CrossMark

Factors influencing the formation of intra-institutional formal research groups: group prediction from collaboration, organisational, and topical networks

Hector G. Ceballos[1] · Sara E. Garza[2] · Francisco J. Cantu[1]

Abstract The factors that foster successful scientific collaboration and teamwork have been studied extensively. However, these factors have been studied in isolation and it is not clear to what extent one factor is more relevant than other in the formation of research groups. In this work we propose a new methodology based on network analysis to simultaneously evaluate multiple factors considered relevant in the conformation of formal research groups. Our methodology is supported on structural, statistical, and correlation analysis. In addition to validating our methodology with a case study at a research-teaching university, we introduce a new network to represent the success of scientific collaboration that produces the best prediction in group formation. Our methodology and the results obtained can be used for organising researchers in a university that seeks to strengthen its research strategy.

Keywords Scientific collaboration · Network analysis · Graph clustering · Research groups · Complex networks

Mathematics Subject Classification 62H30 · 91D30 · 90C27

JEL Classification C61 · C88

✉ Hector G. Ceballos
  ceballos@itesm.mx

  Sara E. Garza
  sara.garzavl@uanl.edu.mx

  Francisco J. Cantu
  fcantu@itesm.mx

[1] Tecnologico de Monterrey, Ave. Eugenio Garza Sada 2501, Monterrey, NL, Mexico

[2] School of Mechanical and Electrical Engineering (FIME), Universidad Autónoma de Nuevo León (UANL), San Nicolás de los Garza, NL, Mexico

Introduction

Universities and other research institutions constantly need to adapt their research strategy to meet quality standards and necessities at a local, national, and global level. A first example concerns *university rankings*, which have become an important performance metric (Tomas-Folch et al. 2015). For prestigious rankings, such as Times Higher Education (THE), Qacquarelli Symonds (QS), Academic Ranking of World Universities (ARWU), CWTS Leiden Ranking, and SCImago Institutions Ranking (SIR), research is considered as one of the most important criteria, and—as a result—the activity of research staff has gained attention among university managers. Motivated by this, institutions have developed research strategies directed to increase their scientific productivity and impact (e.g. citations).

A second example concerns the *triple helix* (Leydesdorff 2013; Etzkowitz and Leydesdorff 1998, 2000) and the *entrepreneurial university* (Etzkowitz 2003; Sam and Van Der Sijde 2014)—models that intertwine the university with the government and the industry and are shifting research from basic to applied and to socially-oriented or to industry-related. Even though the entrepreneurial university model has caused both positive and negative reactions (Martin 2012), it has become a reality and, in Latin American countries, it has started to develop (Heredia and Vinueza 2015) and thus requires support to take root. Two affected entities in this model are the professor and research groups. While the former has passed from (a) teaching to (b) teaching and researching to (c) teaching, researching, and managing research, the latter are now organised as "small businesses" and there is a need, at the institutional level, to encourage these dynamics.

A third example consists of the rise of multidisciplinary, interdisciplinary, and transdisciplinary research (Stokols et al. 2008; Klein 2008). Emerging fields such as Biotechnology require new schemes of research and collaboration to grow and develop. As such, institutional research strategies need to be shaped for coping with these changes and, thus, aid problem solving in these fields.

A key component of research is collaboration. According to Beaver (2001), collaboration has evolved from being dyadic to being in groups. Therefore, an important part of an institutional research strategy heavily relies upon its research groups. In this case, we refer to the *formal intra-institutional* research groups, which are the ones recognised by the university and—more importantly—can be directly strengthened, supported, and nurtured by this institution. Furthermore, research strategies can be directly aimed at these groups. In that sense, so far, other types of groups (e.g. collaborations outside the institution) lie out of our scope and are left for future work.

Strategy alignment sometimes implies group reconfiguration, i.e. adding, eliminating, splitting, or merging groups. To prevent reconfiguring groups arbitrarily, knowledge about the factors that influence group formation, persistence, and success is highly desirable. While there is a body of work that already touches this subject (see "Related work" section) little is known about—specifically—research group formation. Moreover, our key contribution is that—as we will see—we explore several factors under the *network analysis* perspective. This perspective, on one hand, works with objective information. On the other hand, it allows us to produce a unique methodology for combining factors.

In this work, we analyse the influence of seven factors on team formation: past coauthorships, recent coauthorships, previous co-participation in groups, citations to coauthored papers, co-participation in thesis committees, organisational closeness, and topical similarity. The first five factors are related to *previous collaborations*, i.e. collaborations, participation, or memberships that have taken place in the past. We consider that these

collaborations set up a precedent that could impact the construction of current research groups. With respect to organisational closeness, it has been observed that geographical proximity has a positive impact on collaboration (Katz 1994; Liang and Zhu 2002; Sun and Liu 2016); in consequence, it seems relevant to observe whether institutional proximity (which we refer to as *organisational closeness*) holds a similar effect by favouring the creation of intra-institutional research groups. With regard to topical similarity, when there is a thematic affinity among researchers, it seems feasible that these researchers work together; in fact, this should be one of the primary reasons for a research group to take place. Let us note that, for the seven selected factors—unlike for other potential factors—it is possible to extract reliable information (evidence) directly related to these factors (from Scopus, for example, or intra-institutional repositories).

Our aim is to observe whether these seven factors—either alone or combined—influence the formation of intra-institutional formal research groups, i.e. where members have an explicit adherence to the group. To achieve this goal, we use longitudinal data from a case study on the reconfiguration of research groups at a multi-campus private university. We use *network analysis* with these data to create networks that represent the factors and their combinations. Via *graph clustering techniques*, groups of researchers are extracted from these networks and compared against the formal groups of the case study (i.e. we are attempting to identify the formal research groups using solely the information provided by the networks, a task which we refer to as *prediction*). Different types of analyses (statistical, structural, frequency) allow us to determine the influence of the studied factors. In that sense, one contribution of this work is to provide evidence on the relevance of previous collaborations and coauthorships, organisational closeness, topic similarity, and their combinations on the formation of formal intra-organisational research groups through the use of network analysis. Another contribution consists of providing a method for evaluating simultaneously the effect of multiple factors on the formation of formal research groups.

This paper is organised as follows. "Related work" section presents related work on scientific collaboration, team formation and performance, and network analysis for studying groups. "Background" section introduces pertinent notions on graph theory and network analysis. "Data and methods" section presents the data used for the case study and the methods used for predicting group formation. "Results" section shows the relevance of each factor and the respective combinations on the prediction of group formation. Finally, in "Conclusions" section we conclude with closing remarks and future work.

Related work

Related work has been divided into three main categories: *Fostering scientific collaboration*, *Team formation and performance*, and *Group studies through network analysis*. The first category covers work that has been devoted towards discovering collaboration facilitators, such as geographical closeness, author prestige, and government policies; let us note that this section is not limited to collaboration in intra-organisational research groups, but covers dyadic, inter-organisational, and inter-regional collaboration. The second category views a research group as a team and covers work that has discovered factors for individuals joining teams and teams having a better performance; this section is not limited to research groups, but includes work with companies, students, and entrepreneurs, just to mention some. The third category covers studies related to group formation and evolution

from a network analysis perspective; these studies include collaboration and other social networks.

Fostering scientific collaboration

Katz ([1994](#)) summarized some of the main reasons that researchers consider for collaborating with another partner. Among these reasons one can find: (1) scientific popularity, visibility, and recognition of the partner, (2) increasing specialization in science, (3) the need of training researchers, and (4) a decrease in spatial distance. In the end, researchers collaborate to improve their productivity and impact (measured in terms of citations).

The effect of scientific popularity, visibility, and recognition has been studied through the analysis of co-authorship networks (Newman [2001](#), [2004](#)). For instance, it has been observed that the citation count of an article is positively correlated both with the number of authors and the number of institutions involved in its production (Katz and Hicks [1997](#)). It has also been discovered that centrality of an author in this kind of network positively influences the citation of the author's papers and the formation of new scientific collaborations (Uddin et al. [2013](#)). Martin et al. ([2013](#)) found that researchers cite their coauthor's papers more quickly than other researcher's work, and that there is a strong tendency to cite back other authors, specially a previous coauthor. Similarly, centrality of authors in coauthorship networks has been used for predicting whether an article will be highly cited 5 years after its publication (Sarigöl et al. [2014](#)).

On the other hand, geographical distance is a factor that discourages research collaboration. Katz ([1994](#)) observed that the frequency of research collaboration between domestic universities in United Kingdom, Canada, and Australia decreases exponentially with the distance separating research partners. A similar phenomenon has been reported by Liang and Zhu ([2002](#)) for Chinese inter-regional collaboration, as the probability of collaboration between regions dramatically increases when these are geographically close. At the organisational level, Balland ([2011](#)) discovered that geographical, intra-organisational (being part of the same organisation), and institutional closeness (belonging to the same type of organisation, e.g. government, education, company, public institution) all favour collaboration.

Other aspects that have facilitated collaboration at the inter-organisational level are government policies and inter-regional consortia. With respect to the former, it has been reported that government policies have contributed to shape over time the network of collaborations among public research institutes, companies, and universities (Zhang et al. [2016](#)). With regard to inter-regional consortia, these have shown varying degrees of impact in different countries when it comes to new collaborations (Cho et al. [2013](#)).

Hybrid indicators, introduced by Perianes-Rodriguez et al. ([2009](#)), reflect popularity and prestige. These indicators combine the clustering coefficient of a researcher in a coauthorship network with the number of papers and citations this researcher has; these three individual metrics, along with others, are obtained from a single coauthorship network. In our work, we built different networks encoding actual and potential collaboration factors and then combined them for producing hybrid networks, which are the result of aggregating and normalizing these factors. Hence, the two approaches are different.

Team formation and performance

Because the members of a research group work together to pursue common goals and their relationship is based on common values, confidence, and trust, a research group can be

studied as a *team*. According to Katzenbach and Smith (1993a), a team is *a small number of people with complementary skills who are committed to a common purpose, set of performance goals, and approach for which they hold themselves mutually accountable.* Teams, in general, can either be assigned by a competent authority (Reagans et al. 2004), be formed by voluntary participation (Margolin et al. 2012), or lie within these two extremes. While there are teams with stable memberships, others are highly dynamic (Putnam 1992) or are only conformed for completing a specific project (Hahn et al. 2006; Johnson et al. 2009). A team usually begins with a set of separate individuals, which later form a coalition, and finally achieve not only a higher performance but are able to care also for one another (Katzenbach and Smith 1993b). With regard to the process of an individual joining and staying in a team, it has been reported that individuals search for teams that can satisfy their needs and vice versa; in the case of a match, the individual joins the team. Afterwards, the individual attempts to increase the satisfaction provided by the group, while the group attempts to increase the value provided by the individual; if these attempts are accomplished, commitment is strengthened and the group enters a phase of maintenance (Levine and Moreland 1991; Owens et al. 1998).

There is a substantial body of work related to team formation and performance, which has been studied on teams of entrepreneurs (Ruef et al. 2003), students (Hinds et al. 2000), academic institutions (Rey-Rocha et al. 2006; Martín-Sempere et al. 2008), companies (Casciaro and Lobo 2008), gangs (Johnson et al. 2009), online games, (Zhu et al. 2013; Johnson et al. 2009), product teams (Ancona and Caldwell 1992), and innovative and creative teams (Bercovitz and Feldman 2011; Guimerà et al. 2005)—among others. Works related to team formation are mostly devoted towards discovering the motivations behind team affiliation and tie construction, while works related to team performance are mostly devoted towards discovering the factors behind team success, where success can be measured in different forms, such as effectiveness, throughput, satisfaction, production volumes, easiness of communication, etc. As we will see, some factors are present both in formation and performance. Furthermore, some of these studies rely on *network analysis* to gather results.

In an early work, Mattessich and Monsey (1992) encompass factors for successful team performance into six categories: environment, membership, structure (e.g. flexibility and adaptability), communication, purpose (e.g. shared vision), and resources. Cohen and Bailey (1997) report that team performance (measured in their work in terms of effectiveness, member attitudes such as commitment, and behavioural outcomes such as turnover) was found to be influenced by variables related to team composition, such as team size and different types of diversity (e.g. functional, tenure); for instance, coupling team size according to workload has shown to increase effectiveness (Campion et al. 1993).

With regard to member quality affecting team performance, several studies report that individual expertise is positively related to team performance. Volmer and Sonnentag (2011) observed this pattern in software design teams, where team experts would intellectually stimulate the rest, and these authors also noted that expertise would enhance performance in team meetings (Sonnentag and Volmer 2009); in the problem-solving domain, Bonner et al. (2002) found that teams perform at the level of their best members. Verbree et al. (2015) studied the organisational factors that influence the scholarly performance of research groups. They found that the group composition and size as well as the presence of senior staff in the group have a positive influence on productivity and citation.

Several studies agree on the positive impact of previous collaborations for team performance, as these collaborations favour coordination and communication (Bercovitz and Feldman 2011) and reduce the negative effect of distance (Cummings and Kiesler 2008);

moreover, familiarity within teams and role experience has been positively related to team performance (Huckman et al. 2009), and teams composed by friends have shown to out-perform teams of mere acquaintances (Shah and Jehn 1993). Excessive familiarity, how-ever, may hinder results, causing teams of strangers to achieve a better performance when different points of view are required for a task (Gruenfeld et al. 1996). Another factor for team success, as discovered by other studies, is *heterogeneity*—i.e. the presence of com-plementary skills, status, and knowledge in a team—, which encourages communication and innovation (Ancona and Caldwell 1992) and creates synergy among team members (Bercovitz and Feldman 2011).

With regard to team formation, *previous collaborations* (specially successful ones) have also been regarded as an influential factor (Hinds et al. 2000), since uncertainty is reduced in the team (Hahn et al. 2006). Heterogeneity, as well, has played a significant role in team formation (Dahlander and McFarland 2013), as people seek in others complementary skills to complete difficult projects (Johnson et al. 2009). On the con-trary, it has also been observed that *homophily* is a key component in forming teams (Ruef et al. 2003), as people tend to look for similar teammates, either in age, race, skills, or organisation (Hinds et al. 2000; Zhu et al. 2013). In the end, a balance between homophily and heterogeneity seems to be necessary for team formation (Owens et al. 1998). Other influential factors for team formation include affect (Casciaro and Lobo 2008), a good reputation for work (Hinds et al. 2000), and strong relationships such as marriage (Ruef et al. 2003).

In scientific research, where collaboration has evolved from being mostly dyadic to being mostly teamwork over the last decades (Beaver 2001), it has been observed that researchers belonging to consolidated teams have a competitive advantage in terms of JCR articles (Rey-Rocha et al. 2006, 2007), and committed members in these teams have more scientific production (articles, supervised theses, projects, patents) than their colleagues (Martín-Sempere et al. 2008). Research groups that collaborate with top-tier universities have also shown to produce papers with the highest impact (Jones et al. 2008). For research groups in developing countries, collaboration overseas has benefited scientific production (Ordóñez-Matamoros et al. 2010).

Group studies through network analysis

Group formation and evolution have also been studied from a purely structural point of view, i.e. by considering network properties and analysis. For example, based on the analysis of coauthorhsip and phone call networks, Palla et al. (2007) identify six possible events in the evolution of a group: birth, growth, contraction, merging, splitting, and death. Furthermore, depending on group size, there exist different optimal scenarios to ensure longevity. For small groups, stable membership is desirable, while member fluctuations favour large groups; not surprisingly, members with a weak commitment are more likely to abandon the group. In large social networks, the decision of an individual joining a group has been found to correspond to a *diffusion process*, where both the number of friends belonging to the group and the connectivity among these friends are relevant (Backstrom et al. 2006). Kairam et al. (2012) report that groups where a large *clique* (structure where all members are directly related to each other) is present exhibit faster growth and are less likely to disappear. In contrast, groups with a high transitivity are less inclined to grow. Furthermore, two types of groups are identified: those formed on common relationships and those formed on common traits (homophily).

From the network analysis perspective, research groups have been also studied through the detection of *graph clusters* in coauthorship and citation networks. The approach by Calero et al. (2006), for example, extracts coauthorship and topical networks from the field of Spanish nanotechnology; while clustering the former aids the detection of *actual* research groups, clustering the latter produces *potential* research groups, which are recommended as possible future collaborations. Similar works include the use of hierarchical clustering and multidimensional scaling (Yu et al. 2011), a new journal classification methodology based on citation network multilevel clustering (Waltman and Eck 2012), and the creation of clustering algorithms that take advantage of the two-mode (i.e. bipartite) nature of scientific collaboration networks (Garza and Schaeffer 2016).

Discussion

Works referred in previous subsections study the factors that motivate peer-to-peer scientific collaboration, as well as those factors that promote the growth of teams and improve their performance. However, they do not evaluate the possible combinations of factors or whether a factor is occluded by another. In this work we focus on the factors considered by researchers during the conformation of formal research groups. For this purpose we propose a method to combine multiple collaboration factors represented through networks, which are then clustered by automatic methods to determine which factors have a positive or negative influence in group formation.

Background

In this section, we present some basic concepts of graph theory. We also describe the network analysis metrics and graph clustering algorithms used in our work.

Graph theory

A network or *graph* $G = (V, E)$ is a structure that contains a set V of *vertices* (entities) and a set E of *edges* (connections between the entities); we will use n to denote $|V|$ and m to denote $|E|$.

The *degree* of a vertex is the number of edges attached to that vertex, and a graph whose edges have numerical labels is a *weighted graph*. A *connected component* is a subgraph (portion of a graph) where every pair of vertices is connected, i.e. it exists a path between them.

The *density* of a network is the proportion of edges present with respect to the theoretical maximum:

$$\delta = \frac{2m}{n(n-1)}. \tag{1}$$

Graph clustering

Graph clustering, a tool from network analysis, consists of identifying cohesive subgraphs in a network; we will use the term *clusters* to refer to these subgraphs and the term *clustering* to refer to the collection of clusters that result from the application of a

graph clustering algorithm to a network. While there are different definitions for a cluster (Radicchi et al. 2004; Wasserman and Faust 1994), the commonly accepted notion implies that the cluster has more (or stronger) edges between its own vertices and fewer (or weaker) edges between its vertices and the rest of the network. Graph clustering is also known as *cluster analysis* or *community detection* and clusters are also known as *communities*. Clusterings with disjoint clusters are commonly known as *partitions*.

In this work, we use the following five well-known clustering algorithms that generate partitions.

The Girvan–Newman method (Girvan and Newman 2002). This classical algorithm consists of an agglomerative approach that iteratively removes those edges with the highest *betweenness centrality*. Betweenness centrality measures the extent up to which an edge joins clusters.

The Clauset et al. method (Clauset et al. 2004). This algorithm enhances the Girvan–Newman method by using auxiliary data structures (balanced binary trees, max-heaps).

The Louvain method (Blondel et al. 2008). This is a hierarchical algorithm that consists of two main phases. At the first phase, every vertex starts in its own cluster, and a vertex is moved to a neighbouring cluster when a positive increase in *modularity* (a clustering fitness measure) is detected. At the second phase, each cluster is collapsed into a *super-vertex* and the first phase is then repeated. The algorithm stops when there are no more changes and modularity has reached a maximum. It is possible to process weighted graphs with this algorithm.

Label Propagation (Raghavan et al. 2007). This algorithm is based on infectious disease spreading. Each vertex starts with its own label. At each iteration, each vertex is assigned with the label that the majority of its neighbours has and ties are broken randomly. The algorithm finishes when all vertices have the label from the majority of their neighbours. Even though weighted versions of this algorithm exist, we use the original version for its simplicity and rapidness.

Weighted Graph Local Clustering (Schaeffer 2005; Garza and Schaeffer 2016). This algorithm is a bottom-up approach where each candidate cluster starts as a singleton; for each cluster, new vertices are iteratively added (or, alternatively, removed) with the intent of maximizing a fitness function. The candidate cluster is finally accepted or rejected. The algorithm finishes when all vertices have been processed (i.e., considered for clustering). This algorithm is able to process weighted graphs.

Clustering evaluation

There are different forms of evaluating a clustering. One of these consists of *external evaluation*, where there exists a pre-established *result model* and the clustering is compared (matched) against this model. The result model is represented by a set of *reference classes*; these reference classes can be obtained by reporting ground truth groups or by automatic generation in the case of artificially-generated networks. Let us introduce the three external evaluation metrics used throughout the present work: the Adjusted Rand Index, precision, and recall.

The Rand Index is a common metric for comparing partitions. It is based on agreements and disagreements. Let us consider a set $\mathcal{N}$ of n elements and two partitions for this set: $\mathbb{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_r\}$ and $\mathbb{L} = \{\mathcal{L}_1, \ldots, \mathcal{L}_s\}$, where each partition is a set of disjoint groups.

Additionally, let us consider that $\mathbb{C}$ is the clustering obtained from an algorithm and $\mathbb{L}$ is the set of reference classes. The Rand Index RI is then calculated as

$$\text{RI} = \frac{a+b}{a+b+c+d}, \tag{2}$$

where a is the number of pairs of elements that appear in the same group both in $\mathbb{C}$ and $\mathbb{L}$, b is the number of pairs of elements that appear in different groups in $\mathbb{C}$ and in different groups in $\mathbb{L}$, c is the number of pairs of elements that appear in the same group in $\mathbb{C}$ but in different groups in $\mathbb{L}$, and d is the number of pairs of elements that appear in different groups in $\mathbb{C}$ but in the same group in $\mathbb{L}$. As we can see, the numerator accounts for agreements in the partitions and the denominator accounts both for agreements and disagreements. In fact, $a+b+c+d = \binom{n}{2}$.

The Adjusted Rand Index (ARI) is a version of the Rand Index that is corrected for chance. To calculate ARI, a *contingency table* between $\mathbb{C}$ and $\mathbb{L}$ is formed, and each row and column are respectively summed (elements in the form a_i and b_j):

$$
\begin{array}{c|cccc|c}
 & \mathcal{L}_1 & \mathcal{L}_2 & \dots & \mathcal{L}_s & \sum \\
\hline
\mathcal{C}_1 & n_{11} & n_{12} & \dots & n_{1s} & a_1 \\
\mathcal{C}_2 & n_{21} & n_{22} & \dots & n_{2s} & a_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\mathcal{C}_r & n_{r1} & n_{r2} & \dots & n_{rs} & a_r \\
\hline
\sum & b_1 & b_2 & \dots & b_s & n
\end{array}
$$

where n_{ij} is the number of common elements between groups $\mathcal{C}_i$ and $\mathcal{D}_j$, a_i is the sum of row $\mathcal{C}_i$, and b_j is the sum of column $\mathcal{D}_j$. The ARI index is then calculated as

$$
\begin{aligned}
\text{ARI} &= \frac{\text{Index} - \text{ExpectedIndex}}{\text{Max. Index} - \text{Expected Index}} \\
&= \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right]/\binom{n}{2}}.
\end{aligned} \tag{3}
$$

Let us note that, while the Rand Index is bounded between 0 and 1, the Adjusted Rand Index allows values between -1 and 1. The former indicates that the partitions have nothing in common and, conversely, the latter indicates that the partitions are identical.

With respect to precision and recall, which correspondingly evaluate correctness and completeness, let us consider a cluster $\mathcal{C}_i \in \mathbb{C}$ and a reference class $\mathcal{L}_j \in \mathbb{L}$. Precision (denoted by P_i for $\mathcal{C}_i$) is the fraction of cluster elements that actually belong to the reference class (Eq. 4). Recall (denoted by R_i for $\mathcal{C}_i$) is the fraction of class elements that was actually placed in the cluster (Eq. 5):

$$P_i = \frac{|\mathcal{C}_i \cap \mathcal{L}_j|}{|\mathcal{C}_i|}, \tag{4}$$

$$R_i = \frac{|\mathcal{C}_i \cap \mathcal{L}_j|}{|\mathcal{L}_j|}. \tag{5}$$

For example, let us assume that the matching class for cluster $\mathcal{C}_i = \{a, b, c\}$ is $\mathcal{L}_j = \{a, b, d, e\}$. In this case, $P_i = \frac{2}{3} = 0.6$ and $R_i = \frac{2}{4} = 0.5$. Let us note that these scores are bounded between 0 and 1, where 1 is the best score possible. Let us also note that precision and recall are calculated per cluster $\mathcal{C}_i$; to obtain overall scores for clustering $\mathbb{C}$, it is necessary to integrate all individual cluster scores. For example, individual scores can be averaged:

$$P(\mathbb{C}) = \frac{\sum_i P_i}{|\mathbb{C}|}, \tag{6}$$

$$R(\mathbb{C}) = \frac{\sum_i R_i}{|\mathbb{C}|}, \tag{7}$$

where $P(\mathbb{C})$ is the average precision for clustering $\mathbb{C}$ and, likewise, $R(\mathbb{C})$ is the average recall for clustering $\mathbb{C}$.

Data and methods

To determine the influence of the proposed factors in the formation of formal research groups, we use data from a case study. With this data, we create networks that represent the proposed factors. We subsequently cluster these networks with the purpose of obtaining back the formal groups (a task known as *prediction*). Factor influence is then discovered by (1) observing if our obtained clusters resemble the existing formal research groups and (2) analysing which factors and factor combinations were present in the networks whose clusterings are the most similar to the formal groups (see "Results" section). It is important to clarify that, even though formal research groups do not always accurately reflect actual research teamwork, we are using these groups as the reference classes to evaluate our network clusterings because we are specifically interested in discovering factors that influence *formal* research group formation. In that sense, properly obtaining a *ground truth* for actual/potential scientific collaboration is considered as a related problem and is left as future work.

Research groups at Tecnológico de Monterrey

Tecnológico de Monterrey is a comprehensive multi-campus, teaching, and research university with its flagship campus in Monterrey, Mexico. In 2003, a model for fostering the transition from a teaching to a research university was implemented. This model was based on the organisation of researchers into groups lead by a principal researcher (Cantu et al. 2009). These groups, originally called *Research Chairs*, were constituted by 20 researchers in average among researcher professors, graduate students, and postdocs, and were evaluated collectively according to a set of indicators that included scientific publications, grants, patents, and the promotion of researchers in the Mexican Researcher System (SNI by its initials in Spanish).

Currently, the academic staff is composed of approximately 800 research professors and 1600 graduate students at the doctoral and master level from disciplines in engineering,

information technologies, social sciences, arts and humanities, natural sciences, and health sciences.

This program started with approximately 20 research groups and demonstrated that organisation and collaboration improved the scientific productivity of the university (Cantu and Ceballos 2012). Ten years later, the number of research groups grew up to 120, but the number of researchers increased only marginally, provoking the fragmentation of research lines on one hand and the increase of collaboration among groups on the other.

This motivated an exercise of research focus where researchers were distributed among five national schools and organised in 41 national research groups. The formation of these groups was directed by deans who nominated their main researchers and the areas of interest. All other researchers decided which group to affiliate to.

During the first 10 years of this program the information used for evaluating the groups was collected in the *Current Research Information System (CRIS)* developed at the university (Cantu et al. 2005). Collected information permitted to identify coauthorship relationships, co-participation on groups along this time, the journals on which researchers published, and all of this was associated to personal identifiers traceable across institutional databases that enabled to identify, for example, the department on which researchers were employed.

For our case study, we selected a set of 428 researchers that participated in the formation of the first 41 national groups. This set is constituted by 40 leaders, 345 core researchers, and 43 postdoctoral researchers. The sample had 4 groups from the Medicine School, 18 groups from the Engineering School, 5 groups from the Humanities School, 9 groups from the Business School, and 5 groups from the Public Policy School.

A summary of the distribution of researchers per group and their productivity, both per school, is shown in Table 1. As it can be seen, Humanities has the largest groups, followed by Medicine, whereas the Public Policy School has the smallest groups. Table 1 also shows the average number of publications made by researchers of every school, where the productivity of the Engineering School's researchers is the largest.

Table 2 shows the breakdown of the number of publications by document type in each school. Using the full period of time (2003–2014), this table illustrates the difference on publication patterns per school; for instance, researchers of Humanities and Public Policy are those who publish books the most, whereas researchers of Engineering and Medicine Schools are more used to produce journal articles.

Whereas the number of publications in Tables 1 and 2 were obtained from the institutional CRIS, we also identified the publications of these researchers in the Scopus

Table 1 Indicators of groups, researchers and productivity per school

School	Groups	Researchers	Average researcher per group	Avg. docs. (2003–2014) per researcher
Humanities	5	102	20.4	9.5
Business	9	67	7.4	7.6
Public policy	5	24	4.8	6.8
Engineering	18	179	9.9	12.9
Medicine	4	56	14.0	6.4
Total	41	428	10.4	10.3

School	Articles	Books	Chapters	Inproceedings
Humanities	6.3	5.1	11.1	18.9
Business	6.6	2.9	5.6	15.0
Public policy	4.6	6.8	10.2	5.5
Engineering	10.5	2.8	4.7	26.4
Medicine	7.5	1.7	3.2	7.1
Total	8.2	4.0	7.0	20.4

Table 2 Average number of publications (2003–2014) per researcher in every school

database. We counted the number of citations to papers published by these authors between 2010 and 2015, using November 11th, 2015 as cut-off date. In the Scopus database we found all the document types listed in Table 2, as well as citations to these documents. Table 3 shows the average number of citations received by collaborations between pairs of researchers from different or the same school (the total number of citations for all papers written in coauthorship).

Research collaboration networks

From the CRIS, other institutional sources and information obtained from Scopus we generated seven different weighted networks. In all these networks, vertices represent researchers and edges represent a particular relationship (e.g. collaboration), where edge weight represents the intensity of this relationship (for example, frequency, similarity, or closeness). The absence of an edge between vertices indicates the lack of a relationship between researchers in that feature. We obtained the following networks:

Past Coauthorships (PC). In this network, weights represent the number of scientific publications (articles in journals or proceedings, chapters, and books) coauthored by each pair of researchers between 2003 and 2009.

Recent Coauthorships (RC). In this other network, weights represent the number of scientific publications (articles in journals or proceedings, chapters, and books) coauthored by each pair of researchers between 2010 and 2014.

Former Groups (FG). In this network, weights represent the number of research groups in which each pair of researchers co-participated during the 10 year period prior to the formation of the new groups. During those 10 years, researchers moved from one group to another or remained in the same group. In any case, the groups that prevailed at the end of that period are not necessarily subsets of the new groups. This definition does not take into account how much time the co-participation lasted.

Citation-weighted coauthorship (CC). This network represents the reward or success of a coauthorship collaboration, given by the number of citations in the Scopus database to papers published between 2010 and 2015, accounted by November 11th, 2015. The

Table 3 Cites per collaboration between researchers across schools

School	Humanities	Business	Engineering	Medicine
Humanities	0.5	5	2.5	–
Business		10.3	0	–
Engineering			18.5	9.1
Medicine				17.2

number of citations received by publications coauthored by any pair of researchers was normalised in two ways: (1) by discipline, and (2) for linearity. Discipline is represented by the school every researcher is ascribed to and normalization considers the average number of cites received by intra and inter-school collaborations (Waltman 2016). In this way, citations were divided by the average number of cites received by publications coauthored by researchers of the corresponding combination of schools (see Table 3). Discipline normalization values range from 0.052 to 14.108 whenever the collaboration produced at least one citation. Collaborations that received 0 citations were set to 0.005. Then, for producing a linear distribution we used the formula $\text{norm}(C) = \ln(C + 1)/\ln(\max(C) + 1)$, where C is the discipline-normalised citation number and $\max(C)$ represents the maximum value for C. We added 1 to the number of normalised citations for having only positive numbers for each $\text{norm}(C)$. Collaborations without citations finally obtained a $\text{norm}(C) = 0.002$, whereas collaborations having at least one citation had $\text{norm}(C) \geq 0.019$.

Co-participations in thesis committees (TH). This network captures the number of thesis committees in which any two researchers participated between 2000 and 2014. We only considered the relationship between the committee director and any other member of the committee, i.e. we did not count committees were both researchers participated as reviewers.

Topical similarity (TS). This network captures the number of scientific publications (articles in journal or proceedings, chapters, and books) written by two authors about the same topics between 2003 and 2014. This network represents potential collaboration since accounted publications could be written in coauthorship or not. TS edge weight is calculated by:

$$\sum_T n_{iT} + n_{jT} \tag{8}$$

where n_{iT} is the number of papers written by author i classified under category T, likewise n_{jT} for author j. Publications were classified by researchers in one or more topics T of an adaptation of the Hierarchical Interface to the Library of Congress Classification (HILCC) taxonomy, constituted by 59 categories.

Organisational Closeness (OC). This network represents how close two researchers are in the university with respect to the organisational hierarchy, i.e. considering the department where each researcher is affiliated to. If a pair of researchers are in the same department (e.g. the Computer Science Department at Monterrey Campus), the edge between them weighs 1.0; if they both work at sibling departments (e.g. the Computer Science and Mathematics Departments, both depending from the Engineering Division at Monterrey Campus), the edge weighs 0.5; and if they both work on departments depending from sibling departments (e.g. the Computer Science Dept. at the Engineering Division and the Ethics Dept. at the Social Sciences and Humanities Division, both at Monterrey Campus), the edge weighs 0.25; otherwise (e.g. the Computer Science Dept. at the Engineering Division of Monterrey Campus, and the Computer Science Dept. at the Engineering Division of the Mexico City Campus), there is no edge between the pair of researchers. Note that researchers working in the same department form cliques which dramatically increases the density of this network. This network is used as a representation of both organisational and geographical proximity (Balland 2011), since researchers are distributed in 31 campuses located in different cities across Mexico.

Except for the Organisational Closeness (OC) and the Citation-weighted Coauthorship (CC) networks, all other networks were normalised in the range (0,1] using

$$w_i = \frac{\ln(o_i)}{\ln(\max_i)}, \qquad (9)$$

where w_i is the *normalised* weight, o_i represents the *original* weight, and $\max_i$ is the maximum weight found in that network.

The normalised version of these seven networks represents *pure* networks because a single relationship is denoted. Table 4 shows a summary of the properties of these networks. From these properties, three key points stand out. First, we can observe that the number of vertices of CC is relatively low, the reason is that not all vertices had edges (specially since citations were extracted from Scopus and not all researchers had publications indexed in this database) and isolated vertices were removed from the graph (isolated vertices cannot be clustered using graph clustering). The second key point consists of the difference in edges, density, and average degree that TS and OC exhibit in comparison with the rest of the other networks. While both of these networks—strictly speaking—do not represent collaboration relationships, in a broad sense we can state that they reflect *potential collaborations*, since they associate pairs of researchers with a certain type of affinity that could lead them to collaborate (e.g. topical affinity). Because potential collaborations reflect *what could be*—as opposed to *what really is*—, it seems natural for these networks to have more edges and, consequently, a larger degree per vertex and a larger density.

We also generated *hybrid* networks by combining two or more pure networks. To combine networks, we take the union of vertices and the union of edges. The weights of common edges in pure networks are added and every weight w_i is normalised by dividing it by the maximum weight $\max_i$ of the resulting network, i.e.

$$n_i = \frac{w_i}{\max_i}, \qquad (10)$$

where n_i is the normalised weight. We use a *binary scheme*, where pure networks are either present or absent from hybrid networks. We use this scheme because it provides a manageable number of combinations to explore, results are clear and easier to interpret, and it is easier to observe factor influence in formal group formation; using this scheme can be seen as an initial approach to our problem and employing a real-valued scheme is left as future work. Considering the binary scheme and the existence of seven pure networks, $2^7 - 1 = 127$ networks (120 hybrid and the rest pure) are possible.

Table 4 Properties of pure networks

Network	Vertices	Edges	Avg. degree	Density (δ)
PC	178	273	3.067	0.017
RC	257	434	3.377	0.013
FG	302	742	4.914	0.016
CC	141	244	3.461	0.025
TH	256	461	3.602	0.014
TS	315	6943	44.083	0.140
OC	357	8595	48.151	0.135

Vertices represent researchers and edges represent relationships (e.g. actual/potential collaboration) between researchers

A hybrid network resulting of combining k pure networks $(1 \le k \le 7)$ has $\max_i \le k$ as long as each pure network has been previously normalised $(w_i \le 1.0)$. And by normalizing edge weights in the hybrid network we have a normalised network again where $n_i \le 1.0$.

Isolated vertices are removed from both pure and hybrid networks to prevent the detection of clusters containing a single vertex due to the lack of connections. This explains the difference on the number of vertices in pure networks shown in Table 4.

Influence determination through graph clustering

To determine the influence of the studied factors in the creation of formal research groups, we attempt to identify the 41 groups from our case study by clustering each one of the 127 generated networks, a task which is referred to as group *prediction*. In that sense, we assume that, the better the clustering predicts the existing formal research groups, the better that network quality is and—consequently—the more that the factors involved on it are influential. Alternatively, this could be seen as finding the network (factor combination) that maximizes the clustering score obtained by external evaluation.

To have a variegated set of clusterings per network, we use the algorithms described in "Graph clustering" section: the Girvan–Newman method, the Clauset et al. method, Weighted Graph Local Clustering (GLC), the Louvain algorithm, and the Label Propagation Algorithm (LPA). "Results" section discusses our main results.

Results

To discover the extent up to which do previous collaborations, organisational closeness, and research topic similarity play a role in the formation of formal research groups, each of the 127 networks was clustered with the aforementioned algorithms. For algorithms that do not process weighted networks (Clauset et al., Girvan–Newman), unweighted versions of the networks were prepared. The resulting clusters were compared against the existing formal research groups using the Adjusted Rand Index (ARI), explained in "Background" section, Eq. (3). For non-deterministic algorithms (Louvain, GLC, LPA), 30 algorithm repetitions were performed and the best score was taken (we justify this decision in "Best and worst networks" section). To evaluate the quality of a network, and thus extract influential factors, we assumed that this quality was reflected by the score of the clusterings obtained from the network. In other words, the quality of a network was equal to the similarity between the clustering (produced by the network) and the formal research groups, i.e. the score of the network's clustering.

To examine factor influence, we first selected results based on algorithm performance. We then generated an overview of these results, validated them, and conducted a series of analysis. Our analyses revealed strong, weak, positive, and negative influence of the different factors, as well as relevant combinations. We discuss the meaning of these findings.

Clustering algorithm selection

Since clustering results partially depend on algorithm quality, our first step for result analysis consisted on determining if there were significant differences among the scores obtained by the algorithms. Figure 1 summarizes the ARI scores obtained with each algorithm when comparing the formal research groups against the clusterings obtained

from the 127 networks. We conducted a one-way ANOVA using these results, and the test showed that the difference between these clustering algorithms was significant ($p < 0.01$). Since the Louvain algorithm obtained the best scores (see Fig. 1), we selected this algorithm for analysis of the clustering results. We assumed that this selection would mitigate the effect of algorithm performance and that the selected clustering results would be more reliable.

Relevance of factors

To validate the relevance of the selected factors (i.e., as a form of control), we generated 30 random clusterings for each one of the 127 networks. Each random clustering was a modified version of a clustering obtained with the Louvain algorithm; in this modified version, vertices were randomly reassigned to clusters.

Figure 2 compares the best ARI scores obtained from clustering each network with Louvain's algorithm and the best ARI scores obtained by generating random clusterings from each network (note that random clusterings do not exploit network information); Fig. 3 summarizes these scores using a box diagram. As you can see in both figures, the scores achieved using network information and a clustering algorithm are amply superior to the scores obtained by random clusterings. This indicates that both, selected factors (pure networks) and their combinations (hybrid networks), are meaningful for identifying actual formal groups and not a mere act of chance.

Comparison against random clusterings shows, as well, that *indeed* there exists a correspondence between our networks and the current formal research groups; at the same time, we could state that the factors involved in these networks match the formal research groups in varying degrees (i.e., better or worse scores can be mapped to specific factors and interactions). However, let us also note that the obtained clustering scores (0.5 maximum, approximately, and 0.3 on average) are still distant from the ideal, which is 1.0. This may be due to several reasons, including clustering algorithm quality and/or networks with incorrect or incomplete information; these scores suggest the presence of other factors that are currently not being taken into account—a topic which will be discussed later (see "Discussion" section).
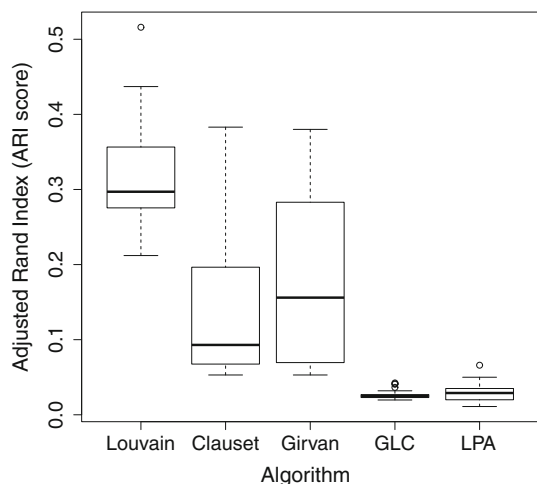

Fig. 1 Clustering algorithms comparison

Fig. 2 Best ARI scores for clusterings of pure/hybrid networks (Louvain) and random clusterings generated from them (Random)
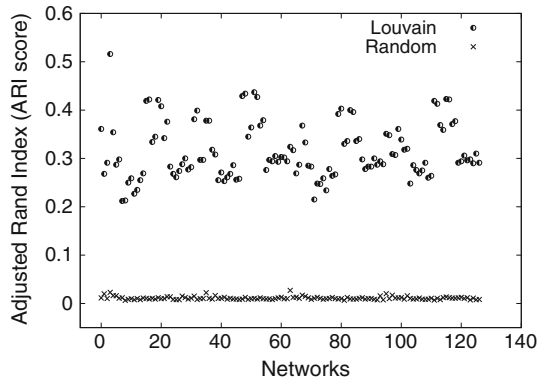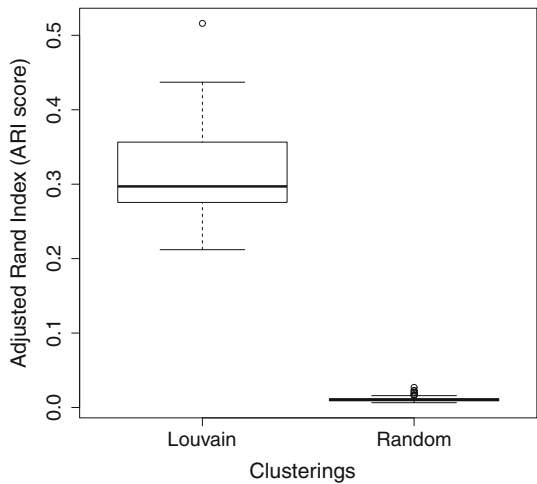

Fig. 3 Box plot of ARI scores for clusterings of pure/hybrid networks (Louvain) and their randomized versions (Random)

Best and worst networks

Tables 5 and 6 respectively present the networks that obtained the highest and lowest ARI scores (consequently, we could regard these networks as, respectively, the best and worst from the 127). In our methodology, the best network is the pure or hybrid network that obtains the best ARI score (closest to 1) when comparing the clustering that produces the Louvain algorithm for that network against the configuration of the formal groups; conversely, the worst network is the one whose ARI score is closest to 0.

As previously stated, we used the best score obtained out of the 30 algorithm repetitions. Nevertheless, in the tables, we include both the best and average scores obtained out of the 30 repetitions to show that, with the exception of some of the worst networks, best scores only vary slightly from average scores. In addition, the best and worst ten consisted of the same elements (in a slightly different order) regardless of whether the best or average scores were used, implying that the scores were considerably consistent from one repetition to another and the best scores were not outliers.

So far, the only clear patterns we can observe are the following: (a) former groups (FG) appears in all but one of the best networks, (b) FG does not appear in the worst networks,

Table 5 Best networks

Rank	Network	Best score	Average score
1	CC	0.516	0.49
2	RC + FG + CC	0.436	0.429
3	RC + FG + TH	0.434	0.424
4	RC + FG	0.429	0.421
5	RC + FG + CC + TH	0.427	0.419
6	PC + RC + FG + CC	0.422	0.415
7	PC + RC + FG + CC + TH	0.422	0.404
8	FG + TH	0.422	0.399
9	FG + CC	0.421	0.42
10	FG	0.419	0.416

Table 6 Worst networks (in descending order)

Rank	Network	Best score	Average score
127	TS	0.212	0.176
126	TH + TS	0.213	0.189
125	PC + TS	0.215	0.188
124	CC + TS	0.227	0.189
123	PC + CC + TS	0.234	0.211
122	CC + TH + TS	0.235	0.215
121	PC + TS + OC	0.247	0.223
120	PC + TH + TS	0.248	0.225
119	PC + RC + TS	0.249	0.214
118	TS + OC	0.252	0.226

(c) topical similarity (TS) appears in all the worst networks, and (d) TS does not appear in the best networks. Nevertheless, to understand why the networks in Tables 5 and 6 are among the best or the worst, we performed distinct types of analysis, which range from factor frequency to formal statistical analysis.

Influence of factors

With the aim of discovering the most influential factors, we performed several types of analysis over the obtained clustering results. These include structural, frequency, correlation, and statistical analyses. Structural analysis—which is performed at the network level—relates network quality to topological properties, and serves the purpose of determining what kind of structure was produced by influential factors. To have an overview on these factors, we calculate their frequency on the best and worst networks. Alternatively, correlation analysis (complemented with violin plots) reveals the relationship between factor presence and clustering scores. Finally, statistical analysis (performed with factorial ANOVA) reveals significant factors and interactions, which are explored by means of post hoc tests. Result interpretation and limitations identified in this work are provided in "Discussion" section.

Structural analysis

Analysing the structural properties of the best and worst networks reveals the relationship between topology and network quality. At the same time, it facilitates appreciating factor influence into clustering results. In this case, we analyse five network structural properties that varied across the 127 networks: number of vertices (n), number of edges (m), density (δ), number of connected components (c), and percentage of vertices in the largest connected component (%vc). To observe how these properties vary in the best and worst networks, we show their values for the top 5, 10 and 20% networks, as well as for the bottom 5, 10 and 20% networks (see Table 7). We also depict the relative value of these properties for the aforementioned networks via spider webs (see Fig. 4).

Recurrent properties of the best networks include a small amount of edges (small size), a low density, and a considerable number of connected components. Networks with these three properties are normally easier to cluster than networks whose structure is more intricate, as less partitions are possible with less edges and graphs that are already fragmented. Taking the former properties into account, the factors involved in the best networks create sparse structures with a few connections; these connections turn out to be the right ones, which causes high scores. However, it seems that there are also connections missing, which could be the cause for the scores still not reaching a peak close to the perfect one. The former can be evidenced if we calculate precision and recall, where the average precision for the clusterings of the best ten networks (Table 5) is 0.74, while the average recall is 0.42. This implies that the clusters produced tend to gather researchers who happen to be together in formal research groups as well (correctness) but are failing to gather *all* the researchers of each formal group (completeness). In addition, the clusterings from the best ten networks produced, on average, 74% more clusters than the amount of formal research groups, which is another indicator of missing connections. As mentioned earlier, the presence of other factors or artificial formal research groups could explain this lack of information. In this sense, the existence of subgroups or non-collaborative members inside a formal group could be easily discovered by analysing publications or other research outputs.

Table 7 Structural properties for the best and worst networks (all the presented quantities are averages, except for the best network and the worst network)

	n	m	δ	c	%vc
Best network (CC)	141.00	245.00	0.02	16.00	67
Top 5%	324.75	1035.25	0.02	9.50	89
Top 10%	329.23	1047.35	0.02	13.53	82
Top 20%	332.03	1055.07	0.02	12.50	86
Bottom 20%	350.08	6928.76	0.11	4.88	96
Bottom 10%	341.84	7090.62	0.12	1.92	99
Bottom 5%	337.00	7058.28	0.13	1.43	100
Worst network (TS)	315.00	6944.00	0.14	1.00	100

While n stands for the number of vertices in the network, m stands for the number of edges, δ for the density of the network, c for the number of connected components, and %vc for the percentage of vertices in the largest connected component
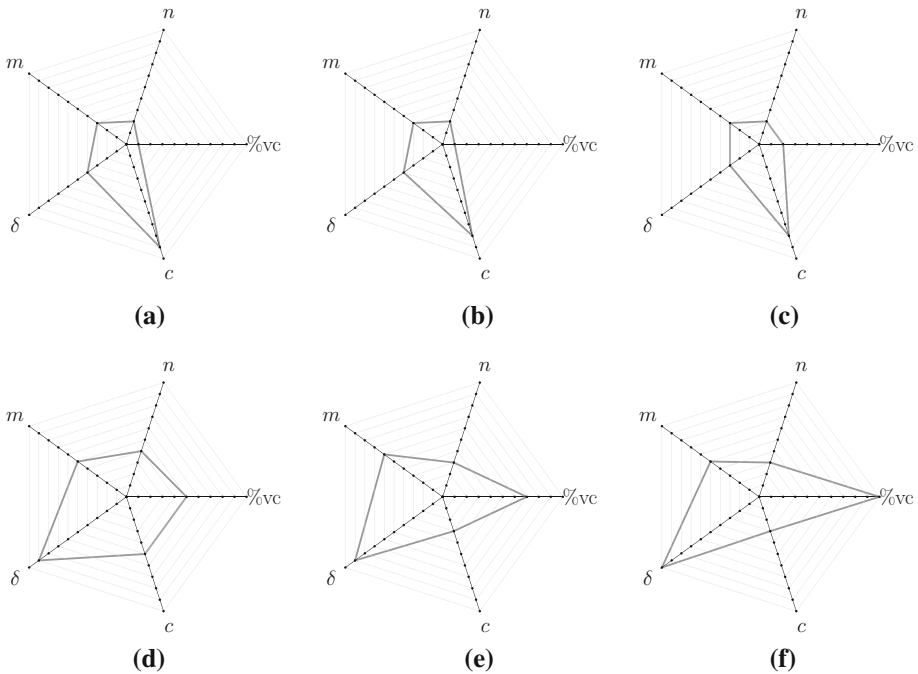
Fig. 4 Relative view of the structural properties for the top and bottom 5, 10 and 20% networks. **a** Top 5%, **b** top 10%, **c** top 20%, **d** bottom 20%, **e** bottom 10%, **f** bottom 5%

For the worst networks, on the other side, the recurrent properties include a relatively larger amount of edges, a relatively higher density, and a single connected component. The factors involved in these networks create denser structures with considerably more connections; these structures are more difficult to cluster and, even though there is valuable information, most of the connections turn out to be spurious. This can be confirmed, once again, by precision and recall, where the former is 0.52 on average for the ten worst networks and the latter is 0.64 on average (much better than for the best ten networks, and surprisingly the TS network—which obtained the lowest score of the 127 networks—has the best recall out of the 127 networks with 0.73). This implies that clusters tend to have a higher degree of completeness, but incorporate a higher number of outliers as well (lack of correctness); this can be appreciated by the number of clusters produced by these networks, which is on average 50% less than the number of formal research groups.

The network edge overlap matrix shown in Table 8, describes what percentage of a network is contained in another. It can be observed, for instance, that TS absorbs most of the other networks (see column TS), but at the same time, the fraction of edges in TS also present in the other networks is very low (see row TS); boths TS row and column are marked in bold in Table 8.

In summary, the factors corresponding to the best networks create sparse, precise structures (having only the necessary or less information); in contrast, the factors corresponding to the worst networks create dense, misleading structures (having information not taken into account during group formation).

So far, the analysis has been focused at the network level. However, the next analyses decompose networks into their underlying factors to reveal the relationship between

Table 8 Network edge overlap

	PC (%)	RC (%)	FG (%)	TS (%)	CC (%)	OC (%)	TH (%)
PC	–	48	29	**79**	34	21	53
RC	30	–	42	**82**	48	20	41
FG	11	24	–	**52**	14	23	20
TS	**3**	**5**	**6**	–	**3**	**4**	**5**
CC	38	85	42	**77**	–	22	43
OC	12	19	38	**56**	12	–	18
TH	31	39	31	**69**	23	18	–

Each cell a_{ij} is calculated as $|E_i \cap E_j|/|E_i|$ (note that the matrix is asymmetrical)

quality and factor presence or absence. At the semantic level, this provides an overview on the factors that help to predict research group formation.

Frequency analysis

Frequency analysis permits to initially discover which factors were present in the networks with the highest or lowest scores. We obtained factor frequencies for networks whose scores were above the mean, below the mean, in the top 20, 10 and 5%, and in the bottom 20, 10 and 5%.

Figure 5a and b show factor frequencies for networks whose scores were, respectively, above and below the mean (66 in the former case, 61 in the latter). As we can see, topical similarity (TS) is mostly absent in networks above the mean and, on the contrary, very frequent in networks below the mean, while the rest of the factors are almost balanced in both cases.
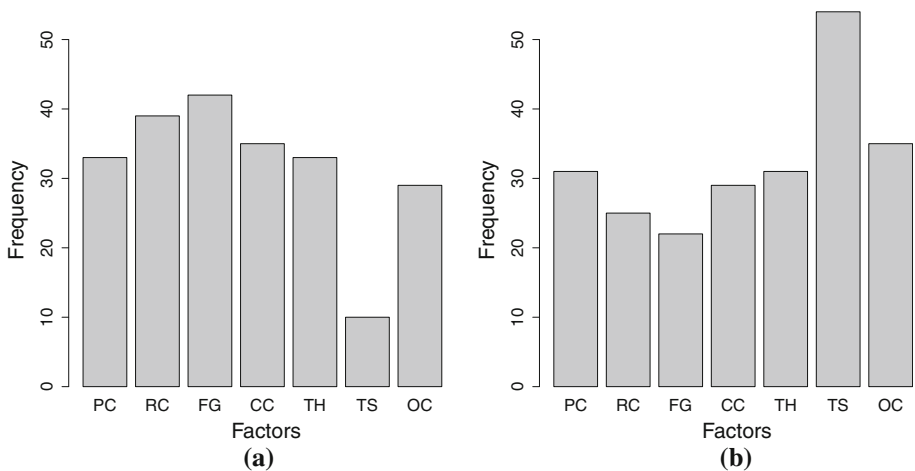

Fig. 5 Factor frequencies for networks whose score is above or below the mean. **a** Networks above the mean, **b** networks below the mean

Figure 6a and b show factor frequencies for the top and bottom 20% (26 networks in both cases). Interestingly, TS is absent from the top 20% and FG is minimally represented in the bottom 20%. Furthermore, FG—along with CC and RC—appears frequently in the top 20%, while TS—along with OC appears frequently in the bottom 20%. PC, RC, CC and TH appear to be balanced.

Figure 7a and b show factor frequencies for the top and bottom 10% (13 networks in both cases), where the scenario is similar to the previous one. TS remains absent from the top but has the highest frequency in the bottom, while FG maintains the opposite behaviour. RC and CC have relatively high frequencies in the top but keep low frequencies in the bottom. OC is absent from the top. PC and TH, once again, appear to be balanced.

Figure 8a and b show factor frequencies for the top and bottom 5% (7 networks in both cases), where we can finally see that FG, RC, and CC are the factors with highest frequency in the top and TS is the factor with highest frequency in the bottom.

In summary, three factors that are related to former collaborations (FG, RC, and CC) have showed to appear frequently in the best networks and factors pertaining to other aspects (topical similarity, organisational closeness) have shown the opposite behaviour. Two other factors related to former collaborations (PC and TH) have no clear pattern up to this point.

Correlation analysis

To confirm the results given by the frequency analysis, we calculated the Pearson correlation coefficient between the obtained network scores (Adjusted Rand Index) and the level (presence/absence, 0/1) of each factor in these. The results, presented in Table 9, indicate a strong negative correlation for TS, a weak negative correlation for OC, a moderate positive correlation for FG, a weak positive correlation for CC and RC, and no correlation for PC and TH. A negative correlation implies that the presence of that particular factor worsens the scores (scores are better without the factor), and a positive correlation implies the opposite (scores are better when the factor is present). These results can be visualised in Fig. 9.
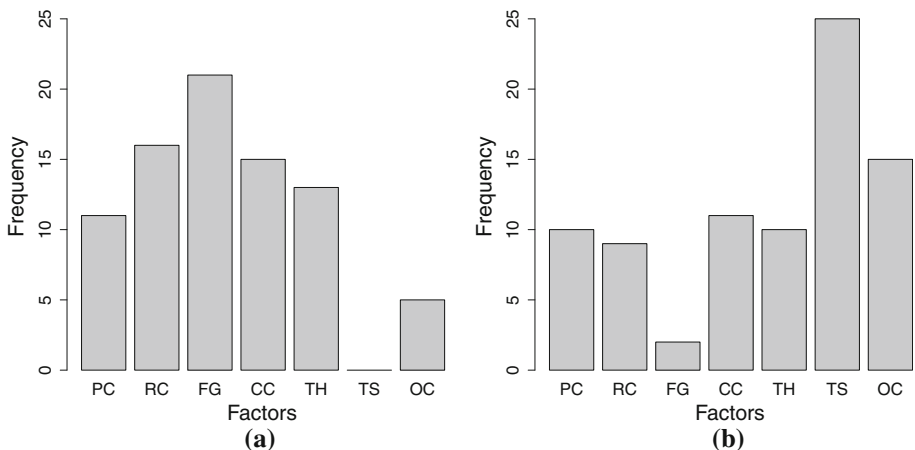

Fig. 6 Factor frequencies for networks in the top or bottom 20%. **a** Top 20% networks, **b** bottom 20% networks
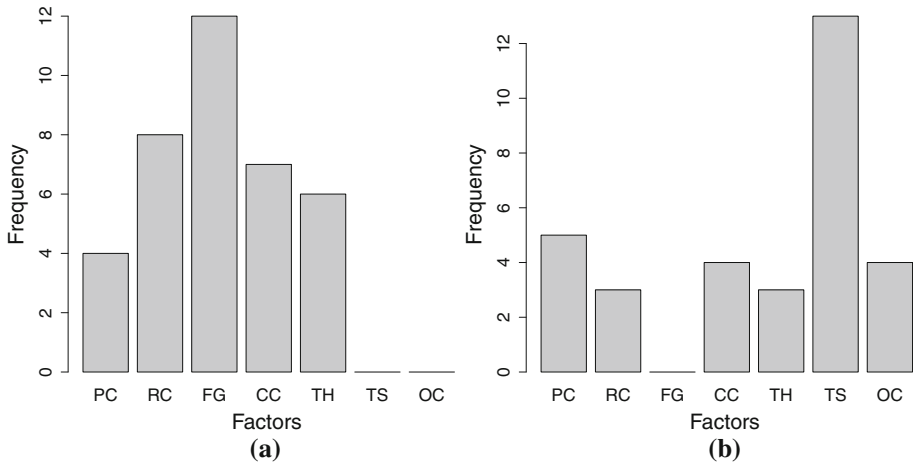
Fig. 7 Factor frequencies for networks in the top or bottom 10%. **a** Top 10% networks, **b** bottom 10% networks
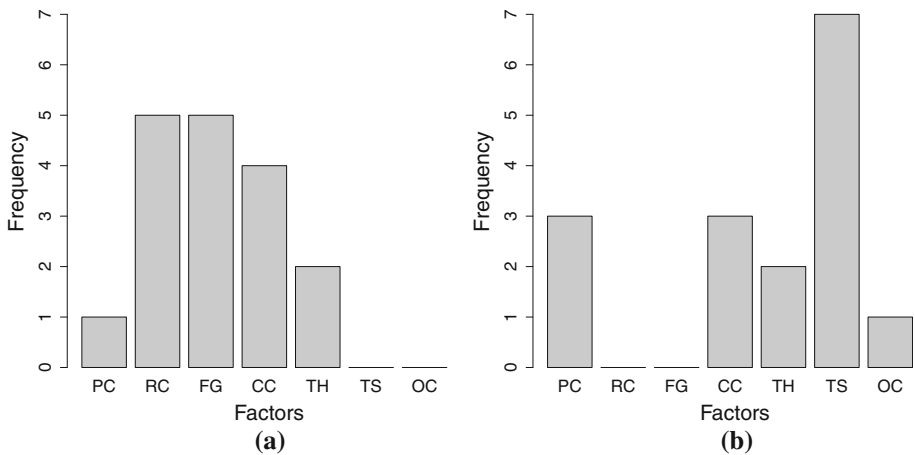

Fig. 8 Factor frequencies for networks in the top or bottom 5%. **a** Top 5% networks, **b** bottom 5% networks

Table 9 Correlations between scores and factor presence/absence

Factor	PC	RC	FG	TS	CC	OC	TH
Correlation	− 0.03	0.15	0.39	− 0.74	0.1	− 0.27	0.02

The two outstanding factors—once again—are TS (negatively) and FG (positively); see the violin plot[1] in Fig. 10a for a better appreciation. If we eliminate the networks that contain the TS factor, it is also possible to observe the negative influence of OC (Fig. 10b).

---

[1] In addition to including a box plot, a violin plot visualises data distribution (the distribution is mirrored and smoothed).
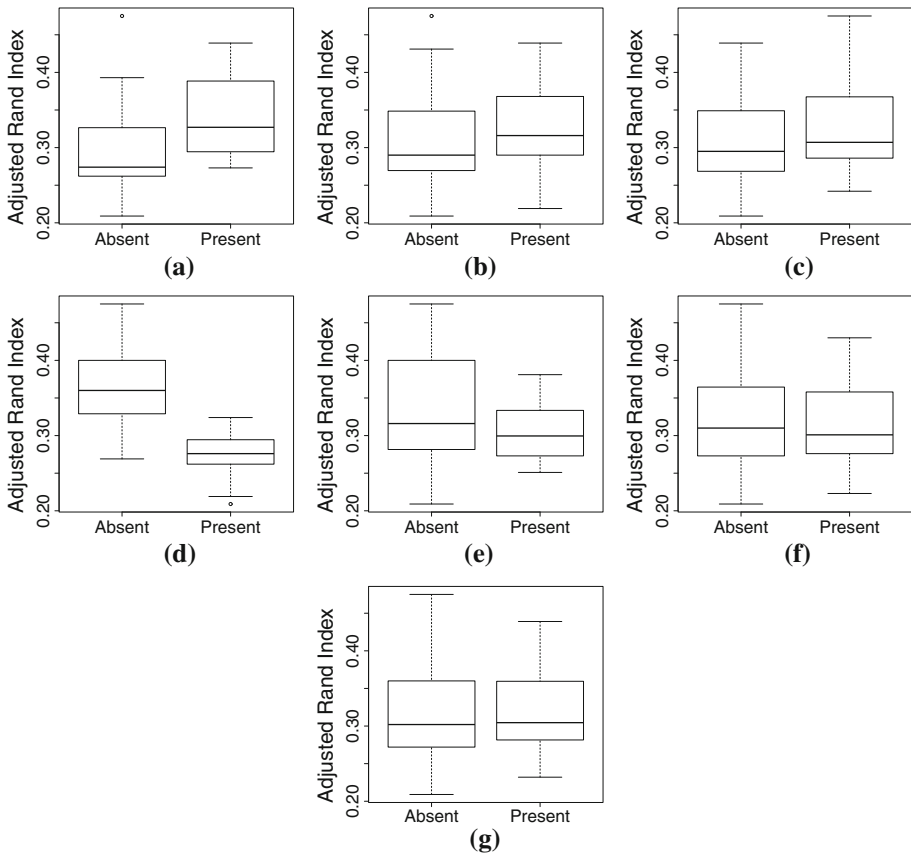
Fig. 9 Network scores with factors present and absent. **a** FG factor $(+)$, **b** RC factor $(+)$, **c** CC factor $(+)$, **d** TS factor $(-)$, **e** OC factor $(-)$, **f** PC factor, **g** TH factor

Statistical analysis

While frequency and correlation analysis have provided a general picture of each factor (whether it is associated positively or negatively with scores, and whether it has a strong or weak participation), statistical analysis provides formal tools for testing the significance of factors and discovering interactions among them. To this end, we performed factorial ANOVA on the factor combinations and their respective scores. In this case, all factors except for TH were significant $(p < 0.01)$. We also discovered significant interactions involving up to five factors (see Table 10), RC being present in each one of them. We explored these interactions visually using violin plots with partitions and by means of Tukey post hoc tests. Our main findings include the impact of TS and OC, as well as positive interactions.

Our first finding, which is consistent with the correlation and violin plots of the previous section, indicates that when TS is present along with one or more factors, the scores decline. The same applies to a lesser degree with OC. However, let us note that no factor combination yields a score lower than the one obtained by pure TS (i.e., the network with only TS present). This kind of interaction is illustrated by Fig. 11, where we can see that
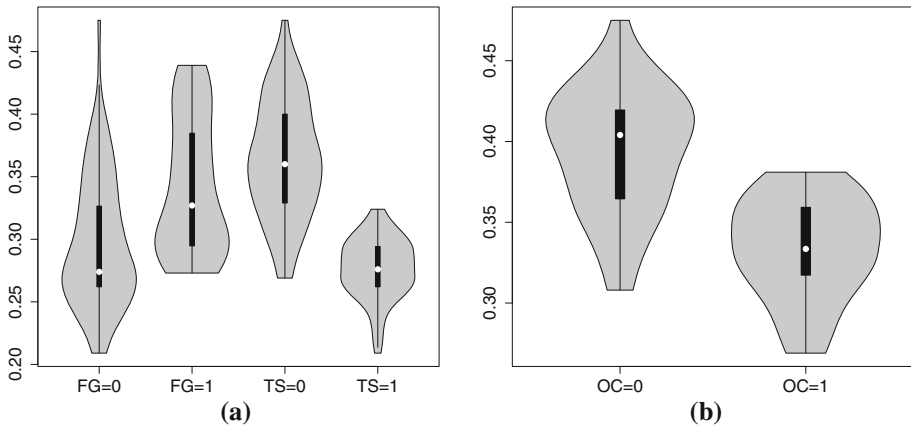
Fig. 10 Violin plots for the most outstanding factors. **a** FG and TS, **b** OC (without TS)

Table 10 Significant interactions ($p < 0.05$) among five factors	PC: RC: FG: TS: OC
	PC: RC: FG: OC: TH
	PC: RC: TS: OC: TH
	RC: FG: TS: OC: TH

the presence of TS hinders the scores of the last four distributions even though FG and RC are present (this decline is significant with $p < 0.01$). Note also in this figure the difference between the first and second pairs of distributions (which is also significant); this other decline results from the presence of OC.

We found, as well, a positive interaction between FG and RC, since better scores are produced when both factors are present. However, this improvement is significant only when TS, OC, and PC are present. As structural analysis has revealed, this happens because the TS and OC networks have more information and links, thus allowing a better synergy to take place between these factors (unfortunately, spurious links hinder the final results). Figure 12 illustrates this interaction (see the fourth and last distributions).

Furthermore, when RC, FG, and TH were present at the same time, the scores improved in comparison to the factors being alone or in pairs. However, the Tukey post hoc test revealed the improvement as non-significant. However, no factor combination was able to produce a score better than the one produced by CC alone.

When analysing the effect of PC, there seems to be no clear pattern. In some cases, its presence improves scores; in other cases, however, scores are worsened. "Discussion" section discusses the implications for this factor.

Best and worst networks revisited

The previous analyses permit to understand the top and bottom ten networks. With respect to the former, we can find that the CC network (ranked 1st) has the fittest structural properties, while the FG network (ranked 10th), on the other hand, has an underlying factor that is positively correlated to the clustering scores; networks ranked 2nd to 9th have
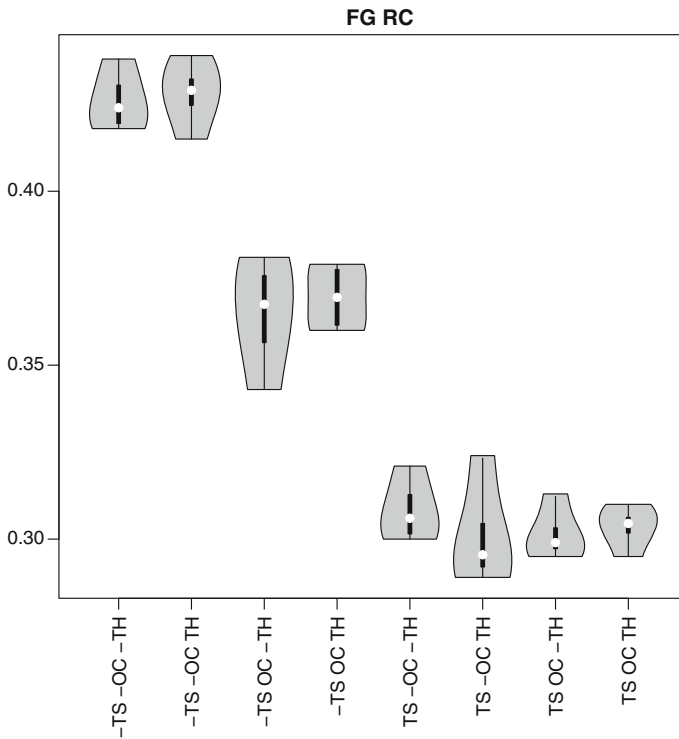
Fig. 11 Interaction between TS, OC, and other factors. The absence of a factor F is denoted as $-F$ in the horizontal axis. In this case, FG and RC are present in the eight distributions, TS is absent in the first four distributions and present in the last four, and the presence of OC and TH is interleaved (no/no, no/yes, yes/no, yes/yes)

underlying positive interactions among factors (note how almost all scores are even better than FG alone, as already stated in the previous section). With respect to the bottom ten networks, it now comes as no surprise the TS network to be ranked as the worst: its structure is misleading and its underlying factor is negatively correlated with the clustering scores.

PC and TH had a balanced presence both in the best and worst networks, this could explain the presence of networks ranked 3rd, 5th-8th, 119th-123rd and 125th-126th. However, if we analyse network edge overlapping (see Table 8), we will note that the TS network (which is the second largest) basically "absorbs" every other network it is combined with. Taking into account that CC, PC, and RC are respectively the networks with the smallest amount of edges, it seems reasonable to see combinations of these networks and TS among the bottom ten.

Discussion

In this section, we first discuss the influence of the selected factors on research group formation according to our methodology and our case of study. Then we discuss additional factors, interpret our results, and describe the limitations of our work.
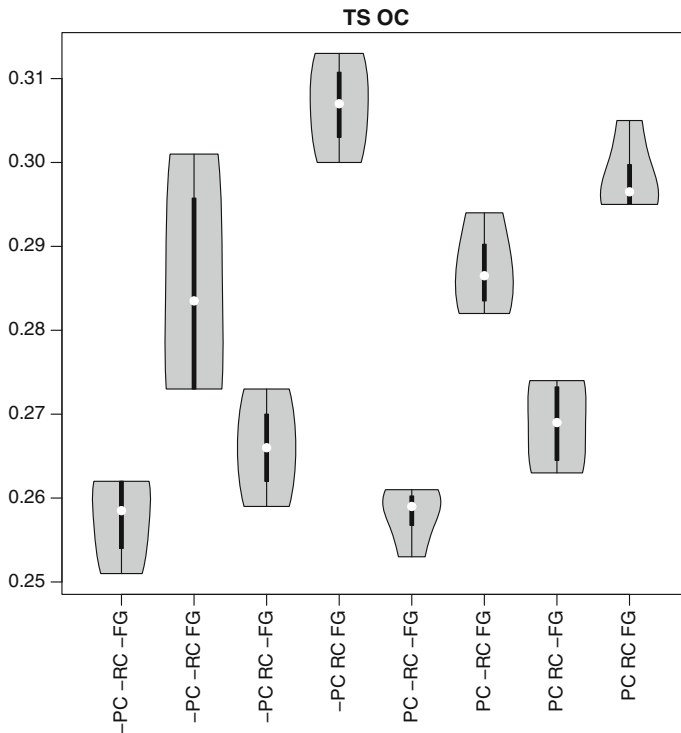
Fig. 12 Positive interaction between FG and RC. The absence of a factor is denoted as $-F$ in the x-axis, where F is the corresponding factor. In this case, TS and OC are present in the eight distributions, PC is absent in the first four distributions and present in the last four, and the presence of FG and RC is interleaved (no/no, no/yes, yes/no, yes/yes)

Analysed factors

Figure 13 shows a summary of the results observed on frequency, correlation and interaction analysis. A greener (darker) background reflects a more positive influence on group formation and a whiter background indicates a more negative influence. On frequency analysis, factor presence was normalised with respect to the number of networks on the top/bottom x percentage of networks (see column Norm); likewise, we normalised with the number of times each factor appeared in interactions.

Among the proposed factors, former groups (FG), citation-weighted coauthorships (CC), and recent coauthorships (RC) showed a positive influence in our clustering results. With respect to FG, it seems logical that researchers who have formed groups in the past tend to keep working together in new groups. A similar phenomenon can be observed with RC, as researchers who have collaborated through coauthorships apparently also tend to gather in research groups. Furthermore, if coauthorships have shown to play a significant role in group formation, it does not come as a surprise that successful coauthorships—such as those receiving citations (CC)—play an important role as well. In that sense, successful collaboration seems to be relevant for group formation. In the overall, these three factors confirm in our case study that former collaborations have an impact on formal research group formation. RC and CC, however, were weaker than FG. We believe this may be due

Criteria	PC	RC	FG	TS	CC	OC	TH	Norm.
Frequency in top 5% networks	14%	71%	71%	0%	57%	0%	29%	7
Frequency in top 10% networks	31%	46%	92%	0%	54%	0%	62%	13
Frequency in top 20% networks	44%	68%	88%	0%	60%	20%	52%	25
Frequency in top 50% networks	63%	72%	80%	19%	65%	54%	63%	54
Frequency in bottom 50% network	44%	34%	30%	74%	40%	48%	44%	73
Frequency in bottom 20% network	38%	35%	8%	100%	42%	58%	38%	26
Frequency in bottom 10% network	38%	23%	0%	100%	31%	31%	23%	13
Frequency in bottom 5% networks	43%	0%	0%	100%	43%	14%	29%	7
Correlation	-0.03	0.15	0.39	-0.74	0.1	-0.27	0.02	
Interactions (times in)	75%	100%	75%	75%	0%	100%	75%	4

Fig. 13 Summary of factor's influence on group formation. (Color figure online)

to coauthorships being less prone to drive people into a group in several cases—for example, when the coauthorship is multidisciplinary but the group needs to be unidisciplinary (restriction) or when the coauthorship is merely a sporadic collaboration (this applies in the case of RC).

Having split coauthorship in two periods permitted to distinguish between the influence of past and recent collaboration. However it is notable that PC appears in the top 10 best networks (Table 5) only when accompanied by RC (two times), whereas RC appears in four additional networks accompanied by other factors (six times in total). Presumably, the potential of a coauthorship deteriorates over time. With regard to thesis committees (TH), researchers merely collaborate in this kind of event and perhaps for this reason TH is not strong enough to influence group formation. Our results indicate that PC and TH are *weak* former collaborations and FG, CC, and RC are *strong* former collaborations, since the latter have greater influence in formal group formation.

Our results show a negative influence of the Topical Similarity (TS) network on group formation that can be explained by both structural and cultural reasons. On one hand, the TS network we used has excessive relationships and despite many of its edges are included in other networks (see Table 8), it might also contain weak or irrelevant relationships that do not influence group formation. On the other hand, researchers are either unaware of whom they share affinities with (and thus do not contact these people to form groups) or simply prefer to join people they already know regardless of affinity, as opposed to joining strangers with similar interests; assuming that the first case is true and even though results were poor using topical similarity, this kind of information has potential for collaboration recommendation or group reconfiguration.

We also calculated a Topic Similarity network based on Scopus subject categories, using journal subject categories as disciplines and using the same Eq. (8) for weighting the network. Nevertheless, 16 authors not publishing in Scopus indexed journals were missed in this network, and an even more dense network was obtained ($\delta = 0.779$). The ARI score of this network was 0.1 when compared against formal groups, lower than the score obtained by the TS network we used (0.171). The advantage of the topic similarity we used is that authors themselves chose the categories for each paper they registered in our database, whereas the Scopus subject category is derived from the journal where documents are published. This effect is accentuated by the fact that Scopus classifies journals into multiple categories.

With respect to organisational closeness (OC), a negative influence was found as well. The reason could be again structural or cultural. This network has the greatest amount of edges and introduces cliques produced by researchers working on the same department,

division, or campus. On the other hand, the reconfiguration of groups in our case study was directed by a nation-wide effort, which caused new groups to incorporate researchers from different campuses. Disciplinary affinity was more important than organisational/geographical closeness.

Another explanation for the behaviour of TS and OC relies on the fact that the networks related to both factors reflect potential collaborations. In that sense, perhaps it is not the structure of these networks that governs their behaviour (influence for team formation), but the semantics ("potential ties") that governs the structure (i.e. causes the excessive amount of edges, high density, high average degree, amount of connected components) and, at the same time, has an impact on the behaviour. As a result, we are observing that *previous ties* have a greater influence over team formation than *potential ties*, with different kinds of previous ties having different degrees of influence.

With respect to interactions, these seem reasonable as well. If FG, RC, and PC have shown a positive effect, their combination having better results comes as no surprise (we believe, nonetheless, that CC's individual score could not be improved because of its structural properties—more than the factor's strength itself). In that sense, if a group has been working together successfully, it is natural for the members of this group to try to stay together. The same applies to FG and RC, and FG, RC, and TH—even though TH is a weaker type of collaboration.

Relation with relevant literature. The work by Dahlander and McFarland (2013) states that, upon reflecting about the (good) quality of a relationship, people tend to keep that relationship (tie persistence). This agrees with our findings, since researchers who have successfully collaborated have formed new groups (evidenced by the CC network). Our findings also support team formation based on previous collaborations, which agrees with the works by Hinds et al. (2000), Hahn et al. (2006) and Bercovitz and Feldman (2011) for student, open software, and inventive teams—respectively. Our results, however, differ from works such as the one by Balland (2011), where institutional proximity showed to favour collaboration. We believe this could be due to the design of the network. Another explanation is that this network, as well as the TS network, mostly reflects potential collaborations. This view matches the work by Calero et al. (2006), where topical network clustering is admittedly used for potential research group identification.

Additional factors

Our discussion does not end with the studied factors. As previously mentioned, our results seem to suggest the presence of other factors for research group formation. We believe these factors include restrictions and policies, acquaintances, and informal work. For example, even when collaborations can be multidisciplinary, research groups are usually restricted by discipline. Also, we observed in our case study that research groups did not have more than one consolidated researcher (SNI 3) despite the existence of other consolidated researchers with a high topical similarity in the institution. At other Mexican universities, there are policies that force all faculty members to affiliate to research groups with the aim of stimulating research; however, we have observed that professors who are not researchers sometimes do not publish nor become connected with actual researchers (hence their representation as isolated vertices in networks). We believe that some of these policies hinder natural group formation and create structures that are arbitrary up to some point. As a result, it becomes difficult to predict research groups based on the information types that we used, which formally evidence natural teamwork and affinities.

In combination with policies are *acquaintances*, which are informal (unlike coauthor-ships or thesis committees, where a solid evidence is created) but could drive researchers towards gathering when there is a need to. Some examples of acquaintances include colleagues (not necessarily related with the researcher's specific area of interest), former teachers and classmates, people who work in the same building (do not necessarily belong to the same department), etc. There are also long-term acquaintances: persons working in the same building for years, researchers being former students of other researchers, being together in the same faculty for a lengthy period, knowing a researcher for years, etc. According to Casciaro and Lobo (2008), people tend to form work ties with whom they appreciate.

Result interpretation

Our first key result is that the CC network obtained the best ARI score. We believe this is due to the network representing successful previous collaborations, which stand for security, stability, and—in the case of this particular factor—quality. In that sense, researchers could feel drawn towards forming groups with colleagues that have helped to produce cited papers. This is represented in the CC network as a collection of punctual connections that give rise to specific structural properties that produce accurate clusters.

Our second key result is that hybrid networks surpass the ARI scores of pure networks (CC being an outlier). At the structural level, this could be due to the combination of collaboration evidence, which in turn could be interpreted as a multiplied attractiveness for team formation. For example, if a pair of researchers have been together in a group (FG), it is more likely for these researchers to be together in a new group if they have coauthored papers (RC, PC) or papers with cites (CC), or have co-participated in thesis committees (TH) as well. The previous collaboration simply becomes stronger. Our third key result is that FG has the most influence in group formation. As we will see in next section, this could be due to the influence this factor receives from other factors.

We expected TS and OC to have a positive influence on group formation as well, but this did not happen. As we previously explained, we believe this is due to the nature of what their corresponding networks reflect: potential collaborations. OC, for instance, represents the colleagues a researcher can relate to at a departmental level, but perhaps this is not attractive enough (or not as strong as a previous collaboration) to conform a research team. Furthermore, network design could also be playing a role in this result. For instance, despite the definition of Topical Similarity (TS) accounts for the production on all the topics where two authors coincide, it does not consider how focused or diversified both researchers' production is across disciplines or how aligned they are.

We also expected, in general, ARI results to be higher. Hence we attribute current results to the presence of other factors (which have been previously explained) that have not yet been taken into account.

Limitations

Derived from the design of our methodology, it can be noted that the factors to evaluate must have an appropriate network representation that encodes the relationship between each pair of researchers. In our case study the population of researchers was chosen from a single institution. However, for an inter-institutional collaboration recommendation, it would be necessary to delimit the population of researchers in the first place and then select factors for which information is available.

In our case study, we focused on identifying groups exclusively constituted by members of the Institution. We did not include external collaborators in network construction, despite the fact that there are collaborative links through them because: (1) there would be missing information for some networks, and (2) the clustering algorithm might produce partitions where external authors predominate and only a few internal researchers be represented.

Given that the information we used in our analysis is prior to group formation we can conclude that the decision of leaders and group members to join a group is a consequence (effect) of the collaborations they had in the past, their thematic affinity and their organisational / geographical proximity, discarding reverse causality. Nevertheless, the network representing joint membership on previous formal groups (FG) could be influenced by coauthorship (PC/RC), citations (CC) and thesis committees (TH) that occurred during the same period of time. This could explain why the FG network has the greatest influence over group conformation on the different analysis we made.

As observed by the results of Topical Similarity and Organisational Closeness, our methodology seems to be affected by the structural properties of the network used for capturing a given factor. Pure and hybrid networks including TS and OC show a small number of connected components and a density greater than 0.1.

It is also important to mention that our case study is based on group reconfiguration. Perhaps group formation from scratch could yield different results.

Finally, we confirmed our initial hypothesis that the combination of pure networks into hybrid networks would produce better predictions of group formation. In the top 10 networks (Table 5) can be found eight hybrid networks containing FG that outperform the pure network FG. Nevertheless we found a single pure network (CC) that obtained the best match with respect to the ground truth, but that in combination with other networks (e.g. five hybrid networks in the top 10) produced a lower fitness.

However, our results are not conclusive to the extent that although we tested all possible combinations of the selected factors, the weighting assigned to each factor was the same. For example, hybrid networks made up of two factors only had a weighting of 0.5 each. It remains to be determined whether there is a hybrid network with a different weighting that could obtain a better rating. The former can be achieved using bio-inspired computing methods, such as genetic algorithms or particle swarm optimization.

Conclusions

We introduced a novel methodology to identify the influence of actual and potential collaboration factors on the formation of formal research groups in an institution. We used a network representation of each factor, combined them and evaluated how similar partitions generated by graph clustering techniques were to actual groups.

After evaluating 127 combinations of factors we found that, in overall, joint membership on previous formal groups (FG) was the factor with the most positive influence on group formation (or reconfiguration of groups in our case study), followed by recent coauthorship (RC) and citations received in coauthored papers (CC). This could be explained by the fact that other analysed factors representing previous collaboration (coauthorship, co-participation in thesis committees and citations in coauthored papers) would already influenced this factor (reverse causality).

We confirmed our hypothesis by finding eight combinations of factors (hybrid networks) containing FG that outperformed the pure network. Nevertheless, the network CC, a novel representation of collaboration success that we introduced in this paper, outperformed all the other combinations of factors we evaluated. The efficiency of CC on group formation makes it worthy of further study.

The main limitation we found in our methodology was that it is susceptible to structural properties of networks used for capturing actual and potential collaboration factors. Results showed that potential collaboration networks, highly dense by definition, e.g. Topical Similarity (TS) and Organisational/Geographical Closeness (OC), occluded sparse networks despite they provide more precise information (e.g. FG, CC, RC).

We also found networks with weak influence on group formation. Past coauthorship (PC) and co-participation in thesis committees (TH) were present in five of the ten networks that better predicted group formation, but they were not good predictors overall.

We believe that the results obtained in this study would be very useful in establishing incentives for intra-institutional research collaboration. University administrators such as deans, provosts, chairmen, and other officials, could utilise these findings in the conformation of new research groups or the adjustment of existing ones. We believe that, even though real collaboration is finally a decision for researchers, their behaviour can be encouraged by appropriate research management policies. For example, by allocating internal funds to researchers who are organised around a strategic research line for the institution and whose funding depends on their joint performance (Ceballos et al. 2017). Likewise, by shedding light on group formation, evaluating agencies (which grant public funds, resources, and awards) can direct their assessment towards factors that reflect actual teamwork.

Future work

The set of factors used in the case study can be extended to detect additional factors that produce a better match with formal research groups. For instance, academic excellence of researchers, quantified through the number of papers, citations and weighted degree in coauthorship networks (Balland 2011), could be used through a proper representation in terms of homophily/heterophily networks. We could also introduce a network that counts the number of articles written in co-authorship with external authors to consider these collaborations even though the latter are not represented in the network. On the other hand, networks with anomalous structural properties could be reformulated to validate if this was actually the cause of the negative influence obtained in the results.

Additionally, an optimization algorithm could be used for finding a hybrid network with a combination of weights for each factor that produce a better match with formal groups. For instance, a combination $\alpha CC + \beta FG$, with $(\alpha, \beta) \neq 0.5$, that outperforms CC, where coefficients α and β would indicate to what extent each factor was taken into account.

References

Ancona, D. G., & Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science, 3*(3), 321–341.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. In: *Proceedings of the 12th ACM SIGKDD International*

Conference on Knowledge Discovery and Data Mining, KDD'06 (pp. 44–54). New York, NY: ACM. https://doi.org/10.1145/1150402.1150412.

Balland, P. A. (2011). Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies*, *46*(6), 741–756.

Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, *52*(3), 365–377. https://doi.org/10.1023/A:1014254214337.

Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, *40*(1), 81–93. https://doi.org/10.1016/j.respol.2010.09.008. Special section on Heterogeneity and University-Industry Relations.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. http://stacks.iop.org/1742-5468/2008/i=10/a=P10008.

Bonner, B. L., Baumann, M. R., & Dalal, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes*, *88*(2), 719–736.

Calero, C., Buter, R., Cabello Valdés, C., & Noyons, E. (2006). How to identify research groups using publication analysis: An example in the field of nanotechnology. *Scientometrics*, *66*(2), 365–376.

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, *46*(4), 823–847.

Cantu, F., & Ceballos, H. (2012). A framework for fostering multidisciplinary research collaboration and scientific networking within university environs. In J. Leibowitz (Ed.), *Knowledge Management Handbook: Collaboration and Scientific Networking, Chap 12* (pp. 207–217). Boca Raton: CRC Press.

Cantu, F., Ceballos, H., Mora, S., & Escoffie, M. (2005). A knowledge-based information system for managing research programs and value creation in a university environment. In *Proceedings of the Americas conference on information systems, AMCIS* (Vol. 2005, pp. 781–791).

Cantu, F., Bustani, A., Molina, A., & Moreira, H. (2009). A knowledge-based development model: The research chair strategy. *Journal of Knowledge Management*, *13*(1), 154–170.

Casciaro, T., & Lobo, M. S. (2008). When competence is irrelevant: The role of interpersonal affect in task-related ties. *Administrative Science Quarterly*, *53*(4), 655–684.

Ceballos, H., Fangmeyer, J., Galeano, N., Juarez, E., & Cantu-Ortiz, F. (2017). Impelling research productivity and impact through collaboration: A scientometric case study of knowledge management. *Knowledge Management Research and Practice*, *15*(3), 346–355.

Cho, P. S., Do, H. H. N., Chandrasekaran, M. K., & Kan, M. Y. (2013). Identifying research facilitators in an emerging Asian Research Area. *Scientometrics*, *97*(1), 75–97. https://doi.org/10.1007/s11192-013-1051-3.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*(066), 111. https://doi.org/10.1103/PhysRevE.70.066111.

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, *23*(3), 239–290. https://doi.org/10.1016/S0149-2063(97)90034-9. A special issue of the journal of management.

Cummings, J. N., & Kiesler, S. (2008). Who collaborates successfully?: Prior experience reduces collaboration barriers in distributed interdisciplinary research. In: *Proceedings of the 2008 ACM conference on computer supported cooperative work, CSCW'08* (pp. 437–446). New York, NY: ACM. https://doi.org/10.1145/1460563.1460633.

Dahlander, L., & McFarland, D. A. (2013). Ties that last tie formation and persistence in research collaborations over time. *Administrative Science Quarterly*, *58*(1), 69–110.

Etzkowitz, H. (2003). Research groups as 'quasi-firms': The invention of the entrepreneurial university. *Research Policy*, *32*(1), 109–121.

Etzkowitz, H., & Leydesdorff, L. (1998). The endless transition: A 'triple helix' of niversity industry government relations. *Minerva*, *36*(3), 203–208.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and 'mode 2' to a triple helix of university–industry–government relations. *Research Policy*, *29*(2), 109–123.

Garza, S. E., & Schaeffer, S. E. (2016). Local bilateral clustering for identifying research topics and groups from bibliographical data. *Knowledge and Information Systems*, *48*(1), 179–199. https://doi.org/10.1007/s10115-015-0867-y.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*(1), 1–15. https://doi.org/10.1006/obhd.1996.0061.

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702. https://doi.org/10.1126/science.1106340.

Hahn, J., Moon, J. Y., Zhang, C. (2006). Impact of social ties on open source project team formation. In *IFIP international conference on open source systems* (pp. 307–317). Springer.

Heredia, R. M., & Vinueza, P. C. (2015). A proposal model to monitor interdisciplinary research projects in Latin American Universities. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, *10*(3), 102–108.

Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational Behavior and Human Decision Processes*, *81*(2), 226–251. https://doi.org/10.1006/obhd.1999.2875.

Huckman, R. S., Staats, B. R., & Upton, D. M. (2009). Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Science*, *55*(1), 85–100. https://doi.org/10.1287/mnsc.1080.0921.

Johnson, N. F., Xu, C., Zhao, Z., Ducheneaut, N., Yee, N., Tita, G., et al. (2009). Human group formation in online guilds and offline gangs driven by a common team dynamic. *Physical Review E*, *79*(066), 117. https://doi.org/10.1103/PhysRevE.79.066117.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, *322*(5905), 1259–1262.

Kairam, S. R., Wang, D. J., & Leskovec, J. (2012). The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on web search and data mining, WSDM'12* (pp. 673–682). New York, NY: ACM. https://doi.org/10.1145/2124295.2124374.

Katz, J. S. (1994). Geographic Proximity and Scientific Collaboration. *Scientometrics*, *31*(1), 31–43.

Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. In *Proceedings of the sixth conference OD the international society for scientometric and informetric* (pp. 163–175).

Katzenbach, J. R., & Smith, D. K. (1993a). *The discipline of teams*. Massachusetts: Harvard Business Press.

Katzenbach, J. R., & Smith, D. K. (1993b). *The wisdom of teams: Creating the high-performance organization*. Massachusetts: Harvard Business Press.

Klein, J. T. (2008). Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine*, *35*(2), S116–S123.

Levine, J. M., & Moreland, R. L. (1991). Culture and socialization in work groups. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 257–279). Washington: American Psychological Association.

Leydesdorff, L. (2013). Triple helix of university-industry-government relations. *Encyclopedia of creativity, innovation and entrepreneurship* (pp. 1844–1851). Springer.

Liang, L., & Zhu, L. (2002). Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics*, *55*(2), 287–316. https://doi.org/10.1023/A:1019623925759.

Margolin, D., Ognyanoya, K., Huang, M., Huang, Y., & Contractor, N. (2012). Team formation and performance on Nanohub: A network selection challenge in scientific communities. In B. Vedres & M. Scotti (Eds.), *Networks in social policy problems* (pp. 80–100). Cambridge: Cambridge University Press.

Martin, B. R. (2012). Are universities and university research under threat? Towards an evolutionary model of university speciation. *Cambridge Journal of Economics*, *36*(3), 543–565.

Martin, T., Ball, B., Karrer, B., & Newman, M. E. J. (2013). Coauthorship and citation patterns in the Physical Review. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. https://doi.org/10.1103/PhysRevE.88.012814, arXiv:1304.0473.

Martín-Sempere, M. J., Garzón-García, B., & Rey-Rocha, J. (2008). Team consolidation, social integration and scientists' research performance: An empirical study in the biology and biomedicine field. *Scientometrics*, *76*(3), 457–482. https://doi.org/10.1007/s11192-007-1866-x.

Mattessich, P. W., & Monsey, B. R. (1992). Collaboration: What makes it work. In *A review of research literature on factors influencing successful collaboration*. Minnesota: ERIC.

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*, 0104209v1.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(Suppl 1), 5200–5205. https://doi.org/10.1073/pnas.0307545100.

Ordóñez-Matamoros, H. G., Cozzens, S. E., & Garcia, M. (2010). International co-authorship and research team performance in Colombia. *Review of Policy Research*, *27*(4), 415–431.

Owens, D. A., Mannix, E. A., & Neale, M. A. (1998). Strategic formation of groups: Issues in task performance and team member selection. *Research on Managing Groups and Teams*, *1*, 149–165.

Palla, G., Barabási, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, *446*(7136), 664–667.

Perianes-Rodriguez, A., Chinchilla-Rodriguez, Z., Vargas-Quesada, B., Olmeda-Gomez, C., & Moya-Anegon, F. (2009). Synthetic hybrid indicators based on scientific collaboration to quantify and evaluate individual research results. *Journal of Informetrics*, *3*(2), 91–101.

Putnam, L. L. (1992). Rethinking the nature of groups in organizations. *Small Group Communication: A Reader*, *6*, 57–66.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., & Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 2658–2663. https://doi.org/10.1073/pnas.0400054101.

Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, *76*(036), 106. https://doi.org/10.1103/PhysRevE.76.036106.

Reagans, R., Zuckerman, E., & McEvily, B. (2004). How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly*, *49*(1), 101–133.

Rey-Rocha, J., Garzón-García, B., & Martín-Sempere, M. J. (2006). Scientists' performance and consolidation of research teams in Biology and Biomedicine at the Spanish Council for Scientific Research. *Scientometrics*, *69*(2), 183–212.

Rey-Rocha, J., Garzón-García, B., & José Martín-Sempere, M. (2007). Exploring social integration as a determinant of research activity, performance and prestige of scientists. Empirical evidence in the Biology and Biomedicine field. *Scientometrics*, *72*(1), 59–80. https://doi.org/10.1007/s11192-007-1703-2.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American Sociological Review*, *68*, 195–222.

Sam, C., & Van Der Sijde, P. (2014). Understanding the concept of the entrepreneurial university from the perspective of higher education models. *Higher Education*, *68*(6), 891–908.

Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, *3*(1), 1–16.

Schaeffer, S. E. (2005). *Stochastic local clustering for massive graphs* (pp. 354–360). Berlin, Heidelberg: Springer. https://doi.org/10.1007/11430919_42.

Shah, P. P., & Jehn, K. A. (1993). Do friends perform better than acquaintances? The interaction of friendship, conflict, and task. *Group Decision and Negotiation*, *2*(2), 149–165. https://doi.org/10.1007/BF01884769.

Sonnentag, S., & Volmer, J. (2009). Individual-level predictors of task-related teamwork processes: The role of expertise and self-efficacy in team meetings. *Group & Organization Management*, *34*(1), 37–66.

Stokols, D., Misra, S., Moser, R. P., Hall, K. L., & Taylor, B. K. (2008). The ecology of team science: Understanding contextual influences on transdisciplinary collaboration. *American Journal of Preventive Medicine*, *35*(2), S96–S115.

Sun, Y., & Liu, K. (2016). Proximity effect, preferential attachment and path dependence in inter-regional network: A case of China's technology transaction. *Scientometrics*, *108*(1), 201–220. https://doi.org/10.1007/s11192-016-1951-0.

Tomas-Folch, M., Mentado-Labao, T., & Ruiz-Ruiz, J. M. (2015). Las buenas prácticas en gestión de la investigación de las universidades mejores situadas en los rankings. *Archivos Analíticos de Políticas Educativas*, *23*(105), 1–23.

Uddin, S., Hossain, L., & Rasmussen, K. (2013). Network effects on scientific collaborations. *PLOS ONE*, *8*(2), e57546. https://doi.org/10.1371/journal.pone.0057546.

Verbree, M., Horlings, E., Groenewegen, P., Van der Weijden, E., & van den Besselaar, P. (2015). Organizational factors influencing scholarly performance: A multivariate study of biomedical research groups. *Scientometrics*, *102*(1), 25–49.

Volmer, J., & Sonnentag, S. (2011). The role of star performers in software design teams. *Journal of Managerial Psychology*, *26*(3), 219–234.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 366–391.

Waltman, L., & Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the Association for Information Science and Technology*, *63*(12), 2378–2392.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.

Yu, Q., Shao, H., & Duan, Z. (2011). Research groups of oncology co-authorship network in China. *Scientometrics*, *89*(2), 553–567.

Zhang, Y., Chen, K., Zhu, G., Yam, R. C. M., & Guan, J. (2016). Inter-organizational scientific collaborations and policy effects: An ego-network evolutionary perspective of the Chinese Academy of Sciences. *Scientometrics*, *108*(3), 1383–1415. https://doi.org/10.1007/s11192-016-2022-2.

Zhu, M., Huang, Y., & Contractor, N. S. (2013). Motivations for self-assembling into project teams. *Social Networks*, *35*(2), 251–264. https://doi.org/10.1016/j.socnet.2013.03.001. Special issue on advances in two-mode social networks.