance in research for PhD students at university. However, we could not generally recommend using one or another model to predict performance. For instance, if a researcher wants to predict performance using exclusively specific nosduocentered network measures, then we would recommend using model 1. Instead, if the researcher is interested in standard measures or size, they could use the model 2. Another possibility is if the researcher is interested in absolute measures or in using a parsimonious model, the one good option would be to use model 3.

In this paper we do not present nosduocentered networks as a cure-all. The ideal situation would be to have the complete network. However, when the complete network is unavailable due to high costs, low accessibility, poor data quality or low response rate, the nosduocentered network still makes it possible to define network measures which are interpretable, which have predictive power on performance, which are easy to compute and which are richer than those would be obtained from egocentered network alone.

References

Anthonisse, J. (1971), *The Rush in a Direct Graph*, University of Amsterdam Mathematical centre, Amsterdam.

Bartus, T. (2000), Fitting Social Capital, Informal Job Search, and Labor Market Outcomes in Hungary, *Connections*, 23, 72-83

Bondonio, D. (1998), Predictors of Accuracy in Perceiving Informal Social Networks, *Social Networks*, 20, 301-330.

Beauchamp, M. (1965), An improved Index of Centrality, *Behavioral Science*, 10, 161-163.

Bonacich, P. (1987), Power and Centrality: A family of Measures, *American Journal of Sociology*, 92, 1170-1182.

Bonacich, P. (1972), Technique for Analyzing Overlapping Memberships, in Costner, H., (Eds.), *Sociological Methodology*, San Francisco, Jossey-Bass. (pp. 176-185).

Burt, R.S. (1983), *Applied Network Analysis,* Beverly Hills, Sage.

Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Cambridge Massachusetts, Harvard University Press.

Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.

Everett, M.G. and Borgatti, S.P. (1999), The Centrality of Groups and Classes, *Journal of Mathematical Sociology*, 23, 181-201.

Faust, K. and Wasserman (1992), Centrality and Prestige: A Review and Synthesis, *Journal of Quantitative Anthropology*, 4, 23-78.

Freeman, L.C. (1979), Centrality in Social Networks: Conceptual Clarification, *Social Networks,* 1, 215-239.

Freeman, L.C., Borgatti, S.P. and White, D.R. (1991), Centrality in Valued Graphs: A Measure of Betweenness Based on Network Flow, *Social Networks*, 13, 141-154.

Freeman, L.C., Roeder, D. and Mulholland, R.R. (1980), Centrality in Social Networks:II. Experimental Results, *Social Networks*, 2, 119-141.

Granovetter, M. (1973), The Strength of Weak Ties, *American Journal of Sociology*, 78, 1360-1380.

Granovetter, M. (1982), The Strength of Weak Ties: A Network Theory Revisited, in Marsden, P.V. and Lin, N., (Eds.), *Social Structure and Network Analysis*, Newbury Park, CA., Sage (pp, 105-130).

Harary, F. (1959), On the Measurement of Structural Balance, *Behavioral Science*, 4, 316-323.

Knoke, D. and Kuklinski, J.H. (1982), *Network Analysis*, Newbury Park, CA, Sage.

De Lange, D., Agneessens, F. and Waege, H. (2004), Asking Social Network Questions: a Quality Assessment of Different Measures, *Metodološki Zvezki. Advances in Methodology and Statistics,* 1, 351-378.

Marsden, P. and Lin, N. (1982), *Social Structure and Network Analysis*, Beverly Hills, Sage.

Nieminen, J. (1974), On the Centrality in a Graph, *Scandinavian Journal of Psychology*, 15, 322-336.

Sabidussi, G. (1966), The Centrality Index of a Graph, *Psychometrika*, 31, 581–603.

Scott, J. (2000), *Social Network Analysis: A Handbook*, Newbury Park, CA, Sage.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, New York, Cambridge University Press.