Duocentered networks

Lluís Coromina\textsuperscript{a}, Jaume Guia\textsuperscript{b}, Germà Coenders\textsuperscript{a,\ast}, Anuška Ferligoj\textsuperscript{c}

\textsuperscript{a} Department of Economics, University of Girona, Faculty of Economics, Campus Montilivi, 17071 Girona, Spain
\textsuperscript{b} Department of Organization, Business Management and Product Design, Faculty of Tourism, Pujada dels Alemanys 4, 17071 Girona, Spain
\textsuperscript{c} Faculty of Social Sciences, University of Ljubljana, Kardeljeva pos čad 5, 1000 Ljubljana, Slovenia

Abstract

When a pair of individuals is central to a research problem (e.g., husband and wife, PhD student and supervisor) the concept of “duocentered” networks can be defined as a useful extension of egocentered networks. This new structure consists of a pair of central egos and their direct links with alters, instead of just one central ego as in the egocentered networks or multiple egos as in complete networks. The key point in this kind of network is that ties exist between the central pair of egos and between them and all alters, but the ties among alters are not considered. Duocentered networks can also be considered as a compromise between egocentered and complete networks. Complete network measurements are often costly to obtain and tend to contain a large proportion of missing data (especially for peripheral actors). Egocentered network data are less costly but a lot of information is lost with their use when a pair of individuals is the relevant unit of analysis.

From the definition of duocentered networks, we develop new social network measures, some of which based on the measures for complete networks such as degree, closeness centrality or density, both absolute and relative, while others are tailored to dealing with specific characteristics of the duocentered network structures.

The proposed measures are used in the analysis of the networks of Slovenian PhD students and their supervisors. We specify three regression models to predict PhD student’s academic performance on the basis of these duocentered network measures for different relations such as advice, collaboration, emotional support, and trust. The results show that absolute duocentered measures predict performance best. When compared with egocentered network measures a higher predictive power of duocentered networks is revealed.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Duocentered networks; Network measures; PhD students; Academic performance; Performance; Central actors

1. Introduction

The most typical network structures found in the literature are complete and egocentered networks, the former is found when the structure of the network as a whole is relevant to a research problem, and the latter when only the ties of a particular actor are considered to be relevant to the problem at hand. In some cases a pair of actors may be central (e.g., husband and wife, buyer and seller) and we may intend to study the behaviour, performance or social capital of these two specific actors in the network. Moreover, there are cases when egocentered networks are difficult to interpret if only one ego is considered, since this ego might have an especially relevant connection with another actor (for instance a PhD student and his/her supervisor). This is certainly the case when predicting PhD students’ academic performance is the problem at hand, which is what the illustration in this article is about. Therefore, in our study the pair of egos is made of the PhD student and his/her supervisor.

For the case when there is a pair of relevant central actors in a network, we suggest a new type of network that we call duocentered network. This kind of network is composed of a pair of central egos and their relationships with alters, while the ties amongst these alters are neglected. An example of a duocentered network is shown in Fig. 1.

Our example fits particularly the concept of duocentered network. The reason of using duocentered networks is that PhD students’ performance cannot be well explained leaving out their supervisor’s influence. Therefore, not only should the students’ egonetwork be analyzed, but also should the supervisors’ egonetwork. If we just take the students’ egonetwork and dismiss the supervisor’s egonetwork, that is, if we consider the supervisor as simply another alter in the student’s egocentered network, we are dismissing some of the supervisor’s contacts who might be very relevant to the student’s performance on the doctoral thesis. Other examples of networks where the relevant unit of analysis is a pair of egos instead of just one individual are: husband and...
network interactions. Similar studies were done by Freeman et al. (1987) focusing on the respondents' accuracy in recalling relations will likely be rated highly on the other relations). In other words, the "halo effect" (Thorndike, 1920) and refers to a rater's tendency to rate a target equally on conceptually distinct aspects (e.g., someone whose contact is rated positively on some relations will likely be rated highly on the other relations). In the 1970s and beginning of the 1980s, Killworth and Bernard (1976), Bernard and Killworth (1977) and Bernard et al. (1980, 1982) studied respondent accuracy and found that respondents seemed to be relatively inaccurate in reporting on their social network interactions. Similar studies were done by Freeman et al. (1987) focusing on the respondents' accuracy in recalling who attended a specific social event. Hlebec and Ferligoj (2002) showed that accuracy also depends on the measurement model and measurement scale. As regards the correlates of inaccuracy, according to Romney and Faust (1982) accuracy depends on the actual frequency of interaction. Johnson and Orbach (2002) also study some characteristics of the actors that make their responses more or less accurate and basically reach the same conclusion. This problem will thus be less likely present in duocentered networks because, if the pair of egos in the network is relevant at all to the research problem, they will also be likely to be central in the network and have frequent contacts.

Unit non-response has been addressed in a number of ways for complete networks. One possibility is to make an extra recall effort in order to gather cooperation from all actors. This solution can be very expensive if telephone or face-to-face contact is used as the mode of data collection. Besides, responses from actors in the periphery of the network with fewer or less frequent contacts (Costenbader and Valente, 2003) are difficult to obtain. Another possibility is to use proxy respondents, who are asked about relationships among third parties (Krackhardt, 1987; de Lange, 2005). An accurate proxy can be a starting point for an imputation procedure to solve missing data problems. One advantage of this solution is a lower cost of the survey, mainly for large networks. However, even if self-reports are far from being a gold standard, there are several problems with proxy data. The burden of responding about relationships among third parties is often related to low respondent motivation and some types of relationships are difficult to respond accurately, sensitive or both (e.g., involving trust, emotional support or social activities outside work, which, proxy respondents may even fail to know about), especially for large networks and peripheral actors (Bernard et al., 1980, 1982; Freeman et al., 1987; de Lange, 2005). Unit non-response problems will thus be much smaller in duocentered networks than in complete networks, because only two individuals – the pair of central respondents – have to be approached in order to obtain the whole duocentered network.

Another alternative to obtain social network measures is, of course, to take a step backwards and to measure egocentered networks. There are two types of egocentered networks, the first type include contacts between the ego and all alters and also contacts among alters. This type of egocentered network will to some extent share the missing data problems of complete networks. The second type includes only contacts between the ego and all alters. This second type of egocentered networks somehow reduces the inaccuracy and unit non-response problems for identical reasons as the duocentered network does it. The problem of egocentered networks is that they fail to account for the second key actor when there is one. By now it will be apparent to the reader that this second type of simpler egocentered networks and duocentered networks are similar in a number of respects. From now on, when we refer to egocentered networks we will refer only to this second type (i.e. where contacts among alters are not accounted for). Duocentered networks can in fact be better understood as a generalization of egocentered networks than as a simplification of complete networks. However, as we will see, duocentered networks make it possible to compute a larger array of network measures than egocentered networks and some of these measures are closer to the complete network equivalents.
Costenbader and Valente (2003) explored the possibility of reducing the measurement costs for complete networks by sampling network members. Our analysis differs from theirs in two fundamental aspects. First, we do not intend to reproduce the complete network, as we argue that for some research problems the duocentered network is interesting per se. Thus, we do not intend to measure the complete network by the subset of its ties that are in the duocentered network. Second, the pair of egos around which the duocentered network is build are not meant to be a sample of network members in the sense given by for instance Frank (1978), but are purposely selected because they are central to a particular research question. The existence of a pair of central actors relevant to a research problem must always be the main argument for using duocentered networks, regardless of all missing data issues that have been raised. The local neighbourhood of these two actors is supposed to be the focus of interest rather than the whole network in which they are embedded.

In this article, we firstly define the duocentered network structure. Secondly, we define social network measures for this network structure based on Freeman’s (1979) complete network measures (degree centrality, closeness, etc.). For some of these measures, standard software for social network analysis such as Pajek developed by Vladimir Batagelj and Andrej Mrvar (de Nooy et al., 2005) or UCINET (Borgatti et al., 2002) can be used. Next we define some tailor-made measures for duocentered network structures for the particular research problem of predicting PhD student’s academic performance. These measures were computed from the network data of Slovenian PhD students, where the measured relations were scientific advice, collaboration, emotional support, and trust. Thirdly, we specify a regression model for PhD students’ academic performance using these duocentered network measures. Finally, other regression models are used to compare egocentered versus duocentered networks in order to find out which of them explains performance best.

2. Definition of duocentered network

A duocentered network is a compromise between an egocentered network and a complete network that can be used when there a pair of central actors is relevant to the research problem at hand. Egocentered networks and complete networks have been widely explained by Granovetter (1973, 1982), Burt (1992), Coleman (1990), Knopke and Kuklinski (1982), Wasserman and Faust (1994), and Scott (2000), among others. We are not primary concerned with discussing the adequacy of different network theories (for instance structural holes, network closure) but by the most commonly used network measures, which we will adapt to the duocentered case. However, we can present our own view of the theoretical relevance of the proposed network measures defined for this particular network structure.

The main characteristic of a duocentered network is that it is built around a pair of egos. Network information is obtained from these two egos and there is no information gathered from alters. The ties among alters in the network are thus not measured. This does not mean that these ties do not exist, but only that they are not observed or taken into consideration. Summarizing, the pair of central egos (from now on we denote them as EgoA and EgoB) provide us with information regarding their mutual relationship and their ties to their alters in the network, but not about relationships among alters.

As we can see in Fig. 1, we can find different ties in the network. In Table 1 the types of ties are shown and named. We have to differentiate between directed and undirected relationships. Fig. 1 and Table 1 present a general example for directed ties. In the undirected case, we would have edges instead of arcs, as the relation would be symmetrical.

In Table 1 sub-index “I” means incoming tie to an ego and “O” outgoing tie from an ego and by definition \(d_I = e_O\) and \(d_O = e_I\). A discrepancy between these quantities can be treated by averaging them or by the maximum frequency, depending on the network relation. Very often, either only ingoing or only outgoing ties are measured, but we want to produce as general an example as possible.

For the undirected case, \(a, b, c\) and \(d = e\) do not have sub-indexes. The fact that there is no distinction between incoming and outgoing ties when the network is undirected, results in the set of \(c\) or common relationships being wider. For instance, if the network in Fig. 1 would be undirected, alter 5 would be connected to both egos.

The following characteristics of duocentered networks should be considered:

- Two main actors (EgoA and EgoB) have to be clearly central and are considered as egos,
- Actors who are not defined as EgoA or EgoB are called alters,
- No relationships are observed among alters,
- Actors who do not have any contact with the egos are considered as isolates. These isolate members are not considered as a part of the duocentered network, so they do not appear in the network.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Types of ties for the duocentered network from Fig. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_I)</td>
<td>Incoming to only EgoA, except contact from EgoB</td>
</tr>
<tr>
<td>(a_O)</td>
<td>Outgoing from only EgoA, except contact to EgoB</td>
</tr>
<tr>
<td>(b_I)</td>
<td>Incoming to only EgoB, except contact from EgoA</td>
</tr>
<tr>
<td>(b_O)</td>
<td>Outgoing from only EgoB, except contact to EgoA</td>
</tr>
<tr>
<td>(c_I)</td>
<td>Shared incoming to EgoA and EgoB</td>
</tr>
<tr>
<td>(c_O)</td>
<td>Shared outgoing from EgoA and EgoB</td>
</tr>
<tr>
<td>(d_I)</td>
<td>EgoA incoming from EgoB</td>
</tr>
<tr>
<td>(d_O)</td>
<td>EgoA outgoing to EgoB</td>
</tr>
<tr>
<td>(e_I)</td>
<td>EgoB incoming from EgoA</td>
</tr>
<tr>
<td>(e_O)</td>
<td>EgoB outgoing to EgoA</td>
</tr>
</tbody>
</table>
• Relationships or ties can be of different types: directed or undirected and valued or binary.

3. Network measures for duocentered networks

To begin with, some social network measures defined by Nieminen (1974), Freeman (1979), Freeman et al. (1980, 1991), Marsden and Lin (1982), Faust and Wasserman (1992) and Everett and Borgatti (1999) are used. The first type is centrality (Bonacich, 1987). There are three major types of centrality measures (Freeman, 1979); degree centrality (how well connected an actor is within the network), closeness centrality (how close an actor is to the alters in the network) and betweenness centrality (the extent to which a particular actor lies between the various other actors in the network). The second type is diversity (the extent to which the two actors in a duocentered network are different), and the third type is density (general level of cohesion in undirected duocentered networks, is called degree centrality, which is a measure of ties (Wasserman and Faust, 1994). The more contacts an ego has, the more central in terms of degree the ego is.

We first adapt these social network measures to the duocentered network concept. Tailor-made measures, which are measures specifically designed to solve our particular research problem are developed next.

3.1. Degree centrality

The first type of centrality, which can be computed for duocentered networks, is called degree centrality, which is a measure that indicates how well an actor is connected within the network. This type of centrality focuses only on direct or adjacent contacts (Wasserman and Faust, 1994). The more contacts an ego has, the more central in terms of degree the ego is.

Nieminen’s (1974) degree measurement counts the number of adjacencies for an actor

\[ C_D(p_k) = \sum_{i=1}^{n-1} t(p_i, p_k) \]  

where \( C_D(p_k) \) is the number of direct contacts to actor \( k \) (in our case Ego \( k \)), \( t(p_i, p_k) \) is a tie from \( p_i \) to \( p_k \) (0 or 1 for binary networks or any non-negative real number for valued networks) and \( n \) is the duocentered network size including both egos and all their alters.

For undirected networks a general measure of degree centrality is obtained for EgoA and EgoB. We have to distinguish between binary and valued networks. For valued data, degree centrality is the sum of the egos’ direct contacts with alters in the network. For binary data, degree centrality is the count of contacts for the considered ego but it can also be computed as the sum of the 0 and 1 values.

For directed networks, depending on the information we have (contacts from the egos, to the egos or both), outdegree \( C_{DO}(p_i) \), indegree \( C_{DI}(p_i) \) or both centralities can be computed, as counts or sums (binary data) or only as sums (valued data).

Freeman (1979) proposed a relative measure of degree centrality, \( C'_{D}(p_k) \), in which the actual count of connections is related to the maximum number that could exist (Scott, 2000).

We obtain the relative degree centrality for \( p_k \):

\[ C'_{D}(p_k) = \frac{\sum_{i=1}^{n-1} t(p_i, p_k)}{n - 1} \]  

For binary data, this relative degree centrality is the percentage of people in the network related to the considered ego. For valued data, it is the mean intensity of contacts to the ego.

Eqs. (1) and (2) can be computed using standard software for social network analysis such as Pajek or UCINET. As an alternative, computation by hand is very simple if we realize that in an undirected duocentered network there are only 4 types of ties \( (a, b, c \text{ and } d = e) \) as shown in Table 1, which only need to be added. This will yield a proper sum (valued networks) or a count (binary networks).

In undirected duocentered networks, we can compute Eq. (1) for EgoA and EgoB respectively, as follows:

\[ C_{D}(p_A) = a + c + d, \quad C_{D}(p_B) = b + c + e \]  

where \( a, b, c \text{ and } d = e \) are defined in Table 1 and \( p_A \) and \( p_B \) refer to EgoA and EgoB.

If the network is directed, outdegree and indegree centralities are obtained separately and sub-indexes will be necessary:

\[ C_{DO}(p_A) = a_0 + c_0 + d_0, \quad C_{DO}(p_B) = b_0 + c_0 + e_0 \]  

Outdegree is indicated by the sub-index “O” and indegree by the subindex “I”. All these centrality measures can be converted into relative ones (Eq. (2)) by dividing with \((n - 1)\).

Some properties of degree centrality measures for duocentered networks are:

• They can be used for directed (asymmetric) and undirected (symmetric) networks and are defined for EgoA and EgoB.
• They can be used for binary and valued network data.
• They can be computed either with standard software for network analysis or as simple functions of the network components defined in Table 1.

3.2. Closeness centrality

Closeness centrality (Harary, 1959; Freeman, 1979) measures how close an actor is to the rest of the network. This centrality is obtained by using the geodesic paths to reach all actors in a network (Sabidussi, 1966; Freeman, 1979). Closeness can be computed as the reciprocal of the sum of distances from an actor to the other actors.

The general definition used comes from Nieminen (1974):

\[ C_C(p_k) = \left( \sum_{i=1}^{n} \text{dist}(p_i, p_k) \right)^{-1} \]  

where \( C_C(p_k) \) is the closeness centrality of the actor \( k \) and \( \text{dist}(p_i, p_k) \) stands for distance, that is the length of the shortest path from actor \( k \) (in our case an ego) to actor \( i \).
From this general expression, we can easily adjust this measure for duocentered networks. Using the following formulae we can obtain the inverse of closeness centrality for undirected binary networks for EgoA ($p_A$) and EgoB ($p_B$), respectively:

$$C(p_A)^{-1} = \sum_{i=1}^{n} \text{dist}(p_i, p_A) = 1(a + c) + d + 2b(d)$$

$$+ (3b(1 - d) + 2(1 - d))(c > 0)$$

$$C(p_B)^{-1} = \sum_{i=1}^{n} \text{dist}(p_i, p_B) = 1(b + c) + e + 2a(e)$$

$$+ (3a(1 - e) + 2(1 - e))(c > 0)$$

Eqs. (7) and (8) have to be inverted in order to obtain closeness centrality.

If both $d = e$ and $c$ are equal to zero, the network is not connected and closeness centrality cannot be computed. This is unlikely to happen as it would mean that EgoA and EgoB have no direct relationship and no tie to a common alter, so that they define two separate egocentered networks.

Comparisons of $C(p_k)$ can be done only for networks of the same size. To overcome this limitation, Beauchamp (1965) suggested a relative definition $C^{'}(p_k)$ for closeness centrality of $p_k$, the inverse of the mean distance between $p_k$ and all alters:

$$C^{'}(p_k) = \left[ \frac{\sum_{i=1}^{n} \text{dist}(p_i, p_k)}{n - 1} \right]^{-1} = \frac{n - 1}{\sum_{i=1}^{n} \text{dist}(p_i, p_k)}$$

These definitions are used for undirected binary duocentered networks. For directed networks, paths must be considered through lines that run in the same direction. In-closeness centrality and out-closeness centrality can thus be obtained. However, it is more likely that a number of actors can be at an infinite distance because a directed duocentered network may fail to be connected. In Fig. 2 there is an example of an unconnected network with infinite distances, in which neither EgoA nor EgoB can reach alter 1.

Some properties of the closeness centrality measures for duocentered networks are:

- They can be used only for binary networks and are defined for EgoA and EgoB. If we have valued network data, we should dichotomize them into 0 and 1.
- They can often lead to infinite distances for directed networks.
- They can be computed either with standard software for network analysis or as simple functions of the network components defined in Table 1.

### 3.3. Betweenness centrality

Betweenness centrality measures the extent to which a particular actor lies on the path “between” the various other actors in the network: an actor of relatively low degree may play an important “intermediary” role and so be very central in the network (Freeman, 1979; Freeman et al., 1991; Scott, 2000). Such an intermediary role is described by Burt (1992) when he defined the concept of structural holes. For instance, the existence of a structural hole allows the relevant actor to act as a broker, also named tertius gaudens by Burt (1992).

Betweenness centrality is defined as the sum of the probabilities $i_j(p_k)$ that the actor $p_k$ is on a geodesic, randomly chosen among the ones which connect $p_i$ and $p_j$:

$$C_B(p_k) = \sum_{i<j}^{n} \frac{i_j(p_k)}{g_{ij}}$$

where $C_B(p_k)$ is the betweenness centrality of actor $k$, $g_{ij}$ the number of geodesics that connect actors $p_i$ and $p_j$, and $i_j(p_k)$ is the number of geodesics which contain $p_i$ and $p_j$ and contain the actor $p_k$.

For duocentered networks it is not possible to calculate this centrality measure because relationships among alters are needed. Any other measure, which depends on relationships between third parties, cannot be computed for duocentered network data.

### 3.4. Diversity

Diversity in a duocentered network indicates how different both egos are. More precisely, we define this measure as the difference between centrality scores of both central actors (EgoA and EgoB). The expression we suggest as diversity for degree centrality is

$$D_D = C_D(p_A) - C_D(p_B)$$

The interpretation of this measure is as follows: If the result is positive it means that EgoA is more central than EgoB; in other words, that EgoA has a larger non shared network. Since we only have two egos, the diversity measure provides all needed information about degree centrality. Depending on the circumstances, in diversity, out diversity or both can be computed by adding the suitable sub-indexes for indegree and outdegree centrality.

The diversity measure can also be computed for closeness centrality using a very similar measure called $D_C$ which is the difference between the closeness centralities of both egos. A positive result means that EgoA is closer to the rest of actors.
in the duocentered network than EgoB. Standard software for
social network analysis can be used to compute the centralities.
The diversity measure is not implemented in standard software
and thus has to be computed separately.

3.5. Density

Density (Burt, 1983) is also a measure for the whole net-
work structure. The simplest idea is that the more actors are
connected to one another, the denser the network is. According
to Wasserman and Faust (1994), the density of a network is the
proportion of ties that are actually present in the network over
the maximum possible number of ties that could be present if the
network were complete. The number of actors determines this
maximum possible number. Since there are n actors in a com-
plete undirected binary network there are \(\frac{n(n-1)}{2}\) possible
unordered pairs of actors, and thus \(\frac{n(n-1)}{2}\) possible ties that
could be present in the network. Density is the ratio of number
of ties present, \(L\), to the maximum possible. The density of an
undirected complete network, denoted by \(\Delta\), is calculated as

\[
\Delta = \frac{L}{n(n-1)/2}
\]

The minimum density of a network is 0, if no ties are present,
and the maximum is 1, if all ties are present.

We can adapt this density measure to a binary undirected
duocentered network. Let us assume that there are \(n\) actors
in the network and relationships among alters are excluded.
Each of the \((n-2)\) alters can be connected to both egos and
both egos can be mutually connected, and thus there are
\((n-2)^2+2\) possible ties in the network. We denote the
density for this type of network by \(\Delta_N\) (duocentered density).
It can be computed as follows:

\[
\Delta_N = \frac{C_{DO}(p_A) + C_{DO}(p_B) - 1}{2n - 3} = \frac{a + b + 2c + d}{2n - 3}
\]

We can easily see that the suggested measure for density for
binary undirected duocentered networks is defined in the same
way as for the complete network case but it is computed differ-
ently. The interpretation can be made in the same way as for the
complete network case, 0 meaning that no ties are present and 1
that all possible ties are present. The logical expression \((d > 0)\)
implies that \(d = e\) is counted only once.

A simpler measure which is not bounded between 0 and 1 is

\[
C_D(p_A) + C_D(p_B)
\]

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 duocentered network.

Several modifications should be made to compute density
for binary directed duocentered networks. It is possible to work
out the density of the network by using indegree and outde-
gree together. The simple measure in Eq. (14) becomes the
sum of outdegree and indegree relative centralities, \(C_{DO}'(p_A) +
C_{DO}'(p_B) + C_{DI}'(p_A) + C_{DI}'(p_B)\). As regards the more usual def-
inition in Eq. (13), all alters \((n - 2)\) can be connected to and from
both egos and both egos can also be mutually connected, thus
\((n - 2)^2 + 2\) ties are possible and thus density for binary
directed duocentered networks is

\[
\Delta N = \frac{-1(d_O > 0) - 1(d_I > 0)}{4n - 6}
\]

The logical expressions in the formula imply that \(d_O = e_I\) and
\(d_I = e_O\) are counted only once.

We can also calculate this density only for a part of the
binary directed duocentered network, either incoming or out-
going relationships. The maximum number of ties becomes
\((n - 2)^2 + 2\) and, for instance for outgoing relation-
ships, the density measure is computed as follows:

\[
\Delta N_O = \frac{C_{DO}(p_A) + C_{DO}(p_B)}{2n - 2}
\]

This partial density is also bounded between 0 and 1 and its
interpretation is the same as for the undirected case.

Density measures can also be computed for valued network
data with a small change in some of the definitions. The deno-
inator in Eqs. (13), (15) and (16) should be changed. In fact, it
should be multiplied by the maximum intensity that a tie can
have. For instance, if the intensity is from 0 (never) to 7 (daily),
then the denominator will be multiplied by 7 in order to cover
the maximum frequency. The interpretation for valued networks 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, the same mean intensity can arise from a large number of
low intensity contacts or from a low number of high intensity
contacts. If researchers are interested in the percentage of exis-
tent contacts they can always dichotomize the valued network.
Standard software may be used to compute centrality but density
for duocentered networks must be worked by hand.

3.6. Tailor-made measures for duocentered networks

The main idea behind these tailor-made measures is to go
back to the origin and make them as closely related as possible
to \(a\), \(b\), \(c\) and \(d = e\) and meaningful to specific research questions.

In our case we have to predict the performance of EgoA (PhD
student) and we created measures for this purpose. Other mea-
sures could be developed to predict the performance of EgoB or
of the team composed by both egos. Researchers can create their
own measures that are useful for their specific study.

For instance, for our specific research question, parameter \(a\)
from Table 1 can be considered as a measure on its own, since
it indicates the alters that are linked to EgoA and to no one else:

- \(a\) is the count or sum of direct contacts of EgoA with alters
  other than EgoB and EgoB’s contacts.

Other measures that can be meaningfully related to the
performance of EgoA could be:
4. Analysis of the Slovenian PhD students’ performance

4.1. Sampling design and network data

There is not a unique way of sampling duocentered networks. We start with our sampling design, and suggest ways in which it can be extended to a more general design, so that researchers can create their own according to their research problem.

The dyads formed by both egos must be defined a priori; for instance a married couple is a married couple even if their ties are weak or non-existent. The duocentered network we are going to study contains the following central actors: EgoA, who is a PhD student and EgoB who is his/her supervisor. A supervisor is officially assigned to each PhD student, whether the supervisor provides frequent advice is another story, but both actors are central in the research problem of PhD students’ academic performance.

Our population of dyads are PhD students who began their doctoral studies at the universities in Slovenia in the academic years 1999/2000 and 2000/2001 and their supervisors. These PhD students obtained a national grant and are research assistants at a university or at a research organization. They have to teach a few hours per week and spend most of the time doing research and work on their dissertations. The sampling frame, that is the list of the PhD students and their supervisors was obtained from the Ministry of Science, Education and Sports of the Republic of Slovenia. In our case, we sample PhD student-supervisor dyads. As an alternative we could have sampled students and later asked them who their official supervisor was. Of course, the sample must be random if confidence intervals or statistical tests are to be used at all. If the sample is more complex than a simple random sample, then adjustments to standard errors of estimates (for instance in the type of regression model that we are using) are needed to account for clustering, stratification or weights.

In our application, alters are the people who belong to the PhD student’s and the supervisor’s research group. Therefore, alters are people who work with the PhD student and his/her supervisor in the research. We organized some focus group sessions in order to create a definition of the term “research group”. The next step was a face-to-face interview with the supervisors, who received name generator questions in order to give a list of influential research group members in connection with the topic of the dissertation of the PhD student. Thus, name generator questions were asked to one or both egos may be one way of specifying the network boundary. Another alternative may be to use some type of a priori group definition. For instance, we could have used the official research groups defined by the Ministry. Our choice was the first one because we were interested in getting the really relevant members to the PhD student’s research work, which could even belong to other departments, universities, institutes or even work outside of the academic world.

Once we got the names for each student’s research group members, a web questionnaire was designed and administered to both PhD students and their supervisors. This questionnaire (de Lange, 2005) was created within the INSOC (International Network on Social Capital and Performance) research group. PhD students and supervisors were asked about background, attitudinal and network questions related to their research group members. Each questionnaire was personalized with the list of their research group member names.

The network data collected included four relations: scientific advice, collaboration, emotional support, and trust. These relations have been selected according to the network literature in the organizational context (Sparrowe et al., 2001; Hansen, 1999). The network questions asked are:

- **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”.

Beside the list of research group members, there was also an additional open list for scientific advice and collaboration...
networks, in case respondents wanted to introduce some other influential persons for them. These additional open lists can be a good source of \( a \) and \( b \) contacts in the duocentered network.

The responses are frequency for scientific advice and collaboration (bounded from 1 “not in the last year” to 8 “daily”), subjective probability for emotional support (from 1 “certainly not” to 4 “certainly yes”) and semantic differential for trust (from 1 “complete distrust” to 7 “complete trust”). Thus, all networks are valued. 1-valued links are considered non-existent when drawing the network.

The initial sample size was 189 PhD students and their supervisors. The response rate was 62% for students, 54% for supervisors and 34% for the student-supervisor dyads. The final sample size was 64 student-supervisor pairs. After carrying out an exploratory data analysis, 60 duocentered networks were finally analyzed.

Scientific advice and emotional support networks are directed networks with incoming relationships because we did not measure if the PhD student and the supervisor were giving the same advice and support, which they were receiving. The trust network is also directed but outgoing because we did not measure if alters trusted the egos. Finally, we consider the collaboration network as undirected, because the relation of working together should be mutual.

Using this information, we were able to compute the centrality, density, diversity and specific tailor-made measures for the duocentered networks separately for the four relations.

Performance was measured mainly by academic publications and conference presentations. PhD students were asked to recall the number of the following research outputs they had authored or co-authored during the past 3 years:

2. Publication without peer reviewing, working paper and conference presentation.

The measure of performance was computed by assigning two points to the first type of output and one point to the second.

The next step is to specify a set of regression models in order to predict PhD students’ academic performance from the duocentered network measures. The significance tests are not used because emphasis is placed on the predictive power of each network previously defined. After that, a predictive power comparison is done with the egocentered networks.

4.2. Models for predicting performance

We specified three different linear regression models for each relation (scientific advice, collaboration, emotional support, and trust) to estimate the influence of duocentered network measures on PhD students’ academic performance. These relations have basically four dimensions \((a, b, c, d = e)\), thus using a larger number of measures will lead to perfect collinearity. It is important to note that the qualitative variable field of study will be used in all models in order to account for field heterogeneity. We made an aggregation of four fields of study, namely sciences, technical studies, arts and social sciences. The three models are:

- **Model 1**: The first model uses some of the specific tailor-made measures created for the duocentered network. The model focuses on the \( a, c \) and \( d \) direct contacts for \( Ego_A \) (PhD student) and on non contacts for \( Ego_A \) which are contacts of \( Ego_B \) (supervisor) centered by the mean intensity of \( Ego_B \)’s contacts \( d(b - \bar{b}) \). The hypothesis for this model is that direct contacts have an influence on PhD students’ academic performance but also supervisor’s contacts are influential if a rather strong tie between the PhD student and the supervisor exists. The model for the simpler undirected case is specified as follows:

\[
Y = f(a, c, d, (b - \bar{b}), F) + U
\]

where \( Y \) is the performance measure, \( F \) stands for field of study and \( U \) is the disturbance term.

- **Model 2**: According to this model, PhD students’ academic performance depends on key characteristics of duocentered networks, which are the relative measures of density and diversity (degree centrality is used) and size. Instead of the degree centrality, closeness centrality could be used when the network is fully connected, which is likely to happen with undirected relationships. Besides, field of study is also included in the model. As argued before, degree or closeness centrality measures are not needed because diversity already provides this information. The specification of the second model for the simpler undirected case is

\[
Y = f(C_D(p_A) + C_D(p_B), C_D(p_A) - C_D(p_B), n, F) + U
\]

We can interpret this in the following way: when we sum centralities (density) we consider all contacts between egos and alters in the network. When we use the difference of centralities (relative diversity), we consider the difference between the networks of \( Ego_A \) and \( Ego_B \). This model construction has the attractive feature that the sum and the difference will tend to have low collinearity.

- **Model 3**: The third model is very similar to model 2, even regarding interpretation, but using the absolute density and diversity measures instead of the relative measures and size. The theoretical foundation of the model is the same as of the previous one. The difference lays mainly in its greater parsimony. The model can be specified as follows for the undirected case:

\[
Y = (C_D(p_A) + C_D(p_B), C_D(p_A) - C_D(p_B), F) + U
\]

4.3. Results

The regression model results regarding the prediction of PhD students’ academic performance are shown in Table 2. It shows the adjusted \( R^2 \) (the first row in each model) and the standardized regression coefficient for each variable in each model.
Table 2
Adjusted $R^2$ (bold if larger than 0.1) and standardized regression coefficients (bold if larger than $\sqrt{0.1}$) for duocentered networks

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Scientific advice relation</th>
<th>Collaboration relation</th>
<th>Emotional support relation</th>
<th>Trust relation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.094</td>
<td>0.170</td>
<td>0.163</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>$a$</td>
<td>0.028</td>
<td>0.172</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>$c$</td>
<td>0.153</td>
<td>0.274</td>
<td>0.415</td>
</tr>
<tr>
<td></td>
<td>$d$</td>
<td>0.116</td>
<td>0.105</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>$(b - b)$</td>
<td>-0.156</td>
<td>0.198</td>
<td>0.095</td>
</tr>
<tr>
<td>Model 2</td>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.063</td>
<td>0.133</td>
<td>0.194</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>Density</td>
<td>0.181</td>
<td>0.179</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>0.088</td>
<td>0.381</td>
<td>0.389</td>
</tr>
<tr>
<td></td>
<td>Relative diversity</td>
<td>0.155</td>
<td>0.158</td>
<td>0.097</td>
</tr>
<tr>
<td>Model 3</td>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.069</td>
<td>0.184</td>
<td>0.192</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td>Absolute density</td>
<td>0.129</td>
<td>0.477</td>
<td>0.419</td>
</tr>
<tr>
<td></td>
<td>Absolute diversity</td>
<td>0.171</td>
<td>0.282</td>
<td>0.116</td>
</tr>
</tbody>
</table>

For simplicity, field of study is omitted from the table because it is only a confounding variable whose effect we wanted to control.

The first model has a substantial adjusted $R^2$ for all networks except the scientific advice relation and the main predictor is $c$ (shared contacts) and even more so for the emotional support relation. Nearly all coefficients have positive signs as expected. Indirect contacts through the supervisor lack substantial predictive power in all networks. The contact with the supervisor was also non predictive but this may be due to the fact that this contact is present and consistently strong in 90% of all networks.

The second model also has a substantial adjusted $R^2$ for all networks except scientific advice. Network size is the main predictive variable for all three relations. Since by definition a duocentered network contains no isolated alters, size by itself is a good summary of the number of contacts within the network and is highly correlated with $c$, the main predictor in the previous model. As expected the sign of the coefficients is consistently positive.

The third model also has a substantial adjusted $R^2$ for the collaboration, emotional support and trust relations. Absolute density is a highly predictive variable for all these three relations.

For comparative purposes, the same regression models are now estimated for the egocentered networks of the PhD students, obviously using the more limited set of measures that can be computed from the egocentered networks. Some of the measures are then adapted. For instance, $a$ and $c$ are undistinguishable at the egocentered level and are thus added together, while degree is used instead of density. Table 3 shows a similar pattern as Table 2. Emotional support, collaboration and trust are the most influential networks. Model 3 is both the most parsimonious and that which best predicts performance. What is most relevant to

Table 3
Adjusted $R^2$ (bold if larger than 0.1) and standardized regression coefficients (bold if larger than $\sqrt{0.1}$) for egocentered networks of PhD students

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Scientific advice relation</th>
<th>Collaboration relation</th>
<th>Emotional support relation</th>
<th>Trust relation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.096</td>
<td>0.149</td>
<td>0.182</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>$a + c$</td>
<td>0.091</td>
<td>0.265</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>$d$</td>
<td>0.184</td>
<td>0.146</td>
<td>0.059</td>
</tr>
<tr>
<td>Model 2</td>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.074</td>
<td>0.130</td>
<td>0.169</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>Relative degree</td>
<td>0.135</td>
<td>0.113</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>0.140</td>
<td>0.291</td>
<td>0.274</td>
</tr>
<tr>
<td>Model 3</td>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.084</td>
<td>0.153</td>
<td>0.196</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>Absolute degree</td>
<td>0.147</td>
<td>0.303</td>
<td>0.367</td>
</tr>
</tbody>
</table>

both central actors to the alters in the duocentered network is a good predictor of performance. This measure is also highly related to $c$ and size. As expected the sign of the coefficients is consistently positive.

If we look at the table as a whole, we see that in general model 3 has the highest predictive power and moreover retains the advantage of being the most parsimonious. We can also see that the relation which least predicts performance with the three models is the scientific advice relation, while emotional support, collaboration and trust have a substantial adjusted $R^2$ for all models. As the literature suggests (Bondonio, 1998; Bartus, 2000), informal networks of support and trust are also important to work performance, not only task-related networks.
the purpose of this article is that adjusted $R^2$ are generally lower than those of Table 2.

Finally, we attempted to combine the duocentered measures for different models and networks into a single model to predict the student’s performance. A first specification was based on model 3 and contained absolute density for all four networks. After performing several specification searches, the model improved when dropping absolute density for trust (because of high collinearity with the same measure for emotional support) and $d$ (advice from supervisor) was substituted for the advice density. Results are shown in Table 4 and suggest that performance can be predicted from the total collaboration and emotional support contacts in the duocentered network (which conceptually would point at group cohesion), and also from advice from the supervisor. An important role of supervisor advice on student’s performance was of course to be expected and again supports the centrality of both actors to the research problem and the usefulness of duocentered networks.

5. Conclusions

The aim of this article was first to define the duocentered network structure. The key characteristic of this network is that it is based on a pair of egos and the relationships between these two egos and alters, but leaving out the relations among alters. Next, we adapted some social network measures of complete networks such as degree, closeness centrality, density and diversity to duocentered networks. Furthermore, we designed specific tailor-made measures. We next applied these duocentered network measures computed for four types of relations (scientific advice, collaboration, emotional support and trust) in order to predict the research performance of PhD students. Measures related to the total intensity of contacts (e.g., size and absolute density) seemed to work particularly well in predicting academic performance and led to very parsimonious regression models with nearly no collinearity. A final model included absolute density for the collaboration and emotional support relations and the direct contact of the PhD student with the supervisor for the advice relation. Duocentered networks also predicted performance better than egocentered networks.

We have, thus, presented duocentered networks as an appropriate network structure in dealing with research problems where a central pair of individuals is involved (e.g., PhD student and his/her supervisor or seller and buyer) and the closer neighborhood of this pair is of interest to the problem at hand. While the complete network structure provides us with more accurate information, when the complete network is unavailable due to high costs, low accessibility, or high unit non-response, the duocentered network still enables us to define network measures which are interpretable, which have predictive power, which are easy to compute and which are richer than those that would be obtained from egocentered networks alone. In our case PhD student and supervisor, definitely constitute a pair of central actors relevant to the problem of predicting PhD student performance.

Even if duocentered networks are arguably less prone to missing data problems than complete networks, some of these problems of course remain, notably those due to boundary specification and respondent inaccuracy. Further research is needed along the path shown by Costenbader and Valente (2003) in order to assess the stability of the duocentered network measures presented in the paper, to the presence of missing data.

The same principles outlined in this article could be used to define appropriate network measures if a triplet, a quartet and so on is central to a given research problem. As in this article, the number of contacts that are absent by definition should be taken into account when defining density. Another common feature would be that the measure of betweenness centrality would not be possible to adapt.

A final extension of the duocentered network concept could be the inclusion of the ties among alters as well. These contacts could be reported by the alters themselves or the egos could act as proxies. These networks would share nearly all properties of complete networks and no specific measures would need to be developed. Measurement cost or missing data problems would likely increase.

Finally, we are aware that the relative merits of duocentered and egocentered networks should be further explored in a variety of settings. The predictive power of duocentered networks should also be compared with that of complete networks. If they can be measured with a reasonable quality, complete networks would of course be expected to perform better. To begin with, there are measures (e.g., betweenness centrality) that only make sense in complete networks.

Acknowledgements

This project was partly supported by the University of Girona grants GRHCS66, BR00/UdG and 4E200304 and by the Slovenian Research Agency, Project P5-0168. It is a partial result of a wider project carried out by 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: 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 an earlier version of this article and made crucial suggestions.

References