PhD Students' Research Group Social Capital in Two Countries: A Clustering Approach with Duocentred Network Measures

Lluís Coromina[1], Germà Coenders[2],
Anuška Ferligoj[3], and Jaume Guia[4]

Abstract

The article examines the structure of the collaboration networks of research groups where Slovenian and Spanish PhD students are pursuing their doctorate. The units of analysis are student–supervisor dyads. We use duocentred networks, a novel network structure appropriate for networks which are centred around a dyad. A cluster analysis reveals three typical clusters of research groups. Those which are large and belong to several institutions are labelled under a bridging social capital label. Those which are small, centred in a single institution but have high cohesion are labelled as bonding social capital. Those which are small and with low cohesion are called weak social capital groups. Academic performance of both PhD students and supervisors are highest in bridging groups and lowest in weak groups. Other variables are also found to differ according to the type of research group. At the end, some recommendations regarding academic and research policy are drawn.

1 Introduction

In our society it is extremely important to produce quality in any professional sector. At the highest level of education, which is the PhD level, academic quality should be given strong emphasis if the society is interested in higher quality researchers at both universities and industry. Much research shows that PhD programmes are ill adapted to the changing and increasing requirements that future

[1] Department of Economics. Faculty of Economics and Business. Campus Montilivi, 17071 Girona, Spain; lluis.coromina@udg.edu
[2] Department of Economics. Faculty of Economics and Business. Campus Montilivi, 17071 Girona, Spain; germa.coenders@udg.edu
[3] Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia; anuska.ferligoj@fdv.uni-lj.si
[4] Department of Organization, Management and Product Design. Faculty of Tourism. Pujada dels Alemanys, 4. 17071 Girona, Spain; jaume.guia@udg.edu

PhDs will face (see Austin, 2002 and references therein). A key point for the academic quality of PhD programmes is that future PhDs achieve high academic performance. In the long run PhDs' performance is evaluated by the broader scientific community by means of attended conferences and published papers. In this article we consider performance from this view point in a way similar to Green and Bauer (1995). Knowledge about the elements that influence performance is also relevant for research groups at universities in order to select the best PhD students and to promote working conditions that foster their performance.

Research groups consist of several experts from different areas who, besides researching and teaching, also perform the role of supervising PhD students. Doctoral theses are written in close interrelationship with the research agenda of the group and under the supervision of one of its senior researchers.

Social ties and networks which emerge in research settings are very important for the performance of researchers since they enable access to knowledge and experiences possessed by other researchers within the group as well as information on where to go outside the primary research group to obtain help when specific problems emerge. They help in establishing contacts with key professionals in the field and provide researchers with social support and positive evaluations, which is especially important in the case of young researchers and PhD students (Ziherl et al. 2006).

The role of social relationships within groups, including trust and communication among social network members has been well established in the literature (Wasserman and Faust, 1994). The basic idea behind this perspective is that an individual's success strongly depends on the relationships the individual has with relevant others inside and outside the organisation (Burt, 2000). The importance of social relations in the network structure concerning individual performance can be captured by the concept of social capital. The key points are the relationships between students and supervisor (Cryer, 1996), the relationships within the research groups (Hemlin et al., 2004), and socialization (Austin, 2002). On the other hand, being isolated in a research group can be one of the main problems for a PhD student (Rudd, 1984).

Ziherl et al., (2006) analyzed whole cooperation network data of PhD students' research groups using a clustering approach. They were able to find meaningful types of network structures and interpreted them in terms of PhD students' social capital and PhD students' performance. The authors found three clusters that could be interpreted as bridging, bonding, and weak social capital.

The bridging social capital cluster consisted of research groups which were large and diverse. They included researchers from different institutions. In addition to being exposed to very diverse internal environment, PhD students also maintained many cooperation ties with people outside the group. The average strength of ties between PhD students and other members of the group was

moderate to low. The network structures showed structural holes which mean that researchers were often brokers between unconnected parts of their networks.

In the bonding social capital cluster the authors found research groups which were very small but with well developed cooperation. The average strength of ties between PhD students and members of their research group was the highest. The overall cohesion of the research group was also the highest. The members of the research group generally came from the same institution.

The weak social capital cluster included small research groups where PhD students and other researchers only rarely cooperated with one another or with researchers outside the group.

In this article we analyze the same data as Capó et al. (2007). These authors used the data to predict PhD students' academic performance from their background, attitudes, supervisors' performance and research group characteristics, research groups being treated as egocentred networks. Capó et al. found that variables related to the research-group network had a negligible explanatory power on student performance once the remaining variables were accounted for, the most important among them being supervisors' performance. However, such network variables are still useful to show different profiles of research-group structure, which may be related to a number of research-group variables, not only to students' performance. The aim of this article is similar to that of Ziherl et al. (2006), namely to study the most typical profiles of structure and composition of the research groups in which PhD students and their supervisors are involved.

As Capó et al. (2007), we use data of PhD student research groups of the University of Girona (Spain) and all universities and research institutes in Slovenia. Whole research-group networks were not at all measured in Girona. In a sizeable number of research groups in Slovenia, whole networks had too many missing values to be usable. Therefore, we opted for a clustering approach with duocentred networks. This novel network structure (Coromina et al., 2008) provides richer network information than egocentred networks, and can be used as a compromise when whole networks are not available, which is our case. Duocentred networks can be used when we find a pair of relevant central actors in a network, which in our case are the PhD student and his/her supervisor. The main characteristic of duocentred networks is that they are composed of a pair of central egos and their relationships with alters, while the ties amongst alters are neither observed nor needed.

In this article, some network measures obtained from duocentred networks are selected in order to be as related as possible to the ones used in Ziherl et al., (2006): strength, cohesion, supervisor frequency of contact with the PhD student, network size and number of different institutions the members of the group belong to. As is the case in Ziherl et al., (2006), these measures are obtained from the network frequency of contact for scientific collaboration and are used for

clustering and uncovering different types of PhD students' research-group organization.

2   Study design

The study reported in this article uses part of the data of a larger project carried out by the INSOC (International Network on Social Capital and Performance) research group, in which researchers of the universities of Ghent (Belgium), Ljubljana (Slovenia), Girona (Spain), and Giessen (Germany) take part. The aim of the INSOC project is to develop models predicting the PhD students' academic performance.

Data for the INSOC project in Girona and Slovenia were obtained by a means of web surveys of PhD students, their supervisors, and in Slovenia also research-group members. The web questionnaires contained a large range of background, attitudinal and social network variables. See Coenders et al. (2007) for details.

Web questionnaires simplify the administration of some complex questions (Tourangeau and Yan, 2007). For example, the software can retain the names of network members given in previous answers and supply them into later questions. Moreover, web questionnaires are self-administered and thus improve data quality for sensitive questions such as those dealing with personal relationships (Dillman, 2007). The reliability and validity of the used web network questions were reported to be high by Coromina and Coenders (2006).

The relevant population is composed of the PhD students who began their doctoral studies at the University of Girona (Spain) and at different universities and research institutes in Slovenia in the academic years 1999/2000 and 2000/2001. We selected only PhD students employed at their universities. This choice was made because these PhD students have frequent contact with other researchers and somehow formally belong to a research group. Most of them had grants (all of them in Slovenia, under the young-researcher programme), the rest being assistant professors or research assistants hired for particular research projects at the University of Girona. Because of the relatively small population size of the PhD students (194 in Slovenia and 86 in Girona), we decided to study the complete populations.

The members of the research group to which the student and the supervisor belonged to had to be identified. In the first phase we asked name generator questions to the PhD students' supervisors in order to obtain the names of people in their research group connected to the research topic of their PhD students:

1. Please name (name and surname) all doctoral students and teaching assistants, whose research work is currently under your supervision.
2. Please name (name and surname) all researchers (not named so far), whose formal supervisor you are and who participate in at least one research project in which you also participate.

3. Please name (name and surname) your colleague professors, researchers and people from private sector, who you co-operate with in research projects in which the doctoral student in question also participates.

The web questionnaires, which were later administered, were personalized and included the names of all research-group members. In both Girona and Slovenia, respondents first received an invitation letter in an official university envelope. Next, personalized e-mail invitations were sent to all respondents with a direct link to their own web questionnaire address. The questionnaires resided in a server at the University of Ghent, and were programmed using the Snap software, Version 7 (Mercator Research Group, 2003). Data collection took place during the last months of 2003 and first months of 2004.

A commonly mentioned threat to web surveys is low response rate. A follow-up design is one of the most efficient techniques to reduce the non-response rate (Kaplowitz et al., 2004). The use of mixed-mode follow-ups increases the response rate for those who are more sensitive to specific modes (De Lange, 2005). At the University of Girona three reminders were carried out by e-mail, letter and phone. In Slovenia two reminders were sent by e-mail. Final response rates are displayed in Table 1.

Table 1: Response rates for PhD students and supervisors of the web survey.

	Response rate PhD Students	Response rate Supervisors	% complete Student–supervisor pairs
Girona	78%	75%	63%
Slovenia	60%	54%	36%

The web questionnaire design was a complex process led by Daniëlle de Lange and involved two years of discussion within the INSOC research group, several international meetings, several focus groups and pre-tests (De Lange, 2005). The fact that we had to produce comparable versions in four languages (Catalan, Flemish, Slovenian, and German) and the differences between the three university systems lengthened the process even further and involved two independent translations, a pre-test of the translated questionnaires and further discussions and modifications. Two different questionnaires were designed, one for PhD students and another for their supervisors and other research-group members, though most of the questions were asked of all.

3 Duocentred network structure

The most typical network structures found in the literature are whole and egocentred networks. The former is used 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 to the other members of the research group is relevant. In some cases, a pair of actors may be central (e.g., husband and wife, buyer and seller, president and prime minister) and we may intend to study the behaviour, performance, or social capital of these two specific actors in the network. In such cases egocentred networks are difficult to interpret because only one ego is considered while this ego has an especially relevant connection with another actor, which is neglected.

In our case we know that one central ego (PhD student) has an especially relevant relationship with another (his/her supervisor). In our study the pair of egos is thus composed of the PhD student and his/her supervisor. These supervisors usually have some important contacts with alters in the network and these relationships can influence the social capital of the PhD students even if they do not belong to the students' egocentred networks.

When there is a pair of relevant central actors in a network Coromina et al., (2008) suggested a new structure called duocentred network. The main characteristic of a duocentred network is that it is built around a pair of central egos. Network information is obtained from these two egos but 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 nor taken into consideration. This means that the pair of central egos (from now on we denote them as Ego_A –the PhD student–, and Ego_B –the supervisor) provide us information about their mutual relationship and their ties to their alters in the network, but not about relationships among alters. An example of duocentred network is shown in Figure 1.
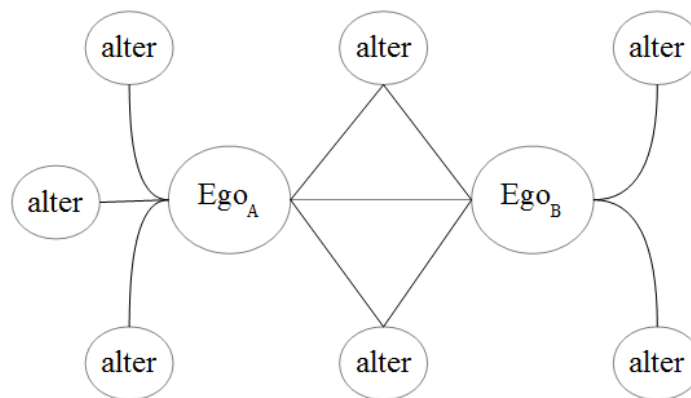

Figure 1: Example of duocentred network around Ego_A and Ego_B.

The following properties of duocentred networks should be considered:
1. Two main actors (Ego_A and Ego_B) have to be clearly central and are considered as egos.
2. Actors who are not defined as Ego_A or Ego_B are called alters.
3. No relationships are observed among alters.

4. Alters who do not have any contact with the egos are considered as isolates. These isolate members are not considered as a part of the duocentred network, so they do not appear in the network.

Three types of alters who belong to the network can be observed in Figure 1 depending on whether they are linked to Ego_A, Ego_B or to both of them.

A few more words have to be told about the relative merits of whole, egocentred and duocentred networks. When possible, researchers search for whole networks, which encompass all the information of egocentred and duocentred networks and are, thus, the richest structure in which to compute social network measures, since they permit the analysis of the relationships among all actors. However, this ideal situation is not easy to reach due to the fact that all actors in the network have to be contacted and must give the information with regard to their relationships with all others. Two major problems for whole networks are thus data collection cost and the presence of missing data. For obvious reasons, these problems are not as much present in duocentred networks, which is a further argument for using this type of networks. Of course, these problems are also reduced in egocentred networks. Duocentred networks can in fact be better understood as a generalization of egocentred networks than as a simplification of complete networks. However, duocentred networks make it possible to compute a larger array of network measures than egocentred networks and some of these measures are closer to the complete network equivalents (Coromina et al., 2008).

4 Duocentred network measures

In this article we use the collaboration network ties because we are interested in the research cooperation between network members. The specific question asked to the PhD student and the supervisor about collaboration is shown in Figure 2. The frequency question was coded in a 1 to 8 scale from "not in the past year" to "daily".

	Not in the past year	Once in the past year	Several times a year	About monthly	Several times a month	Weekly	Several times a week	Daily
Name 1	○	○	○	○	○	○	○	○
Name 2	○	○	○	○	○	○	○	○
Name 3	○	○	○	○	○	○	○	○
Name 4	○	○	○	○	○	○	○	○

Figure 2: Question for collaboration within the research group.

Student and supervisor got the list of research-group members that had previously been obtained through name generator questions to the supervisor (Name 1, Name 2, ... in Figure 2). The next question in the questionnaire was about the existence or not of collaboration relationships external to the research group (Figure 3). If the answer was yes, PhD students and supervisors were asked to provide the names of these people and to rate their collaboration frequency using the same response scale as before. From a methodological point of view, the possibility for both egos to provide additional lists of network members enriches the duocentred network structure as it is a major potential source of alters that are linked to only one of the egos. From a substantive point of view, it makes the results refer not only to different types of research-group structure but also to different ways in which the research group is inserted into the wider scientific community.

Think about all the situations in the past year that required collaboration with other people concerning research (namely since 1 November 2002). Did you collaborate with anyone in the last year besides the people in the abovementioned list? [people from outside the university and/or from abroad can also be mentioned]

    ⦿   Yes

    ○   No

Please fill in the full name of the people besides those in the list with whom you collaborated concerning research in the past year (namely since 1 November 2002)?

Name 1 |_____|

Name 2 |_____|

Name 3 |_____|

Figure 3: Question for collaboration beyond the research group.

Duocentred network measures were computed from these data.. Some of the measures, defined by Coromina et al. (2008) were adapted from the existing measures for whole networks. The authors also defined some tailor-made measures for specific research questions. The purpose of this article is not to review all duocentred network measures suggested by Coromina et al., but only to select those that fit the best into the conceptual framework set up by Ziherl et al. (2006).

A key concept which is included in many network measures is degree centrality, which indicates how well an actor is connected within the network (Wasserman and Faust, 1994). This measure focuses only on direct or adjacent contacts; the more contacts an ego has, the more central the ego is. Most of classical social network studies use this measure (Bonacich, 1987; Everett and Borgatti, 1999; Faust and Wasserman, 1992; Freeman, 1979; Freeman, Roeder, and Mulholland, 1980).

For undirected networks, like the collaboration network used in this study, a general measure of centrality can be obtained for Ego_A and Ego_B. Nieminen's (1974) degree takes into account the adjacencies for an actor p_k:

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

(4.1)

where:

- $C_D(p_k)$ is the centrality of actor k (in the duocentred case Ego_A or Ego_B).
- $t(p_i, p_k)$ is the tie between p_i and p_k (from 1 "not in the past year" to 8 "daily").

- n is the duocentred network size including both egos and all their alters.

This basic concept is relevant in order to understand the duocentred measures we are going to use as variables in the cluster analysis. These variables are explained below:

1. *Strength*: For whole networks, Ziherl et al. (2006) used a measure of what Granovetter (1985) calls tie strength, which is based on the average frequency of collaboration contacts between the PhD student and all members of the network, which can be computed as centrality related to network size. This is what Coromina et al. (2008) call relative centrality of Ego_A (PhD student):

$$C'_D(p_A) = \frac{\sum\limits_{i=1}^{n-1} t(p_i, p_A)}{n-1}$$

(4.2)

2. *Cohesion*: For whole networks, Ziherl et al. (2006) used a measure of what Coleman (1988) calls cohesion, which was measured by the average frequency over all contacts in the network. Cohesion can have a profound impact on the social capital of the network as a whole, by means of shared norms and reputation. We can compute it as the mean of the strength measures for both Ego_A and Ego_B, which is closely related to the measures of density defined in Coromina et al. (2008). If cohesion is larger than strength, then the supervisor has a higher contact average than the student. If cohesion is lower than strength it is the other way around.

3. *Frequency of contact between the PhD student and the supervisor*: This measure was not used by Ziherl et al. (2006). We include it here because it is a very meaningful measure for the specific duocentred case (Coromina et al., 2008), as the PhD student and the supervisor are the two actors around whom the duocentred network is built. This variable refers to the frequency a PhD student and his/her supervisor collaborate together in research. Since the question was asked to both the PhD student and the supervisor and the network is conceptually undirected, we used the average of both responses in order to obtain the final score.

4. *Diversity*: Drawing from Burt (1983), Ziherl et al. (2006) defined diversity of knowledge possessed by the members of the group and measured it through the number and variety of types of actors in the network. We adapt two of the measures used by Ziherl et al. to the duocentred case. (a) *Network size* is the total number of actors in the duocentred network structure. It counts the alters and the two central egos (which coincides with n in (2)). Research-group members that do not have either contact with the PhD student or with the supervisor are called isolated members and they are not considered part of the network. Contacts were considered to exist if respondents selected a frequency of "2: once in the past year" or larger. Size in combination with strength can be used to identify students with a large number of weak ties (Granovetter, 1973) as opposed to students with a small number of strong ties. (b) The *Number of different institutions where research-group members belong* was unavailable for the whole duocentred network and refers only to the research group as defined by the PhD supervisor using the name generator questions presented in Section 2. Social capital theories emphasize the importance of having different types of contacts.

5   Cluster analysis results

For clustering purposes we use the variables strength, cohesion, frequency of contact between the PhD student and the supervisor, network size, and number of different institutions. The data are the complete student–supervisor pairs in the universities of Girona and Slovenia, which were also used by Capó et al. (2007).

After removing two outliers with extreme Cook's distance in the models of Capó et al. (2007), the listwise sample size was 111 complete student–supervisor dyads, 53 from Girona and 58 from Slovenia. The Slovenian sample size is larger than that available to Ziherl et al. (2006), because only complete dyads are required to respond to the questionnaire for the duocentred analysis, not whole research groups. The Pearson correlations among the duocentred network measures are presented in Table 2. The two measures of diversity seem to be positively correlated. Tie strength, cohesion, and student-supervisor contact are also positively correlated. The first group of variables is negatively correlated with the second one, thus showing that *weak ties tend to occur together with high diversity*.

We used Ward's method on the standardized variables (see Everitt et al. 2001). The number of clusters was selected according to the following criteria:

1. No eta coefficient measuring the relationship between the cluster solution and each of the variables should be lower than 0.5.
2. No cluster should contain fewer than 20 cases.

3. The two parts of the cluster that would be subdivided if the number of clusters would be increased by one should not have meaningfully different interpretations.

Table 2: Pearson correlations among the considered variables.

	Student tie strength	Network cohesion	Student Supervisor contact	Network size	Number of different Institutions
Student tie strength	1	.755	.398	-.463	-.162
Network cohesion	.755	1	.567	-.442	-.240
Student supervisor contact	.398	.567	1	-.116	-.073
Network size	-.463	-.442	-.116	1	.307
Number of different institutions	-.162	-.240	-.073	.307	1

As in the case of Ziherl et al. (2006), the three-cluster solution was finally selected. When repeating the three-cluster classification with the *K*-means method, 86% of the cases were identically classified, which argues for a high degree of stability of the obtained solution. The characteristics of these three clusters are shown in Table 3. The labels 1 to 8 of the contact-frequency categories (Section 4) can be used to interpret the means of the first three variables in Table 3.

Table 3: Cluster sizes, means and eta coefficients for considered variables.

	eta	Cluster			Total sample
		Bridging	Bonding	Weak	
Cluster size		29 (26%)	42 (38%)	40 (36%)	111 (100%)
Student tie strength	0.69	1.7	3.2	1.6	2.2
Network cohesion	0.71	2.6	3.8	2.6	3.1
Student supervisor contact	0.54	5.1	6.4	4.5	5.4
Network size	0.64	17.5	9.0	13.3	12.8
Number of different institutions	0.67	4.6	2.4	2.0	2.8

The first cluster (26% of the sample) is composed of student-supervisor pairs in large research groups with members of many different institutions. The collaboration frequencies within the group (cohesion) are slightly below average. The PhD student's average scientific collaboration frequency with other members (tie strength) and with the supervisor are also below average. This cluster corresponds to what Ziherl et al. (2006) call *bridging social capital*.

The second cluster (38% of the sample) is composed of student-supervisor pairs in smaller than average research groups with members of only a few different institutions. The collaboration frequencies within the group (cohesion) are above average. The PhD student's average collaboration frequency with other members

(tie strength) and with the supervisor are also above average. This cluster corresponds to what Ziherl et al. (2006) call *bonding social capital*.

The third cluster (36% of the sample) is composed of student-supervisor pairs in research groups with all variables below or around the average values. This cluster corresponds to what Ziherl et al. (2006) call *weak social capital*.

In all three clusters cohesion is higher than strength, which means that in all clusters supervisors tend to have more frequent average contacts with the network members than their students. This can also be the result of the supervisors having a longer list of external contacts, since the frequency of contact between the student and an alter who is only connected to the supervisor is necessarily zero.

6 Interpretation of the clusters using external variables

Once we have obtained a three-cluster solution (bridging, bonding and weak social capital) we are also interested in analyzing the behaviour of some external variables across clusters. This can help us interpret the different types of social capital of the PhD student and his/her supervisor. These variables are:

1. Performance of both PhD students and their supervisors, measured as a weighted combination of academic publications and conference papers, both published and accepted. Any publication which is either in an international or peer−reviewed medium is assigned a weight equal to two. All other publications and conference presentations and posters are assigned a unit weight.

2. A set of attitudinal variables described in Coromina (2006) having to do with the PhD student's perception of the relationships with the supervisor and within the group as a whole (group atmosphere, too close supervision by supervisor, guidance of supervisor during PhD, promotion of students' contacts by supervisor, and integration of the PhD within the research-group tradition).

3. A set of the attitudinal variables related to the PhD student's attitudes towards several aspects of their job (attitude towards publishing, perceived loneliness of the PhD student's job, and job satisfaction). Attitudinal variables were operationalized by means of summated rating scales of the items displayed in Table 4, all of which had substantial loadings on their dimension in both factor analyses carried out for Slovenian and Spanish data separately.

4. PhD student's and supervisor's age and gender.

5. Field of study, grouped into sciences, technical studies, humanities and social sciences (for the creation of this classification and its comparability across countries see Capó, 2009:49-60).

Table 4: List of summated scales for the attitudinal variables and their items.

Group atmosphere (semantic differential scale)
Distrust-trust
Unpleasant-pleasant
Unfriendly-friendly
Unproductive-productive
Not helpful-helpful
Too close supervision by supervisor
My supervisor gives me enough freedom on the content of my PhD
My supervisor imposes his own opinion all too often
Guidance of supervisor during PhD
My supervisor gives advice concerning the development of my PhD project
My supervisor helps me prepare my publications
Promotion of students' contacts by supervisor
My supervisor introduces me to other researchers
My supervisor encourages me to take educational courses abroad
My supervisor encourages me to attend conferences
Integration of the PhD within the research-group tradition
My PhD concerns a (relatively) new issue in the research tradition of the research group
My PhD concerns a completely new research issue in my field of research
Attitude towards publishing
Publishing is stimulating and motivating
Publishing is an important means of getting feedback
I only publish because I'm supposed to
Publishing is annoying because it is very time-consuming
Publishing is useless
Loneliness of PhD student's job
Working on a PhD is a lonesome activity
During my PhD research I often feel as if I am alone on an island
I often exchange views with my colleagues about my PhD research
Job Satisfaction
I find real enjoyment in my work
I'm often bored with my job
Most of the time I have to force myself to go to work
I definitively dislike my work

6. Country and typology of PhD students, understood as their main activity at university. In Slovenia all PhD students have a grant under the young-researcher programme and research is their main job. An equivalent type of students is also present in Girona; however, in this university there are also PhD students without a grant that are hired by the university for teaching as their main activity. The amount of time devoted to the PhD and to publishing is likely to differ between these two groups.

Table 5 shows the relationships between the external variables and the clustering solution in the first column. The significant ones can be interpreted for each cluster. These variables are: performance of both PhD students and their supervisors, promotion of student's contacts by supervisor, integration of the PhD within the research-group tradition, country, and typology of PhD students.

Table 5: Description of the clusters with external variables (summated scale averages and categorical variable percentages).

	eta/V[1]	Bridging	Bonding	Weak	Total
PhD student performance	0.22*	22.4	15.3	14.5	16.6
Supervisor performance	0.24*	53.8	36.0	40.3	41.8
Group atmosphere	0.11	0.6	0.5	-1.0	0.0
Too close supervision by supervisor	0.13	-0.5	0.4	0.0	0.0
Guidance of supervisor during PhD	0.16	-0.6	-0.3	0.7	0.0
Promotion of students' contacts by supervisor	0.21*	1.2	0.5	-1.4	0.0
Integration of the PhD within the research -group tradition	0.27*	-1.4	-0.4	1.4	0.0
Attitudes towards publishing	0.16	1.3	0.0	-0.9	0.0
Loneliness of PhD student's job	0.19	-1.4	0.2	0.9	0.0
Job satisfaction	0.13	0.9	0.0	-0.6	0.0
Student age	0.19	29.4	28.8	30.9	29.7
Supervisor age	0.07	46.3	45.3	46.5	46.0
% female students	0.05	41%	38%	35%	38%
% female supervisors	0.19	38%	21%	18%	25%
% science field	0.18	48%	50%	53%	50%
% technical field	0.18	38%	33%	24%	31%
% humanities field	0.18	0%	12%	16%	10%
% social sciences field	0.18	14%	5%	8%	8%
% Girona students mainly doing research	0.21*	21%	38%	28%	30%
% Girona students mainly teaching	0.21*	7%	26%	18%	18%
% Slovenia students (all mainly doing research)	0.21*	72%	36%	55%	52%

[1]Measures of association between the given variable and the clustering solution: eta for numeric variables and Cramer's *V* for qualitative variables. Significant associations according to the standard ANOVA *F* test or to the χ^2 test (α=10%) marked with "*". Qualitative variables with more than 2 groups have a common *V* and significance.

The first cluster (bridging social capital) has, by far, the highest performance for both PhD students and supervisors. The promotion of PhD students' contacts by their supervisors is above average (values of the attitudinal variables are mean-centred; therefore the mean of these variables in the total sample is zero), while the integration of the PhD within the research-group tradition is below average, which means that PhD students are involved on a rather new research topic within the research group. This cluster contains the highest proportion of students with a grant mainly doing research and also the highest proportion of Slovene students.

In the second cluster (bonding social capital), supervisors publish below average. The promotion of contacts is above average and the integration of the PhD thesis in the group is slightly below average. The composition of this cluster

in terms of PhD students is opposite to the first, as the second contains the fewest Slovenians and the most students doing mainly teaching.

In the third cluster (weak social capital), PhD students publish below average. The promotion of student's contacts by the supervisor is largely below average and the integration of the PhD thesis in the group tradition is largely above average. The distribution of students regarding country and dedication to teaching and research is close to the distribution in the overall sample.

The remaining external variables make no difference across clusters. These are the attitudinal variables: group atmosphere, too close supervision by supervisor, guidance of supervisor during PhD, attitudes towards publishing, loneliness of PhD student's job and job satisfaction, and the background variables gender, age and field of study.

6   Discussion

The differences encountered among clusters both regarding clustering and external variables enable us to draw some conclusions regarding different ways in which PhD students and supervisors get inserted into the scientific community by means of a research group and of additional scientific collaboration contacts outside the group.

The first cluster (bridging social capital) is composed of student–supervisor dyads that have a large duocentred collaboration networks but low frequency of collaboration with the members, which tend to belong to different institutions. These students and their supervisors have the highest performance. A larger and diverse network seems to foster publications even where coupled with lower collaboration contact frequency. The fact that students have additional contacts due to the influence of their supervisor is highly related to these large networks from different institutions and may contribute to their research on a topic different from the most common in their research group. These characteristics are mainly found in the PhD students from Slovenia (72%) and in almost no PhD student who is mainly teaching in Girona (7%). The high performance of this cluster supports Granovetter's (1973) theory of weak ties. It also fits into Lin's theories (Lin, 1990; Lin et al., 1981), where social capital of individuals is defined as the amount of social resources they have, that is, the number of relationships, the density of the network and the heterogeneity of the contacts. The fact that the student has access to external contacts that may act as a link between different research groups and the fact that the supervisor is a source of these external contacts also comply with Burt's theory of structural holes (Burt, 1992).

The students who belong to the second cluster (bonding social capital) have a high frequency of collaboration with the other members, especially with their supervisors; however their duocentred network is relatively small. As regards the

external variables, the performance of these students and their supervisors is relatively low, and the supervisors are not facilitating contacts to the student, which we could call "sleeping social capital". This cluster basically contains PhD students in Girona who are mainly teaching and for whom publishing may not be the priority goal. This structure with high frequency of collaboration inside a group and very low contacts outside the group results in the topic of the PhD thesis being quite similar to the tradition in the research group. This behaviour is reported in the social capital literature as "network closure" (Coleman, 1990).

The students and the supervisors belonging to the third cluster (weak social capital) have an average network size including relatively few members from different institutions. The frequency of collaboration with the supervisor and with the other network members is about as low as in the bridging cluster, while the number of different institutions is about as low as in the bonding cluster. The contacts that are facilitated by supervisors are perceived as small. The network is thus neither cohesive nor diverse. Besides, the contact between the two main actors in the duocentred network is not frequent. PhD students in this cluster arguably have the lowest social capital, and, in any case, the lowest performance.

Capó et al. (2007) showed that the effect of network variables on PhD student performance is low and counterintuitive, once students' background, students' attitudes and supervisor performance are accounted for. At the bivariate level, network clusters do appear to be correlated to both student and supervisor performance. There are at least two theoretical arguments which can reconcile the apparent contradiction of the obtained results.

It can be argued that the longer careers of supervisors enable them to take greater advantage of their networks than is possible for students. Thus, the network effect on student performance can be argued to be at least partly an indirect effect, via supervisors' performance.

It can also be argued that the obtained clusters make non-linear relationships between networks and performance emerge, which was not possible with the approach of Capó et al. (2007). For instance, Table 3 shows student tie strength, network cohesion, and supervisor contact increase performance when moving from the weak to the bridging cluster and decrease performance when moving from the bridging to the bonding cluster. Kogovšek et al. (2011) also reported analogous non-linear effects.

Even if the information obtained from duocentred networks is less detailed than that from whole networks, it still uncovered meaningful network structures which resembled the ones found by Ziherl et al. (2006), and did so at a much lower data collection cost.

Some conclusions can also be drawn with respect to the educational and the research policies. PhD students and grants should preferably be allocated to the research groups of the bridging type, which have the highest performance. Public funds could also be devoted to help research groups pay for bridging actions. Examples of the latter range from economic support for larger research networks

with external research groups to travel funds for PhD students who attend conferences abroad.

As regards to the limitations of the study, we are aware that the conclusions can be the result of singularities of the universities of Girona and Slovenia. The sample size was small, limited by the small population size. Finally, although publications are awarded an ever increasing importance by institutions evaluating academic performance, we are aware that measuring PhD students' performance only by their publication record is only one of many sensible alternatives.

Acknowledgement

References

[1] Austin, A.E. (2002): Preparing the next generation of faculty. *The Journal of Higher Education*, **73**, 94-122.

[2] Bonacich, P. (1987): Power and centrality: a family of measures. *American Journal of Sociology*, **92**, 1170-1182.

[3] Burt, R.S. (1983): *Applied Network Analysis*. Beverly Hills: Sage.

[4] Burt, R.S. (1992): *Structural Holes*. Cambridge, MA: Harvard University Press.

[5] Burt, R.S. (2000): The network structure of social capital. In R. Sutton and B. Staw (Eds.): *Research in Organizational Behavior*. Greenwich: JAI Press, 345-423

[6] Capó, A. (2009): *Predictors of Knowledge Creation Performance. A Quantitative Qualitative Comparative Study of European Doctorandi*. Unpublished Doctoral Dissertation, University of Girona, Spain. (5 October 2010)   http://www.tdr.cesca.es/TESIS_UdG/AVAILABLE/TDX-0720109-124200/taca1de1.pdf

[7] Capó, A., Coromina, L., Ferligoj, A., Matelič, U., and Coenders, G. (2007): Networks of PhD students and academic performance: a comparison across countries. *Metodološki zvezki*, **4**, 205-217.

[8] Coenders, G., Ferligoj, A., Coromina, L., and Capó, A. (2007): Design and evaluation of a Web survey for the social network data. In G. Loosveldt, M. Swygedouw and B. Cambré (Eds.): *Measuring Meaningful Data in Social Research*. Leuven: Acco, 233-255.

[9] Coleman, J.S. (1988): Social capital in the creation of human capital. *American Journal of Sociology*, **94**, 95-120.

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

[11] Coromina, L. (2006): *Social Networks and Performance in Knowledge Creation. An Application and a Methodological Proposal*. Unpublished Doctoral Dissertation, University of Girona, Spain. (5 October 2010). http://www.tdr.cesca.es/TESIS_UdG/AVAILABLE/TDX-0619106-141917//tlcs.pdf

[12] Coromina, L. and Coenders, G. (2006): Reliability and validity of egocentered network data collected via web. A meta-analysis of multilevel multitrait multimethod studies. *Social Networks*, **28**, 209-231.

[13] Coromina, L., Guia, J., Coenders, G., and Ferligoj, A. (2008): Duocentered networks. *Social Networks*, **30**, 49-59.

[14] Cryer, P. (1996): *The Research Student's Guide to Success*. Buckingham: Open University Press.

[15] Dillman, D.A. (2007): *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

[16] Everett, M.G. and Borgatti, S.P. (1999): The centrality of groups and classes. *Journal of Mathematical Sociology*, **23**, 181-201.

[17] Everitt, B., Landau, S., and Leese, M. (2001): *Cluster analysis 4th ed.* London: Arnold.

[18] Faust, K., and Wasserman, S. (1992): Centrality and prestige: a review and synthesis. *Journal of Quantitative Anthropology*, **4**, 23-78.

[19] Freeman, L.C. (1979): Centrality in social networks: conceptual clarification. *Social Networks*, **1**, 215-239.

[20] Freeman, L.C., Roeder, D., and Mulholland, R.R. (1980): Centrality in social networks: II. Experimental results. *Social Networks*, **2**, 119-141.

[21] Green, S.G., and Bauer, T.N. (1995): Supervisory mentoring by advisers: relationships with doctoral student potential, productivity and commitment. *Personnel Psychology*, **48**, 537-561.

[22] Granovetter, M. S. (1973): The strenght of weak ties. *American Journal of Sociology*, **78**, 1360-1380.

[23] Granovetter, M.S. (1985): Economic actions, social structure: The problem of embeddeness. *Jounal of Sociology*, **91**, 481-510.

[24] Hemlin, S., Allwood, C.M., and Martin, B.R. (2004): *Creative Knowledge Environments. The Influences on Creativity in Research and Innovation*. Northampton, MA: Edward Elgar.

[25] Kaplowitz, M., Hadlock, T., and Levine, R. (2004): A comparison of Web and mail survey response rates. *Public Opinion Quarterly*, **68**, 94-101.

[26] Kogovšek, T., Hlebec, V., and Ferligoj, A. (2011): From busy bees to science geeks and party animals: a typology of Slovenian doctoral students. *Metodološki Zvezki*, **8**, 121-136.

[27] De Lange, D. (2005): *How to Collect Complete Social Network Data? Nonresponse Prevention, Nonresponse Reduction and Nonresponse Management based on Proxy Information.* Unpublished Doctoral Dissertation. Ghent University, Belgium.

[28] Lin, N. (1990): Social resources and social mobility: a structural theory of status attainment. In R. L. Breiger (Ed.): *Social Mobility and Social Structure.* New York: Cambridge University Press, 247-271.

[29] Mercator Research Group (2003): *Snap Survey Software: Version 7.* Bristol: Snap Surveys Ltd.

[30] Lin, N., Ensel, W.M., and Vaugh, J.C. (1981): Social resources and strength of ties: structural factors in occupational status attainment. *American Sociological Review*, **46**, 393-405.

[31] Nieminen, J. (1974): On the centrality in a graph. *Scandinavian Journal of Psychology*, **15**, 322-336.

[32] Rudd, E. (1984): Research into postgraduate education. *Higher Education Research and Development,* **3**, 109-120.

[33] Tourangeau, R. and Yan (2007): Sensitive questions in surveys, *Psychological Bulletin*, **133**, 859-883.

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

[35] Ziherl, P., Iglič, H., and Ferligoj, A. (2006): Research groups' social capital: a clustering approach. *Metodološki Zvezki*, **3**, 217-237.