Scientometrics
DOI 10.1007/s11192-013-1162-x

Atapuerca: evolution of scientific collaboration in an emergent large-scale research infrastructure

Sergi Lozano · Xosé-Pedro Rodríguez · Alex Arenas

Received: 10 May 2013

Abstract We study the evolution of scientific collaboration at Atapuerca's archaeological complex along its emergence as a large-scale research infrastructure (LSRI). Using bibliometric and fieldwork data, we build and analyze co-authorship networks corresponding to the period 1992–2011. The analysis of such structures reveals a stable core of scholars with a long experience in Atapuerca's fieldwork, which would control coauthorship-related information flows, and a tree-like periphery mostly populated by 'external' researchers. Interestingly, this scenario corresponds to the idea of a *Equipo de Investigación de Atapuerca*, originally envisioned by Atapuerca's first director 30 years ago. These results have important systemic implications, both in terms of resilience of co-authorship structures and of 'oriented' or 'guided' self-organized network growth. Taking into account the scientific relevance of LSRIs, we expect a growing number of quantitative studies addressing collaboration among scholars in this sort of facilities in general and, particularly, emergent phenomena like the Atapuerca case.

Keywords Emergent large-scale research infrastructures · Scientific collaboration dynamics · Co-authorship · Network analysis · Atapuerca

---

Sergi Lozano and Alex Arenas are members of the COMPATHEVOL Research Group.

S. Lozano (✉) · X.-P. Rodríguez · A. Arenas
IPHES, Institut Catala de Paleoecologia Humana i Evolució Social, 43007 Tarragona, Spain
e-mail: slozano@iphes.cat

S. Lozano · X.-P. Rodríguez
Area de Prehistoria, Universitat Rovira i Virgili (URV), 43002 Tarragona, Spain

A. Arenas
Departament dEnginyeria Informatica i Matematiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain

A. Arenas
Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Saragossa, Spain

Introduction

Large-scale (or 'world-class') research infrastructures (LSRIs) are complex scientific projects imperative to high-quality, innovative and competitive science (European Commission et al. 2010; OECD 2010). Research associated to this kind of facilities presents three relevant particularities: (a) it addresses key fundamental questions (which justify huge efforts and investment), (b) involves many scientists and technicians from different disciplines, affiliations and countries, and (c) is based on singular data sources and/or unique scientific tools. As a consequence, LSRIs provide a unique potential for the assessment of scientific collaboration dynamics. They are perfect scenarios for the study of general questions about scientific collaboration (e.g. multidisciplinary and multinational collaboration), as well as specific ones concerning research facilities (like patterns of community emergence or the co-evolution of physical and remote collaboration).

In spite of this uniqueness as case studies, LSRIs have received little attention in the literature on scientific collaboration (see Katz and Martin 1997; Havemann et al. 2001, for reviews). Articles in this literature have treated several topics relevant to LSRI such as social and organizational aspects (Hagstrom 1975; Beaver and Rosen 1978, 1979; Hara et al. 2003), physical proximity and informal communication among scholars (Kraut et al. 1988; Katz 1994; Ponds et al. 2007), or scientific collaboration based on management of singular resources or unique scientific data (Wray 2002; Chin et al. 2002). However, to our knowledge, not much have been written about the particularities of collaboration processes taking place in LSRI. This is specially true in the case of the sort of research infrastructures we are interested in this work, namely archaeological sites.

An archaeological or paleoanthropological site is an obvious example of 'single-sited' research infrastructure. Archaeologist and scholars in related disciplines develop their fieldwork in concrete physical locations, which are defined by the empirical record they are excavating and studying. In most cases, archaeological sites are exploited by a modest team of researchers. However there are a few exceptions presenting a huge impact (both in terms of international scientific relevance and volume of allocated resources), which deserve the label of *large-scale*.

One of these singular cases in paleoanthropology is the complex located at Sierra de Atapuerca (Spain). It includes up to 13 sites within an area of no more than 20 km^2, and covers a time spanning between about 1.22 and 0.01 Ma. For almost three decades, it has been the scenario of key findings that have given rise to some of the most important questions about early human evolution in Europe. The oldest human fossils in Europe, dating back to 1.22 Ma ago, have been found in the Sima del Elefante archaeological site (Carbonell et al. 2008). In Gran Dolina site the Atapuerca research team also found human fossils, dated to around 0.9 million years ago (Carbonell et al. 1995). These fossils were classified as a new species: *Homo antecessor* (de Castro et al. 1997). Furthermore these fossils show the, to our knowledge, earliest evidence of cannibalism (Fernández-Jalvo et al. 1996; Carbonell et al. 2010). In the site of Sima de los Huesos more than 6,500 human fossils has been recovered (up to the date the largest sample of human fossils of an extinct species) belonging to at least 28 individuals of *Homo heidelbergensis* (ancestors of Neandertals; Arsuaga et al. 1993, 1999).

This increasing scientific impact has translated into a progressively larger and more complex organizational infrastructure. Along the latest 20 years, Atapuerca has been the temporary working place of up to 945 scholars from 33 different nationalities, collaborating under the supervision of the three co-directors (i.e. Arsuaga, Bermúdez de Castro and Carbonell). Such a long experience of scientific work has lead to the foundation of several scientific institutes, like CENIEH[1], IPHES[2] and the *Centro UCM-ISCIII de Evolución y Comportamiento Humanos*. Moreover other public facilities related to dissemination activities have been established, the *Museo de la Evolución Humana*[3] and *Fundación Atapuerca*[4] being the two most remarkable examples.

Finally, this process has been accompanied (and, to some extend, fueled) by an impressive volume of related popular science publications (the three directors alone have authored around 25 during the period under study in this article; Hochadel 2013a).

This article focuses on the evolution of scientific collaboration at Atapuerca along its emergence as an archaeological LSRI. Specifically, we are interested on the interplay between co-authorship and participation in fieldwork (understood as a proxy of physical collaboration). Starting from bibliometric and fieldwork data, we analyze co-authorship of scientific papers related to Atapuerca from a structural viewpoint. More concretely, we have built and analyzed co-authorship networks corresponding to four successive time windows within the period under study (i.e. 1992–2011).

Despite its limitations (Katz and Martin 1997), article co-authorship networks analysis is a usual method to study scientific collaboration (Melin and Persson 1996; Glänzel and Schubert 2005a). Among several aspects, it has been used to study the role of physical distance and other sorts of spatial biases (Glänzel 2001; Kretschmer 2004; Glänzel and Schubert 2005b; Frenken et al. 2009; Perc 2010; Hennemann et al. 2012), disciplinary particularities (Liu et al. 2005; Hou et al. 2008), as well as multidisciplinarity and performance of research teams (He et al. 2013; Wuchty et al. 2007; Jones et al. 2008; Franceschet and Costantini 2010; Börner et al. 2005). Moreover, co-authorship networks have been commonly used in Science of Networks to illustrate structural signatures (Newman 2003) and to test analysis methodologies (Girvan and Newman 2002).

Focusing on the evolution of scientific collaboration, we find some recent contributions analyzing general patterns of structural dynamics in this sort of networks (Viana et al. 2013; Mali et al. 2012), following a research line that started a decade ago (Barabási et al. 2002). However, much fewer studies have centered on concrete physical scenarios (Perc 2010), as it is our case in this article.

Our results can be summarized as follows. A giant component (composed, basically, by the three directors of Atapuerca and their closer collaborators) has emerged during the period (from a third of the network by the end of 1996, to almost 70 % in 2011). A closer look to such a growing giant component reveals a highly modular structure, with a core composed mainly by researchers with a long experience in Atapuerca's fieldwork, and a tree-like periphery mostly populated by researchers without field experience. Moreover, focusing on the intermediation between communities, we find a steady tendency to unequal betweenness centralities, indicating that few researchers are responsible for most of the collaboration (in terms of publication) linking the core and the periphery.

We conclude that our results correspond to the process of consolidation of a stable, multidisciplinary and Spanish-based research team (i.e. the *Equipo de Investigación de Atapuerca*, EIA), which was planned from the beginning and intentionally promoted by the three directors. This fact leads to additional questions in the line of the "guided" self-organization, and the long-term sustainability, of Atapuerca-related scientific collaboration.

[1] *Centro Nacional de Investigación sobre la Evolución Humana* (*CENIEH*), http://www.cenieh.es/.

[2] *Institut Català de Paleoecologia Humana i Evolució Social* (*IPHES*), http://www.iphes.cat/.

[3] http://www.museoevolucionhumana.com.

[4] http://www.atapuerca.org/.

From our viewpoint, this contribution sets a first step towards a better assessment of scientific collaboration processes in LSRIs in general, and emergent phenomena as Atapuerca in particular.

Data and methodology

Datasets

This work is based on two datasets, namely (a) a list of researchers having participated in at least one campaign at Atapuerca's complex, and (b) ISI bibliometric records of all articles citing at least one of the TOP-20 most cited works about Atapuerca[5] (see the "Appendix" section for a comprehensive list). Details about both datasets follow.

Participation in fieldwork

A list was composed from the compilation of all excavation records corresponding to the period 1992–2011. For each individual participating in Atapuerca's fieldwork (except few cases), the list provides the following details: alphanumerical ID, complete name, birth date, gender, inviting institution, city of origin (specially relevant for Spanish researchers), nationality and participation years (in most cases, including duration of the stay in days).

Bibliometric data

The dataset consists of 1,640 records, corresponding to all articles (in ISI-JCR indexed journals) citing any of the TOP-20 within the period [1992, 2011]. Their distribution by publication year is provided in Table 1. Each record contains bibliographic data such as author names, title, abstracts and keywords. Authors' affiliations are not provided.

The criteria followed to build this dataset (and, in particular, to compose Atapuerca's TOP-20 list) requires a more detailed explanation, as we did not find any precedent in the scientometric literature to base on.

Our aim was to use this dataset to assess the impact of Atapuerca-related research over scientific collaboration dynamics, both internally and with the rest of the scholar community. To this end, first we used Thomson's WEB of Knowledge search tool to identify all papers related to Atapuerca (i.e. 'Atapuerca' as their topic), which were published between 1992 and 2011. Then we took out from the list those ones not 'Made in Atapuerca' (i.e. not reporting empirical findings in the complex or introducing related theories). Finally, we ranked them by number of citations and chose the 20 most cited. This way, our TOP-20 list provides an overview of the most relevant moments in Atapuerca's "research history". Once we had this list, we simply collected the articles citing them as a way to visualize the impact of these key moments on the related literature.

Notice that adopting an alternative, less supervised strategy based on keywords would have probably leaded to partial results. This is because, besides 'Atapuerca', a complete keyword list should have to include a plethora of diverse elements such as all the names of individual archaeological sites (e.g. Sima de los Huesos, Gran Dolina, and so on) and concepts forged at Atapuerca (such as 'Homo antecessor'), which were introduced by articles that can actually be found in our TOP-20.

[5] According to their citation count by May 2012.

Table 1 Some characteristics of the bibliometric dataset

Distribution by periods of articles in the TOP-20 list and the bibliometric dataset, as well as the average number of authors per paper

Periods	Publications in TOP-20	Publications in dataset	Average number of authors
1992–1996	4	53	2.39
1997–2001	12	426	3.10
2002–2006	2	319	4.21
2007–2011	2	842	5.34

Matching names

Combining the above datasets is one of the innovative aspects of this work. Such a combination required matching author names from publications with researcher names from excavation records. All the operations described in the following were implemented by means of Python scripts.[6]

Our name matching procedure was based on two starting assumptions, namely:

– Researcher identities in excavation records were unique and differentiated.
– Since the goal was to build co-authorship networks, all author names (but not all names in excavation records) were relevant.

Consequently, we took each of the authors' names in the bibliometric dataset and tried to assign them researcher identities from the excavation record. Specifically, this was achieved by following three steps:

(1) Author names were extracted from the 1,640 records to a single list ordered alphabetically. Such a list was checked manually in order to unify divergent expressions of the same author names. A paradigmatic example of this kind of cases is *José María Bermúdez de Castro*, one of Atapuerca's directors, whose name was written in up to four different ways in the list. Notice that this manual checking could not solve author name disambiguation (i.e. different authors with the same name). However we must recall that the potential impact of name ambiguity on our study is limited, since the scholar community involved was very specific.
(2) Author names in the list resulting from the previous step were matched against first names and first family names in the excavation record (notice that a second family name is used in several countries, including Spain). When this matching was possible, authors were assigned the corresponding excavation alphanumerical ID. Otherwise, the author name was sent to a separate list of items pending assignment.
(3) The composition of this second list revealed that almost all cases of failed matching were related to the treatment of the two family names of Spanish authors. An author could be identified, for instance, only by the first family name, with both of them hyphened or with the first one treated as middle name. Solving this required a more flexible matching strategy, beyond the simple comparison in the previous point. We adopted a solution based on the *Sequencematcher* class of Python's *difflib* module.[7] Given two strings, this class allows to run a number of comparison algorithms and to obtain a numerical measure of similarity between them. Using this class, we compared the unmatched author names with identities in the fieldwork dataset, and set matchings with a reasonable high similarity score. Author that were not assigned

6 http://www.python.org/.

7 See http://docs.python.org/2/library/difflib.html.

any excavation ID after this step, were considered scholars not having participated in Atapuerca fieldwork and given a new unique ID (see "Construction of co-authorship networks" section for more details on unique ID assignation).

Construction of co-authorship networks

The bibliometric dataset on Atapuerca-related work has been used to build four co-authorship networks corresponding to consecutive 5-years time windows within the period under study: [1992, 1996], [1997, 2001], [2002, 2006] and [2007, 2011]. Vertices in these networks correspond to authors of articles having cited at least one reference in Atapuerca's TOP-20 (see "Appendix" section), and weighted links among them represent the number of papers they have co-authored during the corresponding period.

Notice that this time division separates different stages in the emergence of Atapuerca as an paleoanthropological LSRI, and that 5 years is a long-enough time interval to capture relevant publication activity. Moreover, it is worth mentioning that co-authorship networks are built from scratch every time, so links are not accumulated from one time window to the following one. This way, we make sure that the structure captures the particularities of co-authorship during the corresponding period.

Finally, in order to make it possible to track individual authors along the four temporal intervals, each vertex has been assigned a unique ID. Such numerical IDs have also been used to codify information about authors' participation in Atapuerca's fieldwork. More concretely, values below 2,000 correspond to authors that had collaborated in Atapuerca's field works during the interval corresponding to the network or before. Conversely, values above 5,000 indicate that an author has never joined any of Atapuerca's teams. The range of values in between 2,000 and 5,000 were used as temporal IDs, corresponding to authors who would collaborate at some point in time posterior to the network one.

Network analysis

Co-authorship networks obtained above were analyzed using the *igraph* library of R statistical package.[8] In the following, we provide descriptions of the structural features analyzed in our study.

Basic structural characteristics

Among other basic structural characteristics, our study focused on those listed bellow. More extended explanations and examples can be found in any manual of social networks analysis like, for instance (Wasserman and Faust 1994).

- The *Average degree* is the number of neighbors of a vertex (other nodes it is connected to) on average. In our case, it corresponds to the average amount of co-authors a scientist has.
- The *Clustering coefficient* measures the ratio of closed triads in the network, i.e. cases where two neighbors of a node A (B and C) are also connected. It is a very important characteristic in social networks analysis, as it measures to what extend social actors represented by nodes are clustered together in densely connected groups or not. In a co-

[8] See http://igraph.sourceforge.net/.

authorship network, high clustering coefficient values correspond to groups of authors usually publishing with the same close list of scholars.

- In graph theory, the geodesic distance between two nodes is the number of edges in the shortest path linking them. The *Average* (*geodesic*) *path length*measures the average of such a minimum path in the network. In a co-authorship network, this measure provides a proxy of 'closeness' among authors.

Assortative mixing

Generally speaking, assortative mixing is the level of connectivity existing among nodes in the network with similar characteristics. The higher it is, the greater the bias towards connections between similar nodes and, therefore, the worse mixed is the population (i.e. the more separated are the different groups in it). It's value lies in between 1 and -1, which correspond to the extremal cases (i.e. completely separated and completely mixed groups, respectively). A value equal to 0 would correspond to a mixing resulting from a random matching of nodes.

In social networks analysis it is typical to calculate the assortative mixing (also called homophily) according to attributes such as race, gender or age. In our study, we have focused on the mixing degree among scientists involved or not in Atapuerca's fieldwork.

Betweenness centrality

Betweenness centrality measures the extent to which a vertex lies on paths between other vertices (i.e. its intermediation power; Freeman 1977).

In its *classical* form, it is obtained by computing the shortest paths connecting each possible pairs of nodes in the network, and calculating how often does a node belong to one of them. For our study, we have also computed a variant of the measure specifically thought to quantify intermediation power of individuals bridging populations of researchers with and without fieldwork experience in Atapuerca. In such a variant, we only take into consideration those paths connecting one author of each type.

Results

Co-authorship networks

Figure 1 shows the resulting networks ordered chronologically left–right and top-down. In all four cases we observe a giant component with strong collaborations and a collection of smaller subgraphs (dyads and triads in most cases) with thinner links. As one would expect, the giant component is composed by the three directors of Atapuerca, their closer collaborators and other coauthors they had during the period. The rest of the network basically corresponds to different research teams that have also cited ATAs TOP-20 separately.

When comparing across these networks, we can easily see the growth of the giant component. Notice that this growth does not correspond to the cumulation of links along periods (since networks are built from scratch every period, see "Construction of co-authorship networks" section), but to an actual increase of scientific collaboration around Atapuerca findings. As a consequence, this comparative analysis should reveal the characteristics of the underlying collaboration processes.

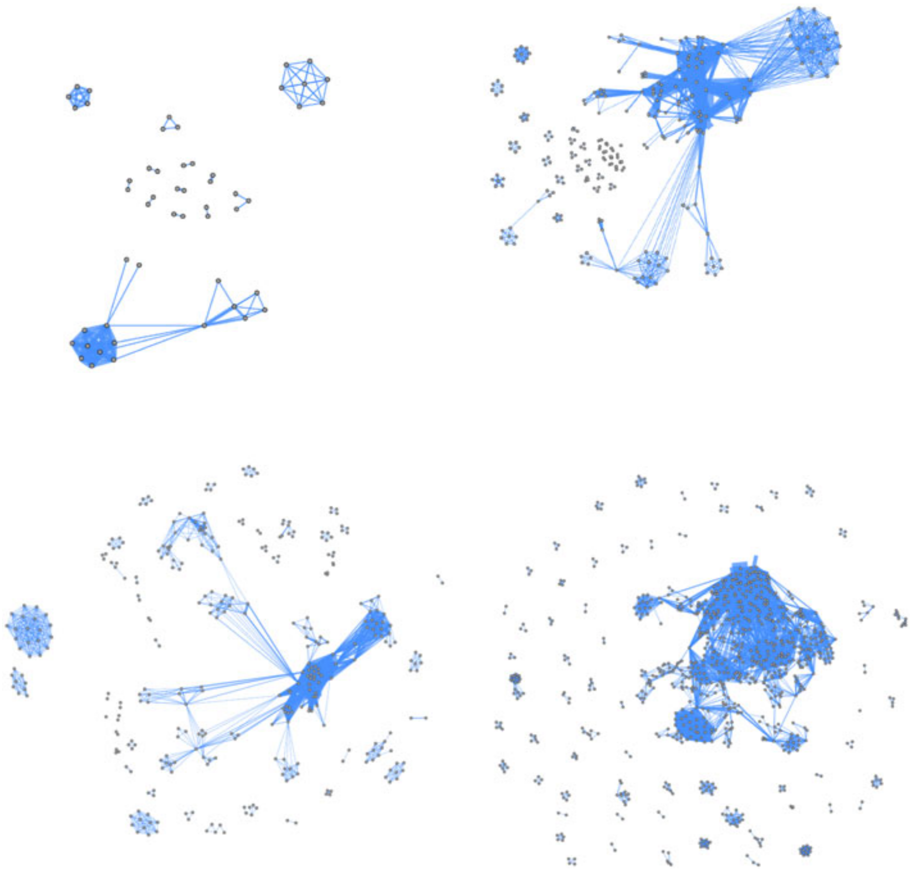

Fig. 1 Co-authorship networks corresponding to the four periods (*top–bottom* and *left–right*): 1992–1996, 1997–2001, 2002–2006, 2007–2011. *Nodes* correspond to researchers and a link between two of them indicates co-authorship of, at least, one paper (the actual number of collaborations in represented by the thickness, which is proportional to the number of collaborations

A more detailed view of the evolution of the giant component is provided in Fig. 2. Participation in fieldwork is also represented in these figures by the gray-scale vertices' coloring (i.e. black meaning participation in all the campaigns up to the end of the time interval corresponding to the network, and white corresponding to zero field experience at Atapuerca). Two main observations can be made. First, the resulting topology is strongly modular (i.e. nodes are organized in communities much more densely connected internally than among them), and such modules are highly homogeneous in terms of coloring. Second there is a dense core dominated by black nodes, and a tree-like periphery mainly populated by white ones.

Network visualization is a good starting point, but studying more closely the growth of the giant component and the relationship between fieldwork participation and co-authorship, requires a quantitative analysis of these networks. Table 2 provides some details about the networks under study. More concretely, for each network we have computed the following information: the size of the giant component (absolute and relative to the size of the whole network), the average degree (so, the average number of collaborators within the time interval),

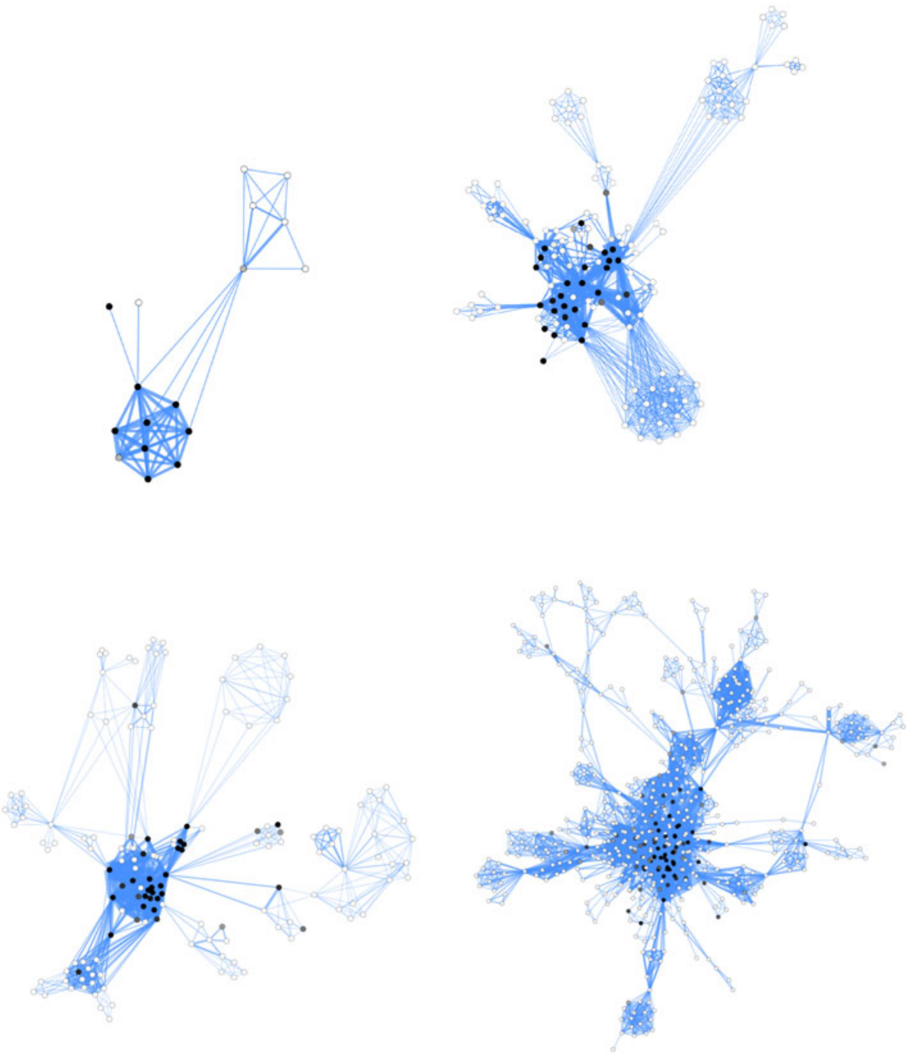

Fig. 2 Giant components of the four co-authorship networks (ordered as in Fig. 1). *Gray-scale color* of nodes represents authors' degree of involvement in fieldwork (*black* all campaigns, *white* not having participated at all)

clustering coefficient and average path length in the giant component, as well as its composition in terms of the proportion of scholars having participated in Atapuerca's fieldwork.

By taking a look to these results, we can confirm our observations and identify some trends across periods. The giant component grows from a third of the whole network in the first period to represent most of it (almost 70 %) by the end, possibly signaling the emergence of a 'community' or 'invisible college' (Crane 1972; de Price 1986; Zuccala 2006) around the work developed at Atapuerca. However, the composition of researchers actually having contributed to fieldwork at the site has diminished monotonically to a 16 %. This is not strange, taking into account that time and resources' constraints affecting fieldwork are stronger that those in paper writing collaborations.

Table 2 Structural features of the four giant components under study

Periods	N_{GC}	$\frac{N_{GC}}{N}$	Average degree	Clustering coefficient	Average path length	Percent at Atapuerca's fieldwork
1992–1996	17	0.33	6.35	0.77	1.86	59
1997–2001	139	0.55	11.92	0.85	2.94	25
2002–2006	151	0.44	10.54	0.88	3.64	22
2007–2011	570	0.67	13.05	0.83	4.23	16

Descriptions are provided in "Network analysis" section. The size of the giant component (absolute and relative to the size of the whole network), the average degree (so, the average number of collaborators within the time interval), clustering coefficient and average path length in the giant component, as well as its composition in terms of the proportion of scholars having participated in Atapuerca's fieldwork

Table 3 Connectivity patterns within the four giant components: average degree and assortative mixing

Periods	Average degree ATA	Average degree NOT ATA	Assortative mixing
1992–1996	7.9	4.14	0.29
1997–2001	17.22	10.14	0.25
2002–2006	17.76	8.44	0.37
2007–2011	24.78	10.82	0.38

Descriptions are provided in "Network analysis" section. Scholars having participated in Atapuerca's fieldwork present, on average, a higher average degree (i.e. more co-authors). Homophily is high

In order to shed further light on the influence of fieldwork participation over information diffusion and collaboration dynamics inside the (growing) giant component, we have computed the average degree separately for participants and non-participants, as well as the assortative mixing (see Table 3).

A priori, the strong growth of the giant component would provide authors in it with higher possibilities for collaboration with other members, as co-authors tend to share information about their work. In our case, the average degree does tend to increase from one period to the following one. However this increase is much slower than that of the size of the giant component. Actually, it could just correspond to a general tendency for papers with increasing number of authors. If we compare the evolution of average degrees in Table 3 with that of the average number of authors per paper in our bibliometric dataset (see Table 1), we observe that their growth rates are similar.

This apparent lack of influence of the giant component's growth can be explained as a result of the strong modularity of these networks. Such a structural feature would have limited the diffusion of information to different parts of the giant component, thus preventing further collaborations among groups. The observed positive (and increasing) values of assortative mixing in all four networks would support this hypothesis. Scholars who had participated in Atapuerca's fieldwork (in most cases, for several years) would have promoted (and strengthen) stable collaborations among them. Moreover, new coauthors without field experience in Atapuerca would have preferentially joined communities growing at the periphery of the network composed basically by scientists also without field experience. Notice that this influence of homophily on co-authorship decisions has already been reported in other studies addressing the dynamics of scientific collaboration (Ding 2011).

These results justify taking a closer look to structural features related to information diffusion and, more specifically, to intermediation between these communities.

Betweenness

Betweenness centrality quantifies the intermediation power of a vertex in a network (Freeman 1977). Consequently, it has been used as an indicator of the role played by an individual sitting in that position concerning processes of information diffusion (Burt 2004) and network growth (Abbasi et al. 2012).

In our case, besides obtaining the *classical* betweenness for all nodes in the giant component of our networks, we have calculated a variant of the measure specifically thought to quantify intermediation power of individuals bridging populations of researchers having and having not collaborated in Atapuerca. This way, individuals with the highest values of this modified betweenness should correspond to those in privileged positions in terms of intermediation between the core (mostly black in Fig. 2) and peripheric communities. We have obtained the TOP-10 ranking of researchers at each network according to partial betweenness. Not surprisingly, Atapuerca's three directors are present in all four lists. Moreover, all other members of these lists (except one case in the first period) had also participated in Atapuerca's fieldwork. These observations highlight the role of fieldwork participation on the control of information flows and scientific collaboration in Atapuerca as a particular case in archeology, but also probably in other LSRI (see the "Discussion and conclusions" section).

Beyond identifying key individuals, one can obtain more information about the influence of the structure on diffusion processes by making a macroscopical reading of betweenness values. Such a macroscopic view is provided by Fig. 3, which shows how this magnitude is distributed across the four networks. Partial betweenness values are not uniformly but skewedly distributed (i.e. few individuals show high values while many have low ones). Such a distribution indicates that interactions with external researchers are channeled through a limited number of individuals, while many coauthors usually participating in Atapuerca's field work tend to collaborate in a quite endogamous way.

This unequal distribution is not limited to collaborations with external researchers, but general (i.e. classical betweenness' distribution is also skewed). These results are coherent with the scenario of high modularity observed: collaborations are normally restricted to more or less closed and homogeneous groups, which are eventually bridged by a small set of individuals (i.e. those with a high betweenness).

Finally, also notice that distributions corresponding to the later three periods present very similar plots. One could expect collaborations to emerge among researchers usually attending Atapuerca's summer campaigns, even if they belong to different groups. In that case, betweenness' distributions would tend to get flatter as time goes by. However this is not the case. Why? What mechanism makes this distribution to stay stable despite the strong growth of the network?

Discussion and conclusions

In order to get a better understanding of our findings, we analyzed them from the perspective of the personal experience of one of the authors (i.e. X. P. Rodríguez), as well as of other colleagues having participated in Atapuerca fieldwork since the early 1990s. They concluded that the emergence of a dense core dominated by black nodes (i.e. researchers

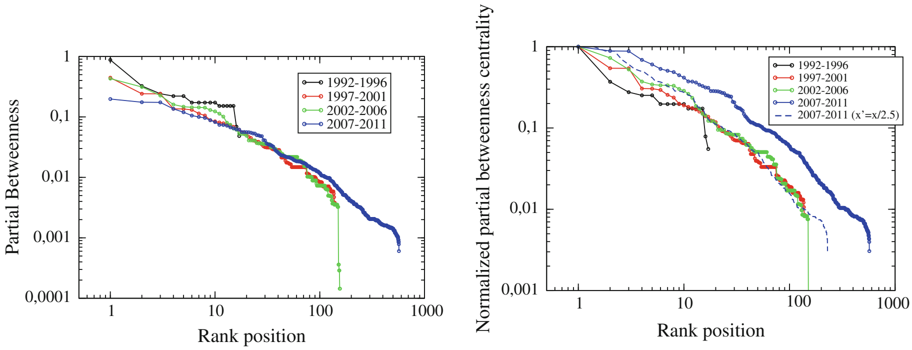

Fig. 3 Distribution of partial betweenness for all networks. Absolute (*left*) and normalized (*right*) values are provided for comparison. Notice the nice coincidence of distribution curves (in some cases, with some rescaling indicated with *dashed lines*)

with long experience in Atapuerca) corresponded to the process of consolidation of a stable, multidisciplinary and Spanish-based research team around Atapuerca's archaeological complex. Such a team, usually named EIA (Hochadel 2013a, b), have attracted the attention of 'external' collaborators (in, paleoanthropology, for instance) and received support by specialists in fields or techniques not well represented in the EIA (e.g. geochronology). These two later groups of scholars would correspond to peripheral, and colored in gray or white, nodes in our networks.

The central role of fieldwork observed in our case is clearly related to the uniqueness of empirical data (i.e. the archaeological record), and the costly procedures to obtain them. This particularity (which is a commonality in archeology-related disciplines), will be probably found also in other LSRIs, since they are usually built as huge scientific tools to perform experiments that cannot be developed anywhere else.

Interestingly, the idea of creation of the EIA was apparently planned from the beginning by Atapuerca's first director (Emiliano Aguirre) and intentionally promoted by the current three directors. Taking into account that co-authorship is (in most cases) a de-centralized phenomenon, basically driven by individual interactions among authors, it is surprising to see how a pre-designed organizational pattern could succeed. This leads to additional questions about the mechanisms set by the three directors to 'guide' or 'orient' the (in principle) self-organized emergence of the scientific collaboration structure.

Finally, the results of our structural analysis brings the discussion towards the robustness or resilience of Atapuerca's collaboration network. Several studies in Science of Networks have analyzed how removing a structural element of a system (intentionally or accidentally) can affect its functionality and survival (see Albert and Barabási 2002; Newman 2010 for examples in different kind of systems). In our particular case, the key intermediation role of a bunch of scholars occupying central positions (with the three directors as their paradigmatic example) could be seen as a structural debility, since the eventual 'disappearance' or 'removal' of these agents could disconnect peripheral communities, then affecting collaboration processes.

The structural analysis presented in this work opens the door to many other studies on the evolution of scientific collaboration along Atapuerca's emergence. Just to mention three possible extension lines, we envision: (a) additional works focusing on the dynamics that have shaped Atapuerca's co-authorship structures (e.g. the role of disciplinary protocols on inclusion and ordering of co-authors as micro-macro links Coleman 1990, aligning self-organized scientific production with general design principles), (b) further exploration of the consequences of our results regarding the future evolution of scientific cooperation around Atapuerca (i.e. its apparent lack of robustness), as well as the elaboration of *take-home messages* for stakeholders involved in similar processes, and finally (c) complementary studies incorporating either information already available but not used (e.g. nationality of scholars involved in Atapuerca's fieldwork), or new sort of data (e.g. project funding).

We expect these and other similar contributions to seed a growing literature addressing scientific collaboration processes in LSRIs in general, and emergent phenomena like Atapuerca in particular.

Acknowledgments The authors want to thank Javier Fernández López de Pablo and Ignasi Pastó for fruitful discussions, and the two anonymous referees for their valuable comments on the manuscript. This work has been partially supported by the Spanish DGICYT Grants CGL2012-38434-C03-03 and FIS2012-38266, the Generalitat de Catalunya Grants 2009-SGR-188 and 2009-SGR-838, as well as the FP7-FET Project PLEXMATH. Sergi Lozano acknowledges the Beatriu de Pinós post-doctoral research fellowship (Generalitat de Catalunya and COFUND Marie Curie Actions, EU-FP7), and Alex Arenas the financial support from ICREA Academia and James S. McDonnell Foundation. Fieldwork at Atapuerca is supported by Junta de Castilla y León and Fundación Atapuerca.

Appendix: TOP-20 articles of Atapuerca

As explained in the main text, the bibliometric dataset was based on the following list of 20 high impact publications reporting the most important empirical findings at the archaeological complex and introducing theories based on them.

- Carbonell, E.; Bermúdez de Castro, J.M.; Arsuaga, J.L.; Díez, J.C.; Rosas, A.; Cuenca-Bescós, G.; Sala, R.; Mosquera, M. & Rodríguez, X.P. (1995). Lower Pleistocene hominids and artefacts from Atapuerca-TD6 (Spain). Science 269 (5225), 826–830.
- Arsuaga, J. L., I. Martinez, A. Gracia, and C. Lorenzo. 1997. The Sima de los Huesos crania (Sierra de Atapuerca, Spain). A comparative study. Journal of Human Evolution 33:219–281.
- Falguères, C., J.-J. Bahain, Y. Yokoyama, J. L. Arsuaga, J. M. Bermúdez de Castro, E. Carbonell, J. L. Bischoff, and J.-M. Dolo. 1999. Earliest humans in Europe: the age of TD6 Gran Dolina, Atapuerca, Spain. Journal of Human Evolution 37:343–352.
- Arsuaga, J.L.; Martínez, I.; Gracia, A.; Carretero, J.M. & Carbonell, E. (1993): Three new human skulls from the Sima de los Huesos in the Sierra de Atapuerca (Burgos, Spain). Nature 362(6420):534–537.
- Parés, J. M., and A. Pérez-González. 1995. Paleomagnetic age for hominid fossils at Atapuerca archaeological site, Spain. Science 269:830–832.
- Carbonell, E.; J. M. Bermúdez de Castro, J. M. Parés, A. Pérez-González, G. Cuenca-Bescós, A. Ollé, M. Mosquera, R. Huguet, J. van der Made, A. Rosas, R. Sala, J. Vallverdú, N. García, D. E. Granger, M. Martinón-Torres, X. P. Rodríguez, G. M. Stock, J. M. Verges, E. Allué, F. Burjachs, I. Cáceres, A. Canals, A. Benito, C. Díez, M. Lozano, A. Mateos, M. Navazo, J. Rodríguez, J. Rosell, J. L. Arsuaga. 2008. The first Hominin of Europe. Nature. 452, 465–469.
- Fernández-Jalvo, Y., and P. Andrews. 1992. Small mammal taphonomy of Gran Dolina, Atapuerca (Burgos), Spain. Journal of Archaeological Science 19:407–428.
- Bermúdez de Castro, J. M., J.L. Arsuaga; Carbonell, E; A. Rosas, I. Martínez & M. Mosquera (1997). A Hominid from the Lower Pleistocene of Atapuerca, Spain: Possible Ancestor to Neandertals and Modern Humans. Science 276:1392–1395.

- Arsuaga, J. L., I. Martinez, A. Gracia, J. M. Carretero, C. Lorenzo, N. Garcia, and A. I. Ortega. 1997. Sima de los Huesos (Sierra de Atapuerca, Spain). The site. Journal of Human Evolution 33:109–127.
- Parés, J. M., and A. Pérez-González. 1999. Magnetochronology and stratigraphy at Gran Dolina section, Atapuerca (Burgos, Spain). Journal of Human Evolution 37:325–342.
- Bischoff, J. L., R. W. Williams, R. J. Rosenbauer, A. Aramburu, J. L. Arsuaga, N. García, and G. Cuenca-Bescós. 2007. High-resolution U-series dates from the Sima de los Huesos hominids yields: implications for the evolution of the early Neanderthal lineage. Journal of Archaeological Science 34:763–770.
- Bischoff, J. L., J. A. Fitzpatrick, L. León, J. L. Arsuaga, C. Falgueres, J. J. Bahain, and T. Bullen. 1997. Geology and preliminary dating of the hominid-bearing sedimentary fill of the Sima de los Huesos chamber, Cueva Mayor of the Sierra de Atapuerca, Burgos, Spain. Journal of Human Evolution 33:129–154.
- Arsuaga JL, Lorenzo C, Carretero JM, Gracia A, Martinez I, Garcia N, Bermúdez de Castro JM, Carbonell E (1999) A complete human pelvis from the Middle Pleistocene of Spain. Nature 399(6733):255–258.
- Cuenca-Bescós G, Laplana C, Canudo J.L. (1999). Biochronological implications of the Arvicolidae (Rodentia, Mammalia) from the Lower Pleistocene hominid-bearing level of Trinchera Dolina 6 (TD6, Atapuerca, Spain). Journal of Human Evolution 37(3–4):353–373.
- Bischoff JL, Shamp DD, Aramburu A, Arsuaga JL, Carbonell E, de Castro JMB (2003). The Sima de los Huesos hominids date to beyond U/Th equilibrium (>350 kyr) and perhaps to 400-500 kyr: New radiometric dates. Journal of Archaeological Science 30(3):275–280.
- Arsuaga, J. L., I. Martinez, C. Lorenzo, A. Gracia, A. Muñoz, O. Alonso, and J. Gallego. 1999. The human cranial remains from Gran Dolina Lower Pleistocene site (Sierra de Atapuerca, Spain). Journal of Human Evolution 37:431–457.
- Arsuaga, J. L., Carretero, J.M., Lorenzo, C., Gracia, A., Martínez, I., Bermúdez de Castro, J.M. & Carbonell, E. (1997): Size variation in Middle Pleistocene humans. Science 277:1086–1088.
- Fernández-Jalvo, Y., Díez, C., Cáceres, I. & Rosell J. 1999. Human cannibalism in the Early Pleistocene of Europe (Gran Dolina, Sierra de Atapuerca, Burgos, Spain). Journal of Human Evolution 37(3–4):591–622.
- Martínez I, Arsuaga JL (1997) The temporal bones from Sima de los Huesos Middle Pleistocene site (Sierra de Atapuerca, Spain). A phylogenetic approach. Journal of Human Evolution 33(2–3):283–318.
- Martinón-Torres, M., M. Bastir, et al. (2006). Hominin lower second premolar morphology: evolutionary inferences through geometric morphometric analysis. Journal of Human Evolution 50(5):523–533.

References

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics, 6*(3), 403–412.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics, 74*, 47–97.

Arsuaga, J. L., Lorenzo, C., Carretero, J. M., Gracia, A., Martínez, I., García, N., de Castro, J. M. B., et al. (1999). A complete human pelvis from the Middle Pleistocene of Spain. *Nature, 399*(6733), 255–258.

Arsuaga, J. L., Martinez, I., Gracia, A., Carretero, J. M., & Carbonell, E. (1993). Three new human skulls from the sima de los huesos middle pleistocene site in sierra de atapuerca, Spain. *Nature, 362*(6420), 534–537.

Barabási, A., Jeong, H., Nda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(34), 590–614.

de Beaver, D., & Rosen, R. (1978). Studies in scientific collaboration. *Scientometrics, 1*(1), 65–84.

de Beaver, D., & Rosen, R. (1979). Studies in scientific collaboration—Part ii. Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799–1830. *Scientometrics, 1*(2), 133–149.

Börner, K., Dall'Asta, L., Ke, W., & Vespignani, A. (2005). Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. *Complexity, 10*(4), 57–67.

Burt, R. S. (2004). Structural holes and good ideas1. *American Journal of Sociology, 110*(2), 349–399.

Carbonell, E., Cáceres, I., Lozano, M., Saladié, P., Rosell, J., Lorenzo, C., Vallverdú, J., et al. (2010). Cultural cannibalism as a paleoeconomic system in the European Lower Pleistocene. *Current Anthropology, 51*(4), 539–549.

Carbonell, E., de Castro, J. B., Arsuaga, J., Diez, J., Rosas, A., Cuenca-Bescos, G., et al. (1995). Lower Pleistocene hominids and artifacts from Atapuerca-TD6 (Spain). *Science, 269*(5225), 826–830.

Carbonell, E., de Castro, J. M. B., Parés, J. M., Pérez-González, A., Cuenca-Bescós, G., Ollé, A., et al. (2008). The first hominin of Europe. *Nature, 452*(7186), 465–469.

Chin, G. Jr., Myers, J., & Hoyt, D. (2002). Social networks in the virtual science laboratory. *Communications of the ACM, 45*(8), 87–92.

Coleman, J. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.

de Castro, J. M. B., Arsuaga, J. L., Carbonell, E., Rosas, A., Martínez, I., & Mosquera, M. (1997). A hominid from the Lower Pleistocene of Atapuerca, Spain: Possible ancestor to neandertals and modern humans. *Science, 276*(5317), 1392–1395.

de Price, D. J. S. (1986). *Little science, big science... and beyond*. New York: Columbia University Press.

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics, 5*(1), 187–203.

European Commission, et al. (2010). *A vision for strengthening world-class research infrastructures in the ERA. Report of the expert group on research infrastructures*. Brussels: European Commission.

Fernández-Jalvo, Y., Díez, J. C., de Castro, J. M. B., Carbonell, E., & Arsuaga, J. L. (1996). Evidence of early cannibalism. *Science, 271*, 277–278.

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics, 4*(4), 540–553.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics, 3*(3), 222–232.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Science of the USA, 99*(12), 7821–7826.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics, 51*(1), 69–115.

Glänzel, W., & Schubert, A. (2005a). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research* (pp. 257–276). Dordrecht (NL): Kluwer Academic Publishers.

Glänzel, W., & Schubert, A. (2005b). Domesticity and internationality in co-authorship, references and citations. *Scientometrics, 65*(3), 323–342.

Hagstrom, W. O. (1975). *The scientific community* (Vol. 130). New York: Basic Books.

Hara, N., Solomon, P., Kim, S. L., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology, 54*(10), 952–965.

Havemann, F., Wagner-Döbler, R., & Kretschmer, H. (2001). In *Proceedings of the Second Berlin Workshop on Scientometrics and Informetrics*, *Collaboration in Science and in Technology*.

He, B., Ding, Y., Tang, J., Reguramalingam, V., & Bollen, J. (2013). Mining diversity subgraph in multidisciplinary scientific collaboration networks: A meso perspective. *Journal of Informetrics, 7*(1), 117–128.

Hennemann, S., Rybski, D., & Liefner, I. (2012). The myth of global science collaboration—Collaboration patterns in epistemic communities. *Journal of Informetrics, 6*(2), 217–225.

Hochadel, O. (2013a). A boom of bones and books: The popularization industry of Atapuerca and human-origins research in contemporary Spain. *Public Understanding of Science, 22*(5), 530–537.
Hochadel, O. (2013b). *El Mito de Atapuerca: Orígenes, Ciencia, Divulgación*. Barcelona: Edicions UAB.
Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in scientometrics. *Scientometrics, 75*(2), 189–202.
Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science, 322*(5905), 1259–1262.
Katz, J. (1994). Geographical proximity and scientific collaboration. *Scientometrics, 31*(1), 31–43.
Katz, J., & Martin, B. R. (1997). What is research collaboration?. *Research Policy, 26*(1), 1–18.
Kraut, R., Egido, C., & Galegher, J. (1988). Patterns of contact and communication in scientific research collaboration. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work, CSCW '88*, *CSCW '88* (pp. 1–12).
Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the web. *Scientometrics, 60*(3), 409–420.
Liu, X., Bollen, J., Nelson, M. L., & de Sompel, H. V. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management, 41*(6), 1462–1480.
Mali, F., Kronegger, L., Doreian, P., & Ferligoj, A. (2012). Dynamic scientific co-authorship networks. In A. Scharnhorst, K. Brner, & P. Besselaar (Eds.), *Models of science dynamics, understanding complex systems* (pp. 195–232). Berlin: Springer.
Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics, 36*(3), 363–377.
Newman, M. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.
Newman, M. (2010). *Networks: An introduction*. Oxford: Oxford University Press.
OECD. (2010). *Establishing large international research infrastructures*: *Issues and options*. Report of the Global Science Forum.
Perc, M. (2010). Growth and structure of Slovenias scientific collaboration network. *Journal of Informetrics, 4*(4), 475–482.
Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science, 86*(3), 423–443.
Viana, M. P., Amancio, D. R., & da F Costa, L. (2013). On time-varying collaboration networks. *Journal of Informetrics, 7*(2), 371–378.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications (structural analysis in the social sciences) Author: Stanley Wasserman*. Cambridge, MA: Cambridge University Press.
Wray, K. B. (2002). The epistemic significance of collaborative research. *Philosophy of Science, 69*(1), 150–168.
Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039.
Zuccala, A. (2006). Modeling the invisible college. *Journal of the American Society for Information Science and Technology, 57*(2), 152–168.