Italian sociologists: a community of disconnected groups

Akbaritabar, A.; Traag, V.A.; Caimo, A.; Squazzoni, F.

Scientometrics (2020) 124:2361–2382
https://doi.org/10.1007/s11192-020-03555-w

Italian sociologists: a community of disconnected groups

Aliakbar Akbaritabar[1] · Vincent Antonio Traag[2] · Alberto Caimo[3] · Flaminio Squazzoni[4]

Received: 14 February 2020 / Published online: 13 June 2020

Abstract

Examining coauthorship networks is key to study scientific collaboration patterns and structural characteristics of scientific communities. Here, we studied coauthorship networks of sociologists in Italy, using temporal and multi-level quantitative analysis. By looking at publications indexed in Scopus, we detected research communities among Italian sociologists. We found that Italian sociologists are fractured in many disconnected groups. The giant connected component could be split into five main groups with a mix of three main disciplinary topics: sociology of culture and communication (present in two groups), economic sociology (present in three groups) and general sociology (present in three groups). By applying an exponential random graph model, we found that collaboration ties are mainly driven by the *research interests* of these groups. Other factors, such as *preferential attachment*, *gender* and *affiliation homophily* are also important, but the effect of gender fades away once other factors are controlled for. Our research shows the advantages of multi-level and temporal network analysis in revealing the complexity of scientific collaboration patterns.

Keywords Italian sociology · Coauthorship networks · Preferential attachment · Community detection · Exponential random graph model (ERGM)

✉ Aliakbar Akbaritabar
akbaritabar@dzhw.eu

Vincent Antonio Traag
v.a.traag@cwts.leidenuniv.nl

Alberto Caimo
alberto.caimo@tudublin.ie

Flaminio Squazzoni
flaminio.squazzoni@unimi.it

[1] German Centre for Higher Education Research and Science Studies (DZHW), Schützenstr. 6a, Berlin 10117, Germany

[2] Center for Science and Technology Studies (CWTS), Leiden University, Leiden, The Netherlands

[3] School of Mathematical Sciences, Technological University Dublin, Dublin, Ireland

[4] Department of Social and Political Sciences, University of Milan, Milan, Italy

Introduction

Connections between scientists are important to scientific progress (Garvey 1979; Zhang et al. 2018). Teamwork is paramount in science today more than ever and this is true in both hard and social sciences (Wuchty et al. 2007). Having larger collaboration networks seems to increase the number of publications and citations of scientists, with important implications on tenure and promotion (Grant and Ward 1991; Long 1992; Leahey et al. 2010), as well as on funding (Nederhof 2006; Edwards and Roy 2017). In addition, collaboration also plays a role in recognition and academic reputation (Merton 1968).

Quantitative analysis of coauthorship of scientific publications has been one of the most frequent means of studying scientific collaboration (Katz and Martin 1997; Batagelj et al. 2017). This type of study reveals conditions and effects of collaboration across a wide spectrum of scientific activities, including grant proposal and funding (Bellotti et al. 2016; Sciabolazza et al. 2017). Coauthorship networks can also reveal the structure of the scientific community, the evolution of its epistemic field, the degree of cohesiveness or fragmentation and the co-existence of scholarly communities.

In an influential article, Moody (2004) questioned whether sociology became a discipline more socially integrated in the last decades. He examined coauthorship networks of sociologists using all sociological abstracts in English language from 1963 to 1999, amounting to 197,976 abstracts. He explored three competing hypotheses on the nature of scientific collaborations in sociology: (1) Collaboration represents a small world of distant communities of sociologists focusing on their substantive research areas, while scholars are connected through short paths; (2) A large periphery of scholars gathered around a core of a few star scientists; or (3) A structurally cohesive network, as suggested by Abbott (2001), with wide-ranging collaboration between different specialists. Abbott argued that the peculiar position of sociology, which has always been surrounded by adjacent disciplines and idea spaces, could make it permeable to external theories, methods and concepts, making *wide-ranging* collaboration especially around quantitative research more likely to happen.

Moody (2004) found that sociology was characterized by a structurally cohesive core, which has grown steadily over the time. Collaboration depended on research specialization with quantitative researchers more engaged in collaboration. Finally, he suggested that a scientist's probability of being embedded in the core network depended more on collaboration trajectories than on his or her research specialty.

Sciabolazza et al. (2017) used a modularity algorithm (Newman and Girvan 2004) to study coauthorship networks between scholars from the University of Florida in 2013–2015. They used an exponential random graph model (ERGM) and found that similar institutional affiliation, spatial proximity, transitivity effects, and use of similar research services provided by the university predicted higher rates of collaboration.

Zhang et al. (2018) recently proposed an interesting ERGM specification that is relevant for studying coauthorship networks. By looking at 633 prolific authors in computer science, they analyzed different factors influencing coauthorship tie formation, including homophily, transitivity and preferential attachment. Tie formation was found to be a complex process, often dominated by transitivity (i.e., the tendency of authors to collaborate with their coauthors' collaborators is strong) and preferential attachment (i.e., the more coauthors one has, the more new collaborators (s)he will attract). These factors may contribute to the so-called "*Matthew effect*", leading to cumulative processes of academic recognition and prestige (Merton 1968).

In regards to collaboration patterns, the case of sociology is of particular interest. While sociologists collaborate more than humanities scholars and less than physicists (Babchuk et al. 1999), they are fragmented in small groups with weak epistemic and methodological coherence and a contested subject (Abbott 2000, 2001; Turner 2006). The lack of prominent and universally accepted paradigms in the field (Wallerstein 2000; Hargens 2004) and the competition with other specialists, such as economists and political scientists (Wallerstein 2000), could lead to complex collaboration patterns. Furthermore, the case of Italian sociologists is of special interest, considering the limited size of the community and its fracture between more internationalized and more local scholars (Akbaritabar et al. 2018).

To examine these patterns, we constructed the coauthorship network based on publications indexed in Scopus with a "*complete*" or "*sociocentric*" (Marsden 2002) network approach. Uncovering patterns in these coauthorship networks requires advanced quantitative analysis. Here, we followed Sciabolazza et al. (2017) to detect research communities and Zhang et al. (2018) in using an ERGM (Lusher et al. 2013) to analyse the networks. We followed previous research on Italian sociologists by Akbaritabar et al. (2018) and Bellotti et al. (2016), and used a multi-level approach (Lazega et al. 2008). We used a model to check interaction between different factors while controlling for individual scientist attributes, along with covariate attributes, communities and network level characteristics.

With this multi-level design, we aimed to understand whether Italian sociologists tended to collaborate preferably with well-known and more prolific colleagues (preferential attachment), with colleagues of their same gender (gender homophily), and with their same affiliation country (affiliation homophily). Furthermore, we wanted to understand if they were inclined to collaborate more with colleagues with the same research productivity level and/or who shared the same substantive focus of research.

The structure of the paper is as follows: In "Data and method" section, we presented our data and methods. In "Results" section we presented our findings, while we discussed our main results in "Conclusions and discussion" section.

Data and method

We gathered data from the website of the Italian Ministry of Education, Universities and Research (MIUR) for all currently hired sociologists in Italian universities and research centers. This included information about the subject's current academic position (i.e., assistant, associate or full professor), the "scientific disciplinary sector"[1] in which (s)he has been formally hired, gender, affiliation, department, and last and first name (Akbaritabar et al. 2018).

We then extracted all publications by Italian sociologists (3168 including *Article* 1912, *Article in Press* 54, *Book* 85, *Book Chapter* 477, *Conference Paper* 113, *Editorial* 116, *Erratum* 4, *Letter* 3, *Note* 35, *Review* 355, *Short Survey* 6 and 8 without a document type) from Scopus in September 2016.[2] We did not apply time or document type limitation to

[1] Sectors established by MIUR are as follows: General sociology (SPS/07), Sociology of culture and communication (SPS/08), Economic sociology (SPS/09), Environmental sociology (SPS/10), Political sociology (SPS/11) and Sociology of law and social change (SPS/12).

[2] We wrote R (2016) scripts to interact with the Scopus API. It searched each author's last and first name in Scopus and extracted all publications records. Data gathering started by sending search queries to Scopus API on July 27th 2016, while from September 8th 2016 we gathered Scopus CSV exports of all available information on publications through Scopus web interface to cover shortages with data from API. To

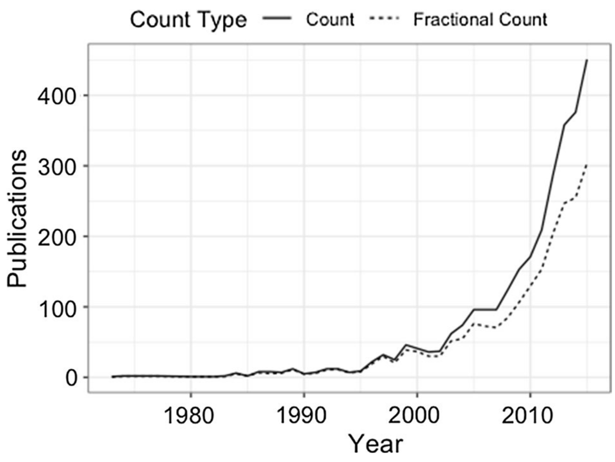

Fig. 1 Total number of publications (count and fractional count based on number of authors) 1973–2015 (Scopus data)

include all possible scientific output. Data included articles' title, keywords, abstract, publication year, authors' names and affiliations and number of citations received. Figure 1 shows the growth in number of publications in all journals over time with a similar trend in both *counts* and *fractional counts* based on the number of authors. The evolution of fractional counts shows that Italian sociology has moved towards higher number of coauthorships from 2003 onward (note the gap between solid and dashed lines) in line with literature (e.g., Wuchty et al. (2007)).

For authors whose gender was missing from the MIUR website, we searched for an online profile and photo. After careful checking, only 15 cases with missing gender remained. We also assigned each author's continent based on country of affiliation.

We constructed a coauthorship network from articles as undirected ties. Ties were weighted for repeated coauthorships (Newman 2001a, b) using full counting (Perianes-Rodriguez et al. 2016). This allowed for the projection of the bipartite network of the ties between authors and papers. We used the author identification number provided by Scopus[3] to treat name disambiguation (De Stefano et al. 2013). We looked at all publications of sociologists in the list extracted from the MIUR website. We collected their collaborations with scientists outside Italy or in other fields of science. However, we did not collect the full publication list for each collaborator. This implies that any collaborator existed in the coauthorship network only because he/she coauthored an article (or more) with an Italian sociologist.

Footnote 2 (continued)

process and analyse the data, we used base (2016), dplyr (2016), igraph (2006), ERGM (in Statnet) (2008, 2016), stargazer (2015), ggplot2 (2009), tidyverse (2017), jsonlite (2014) and stringdist (2014) packages in R (2016).

[3] We controlled for IDs duplication, homonyms and multiplicity by cross-checking Scopus web interface data with information extracted from the API using the R scripts described before and manually with help of research assistants.

Community detection

In order to detect coauthor communities, we used the Leiden algorithm for community detection (Traag et al. 2019) as implemented in the leidenalg library in Python (see here[4] for *how-to-use* and technical descriptions). This library allows to apply different community detection methods on network graphs as elaborated in Traag (2014). We specifically used the *Constant Potts model* (CPM) (Traag et al. 2011), which is a specific version of the more general Potts model suggested by Reichardt and Bornholdt (2004). CPM was proposed by Traag et al. (2011) as a *resolution-limit-free* method to overcome the resolution limit in modularity (Newman and Girvan 2004) and other methods for community detection. This limit impedes the detection of small communities in large networks and affects the efficiency of the community detection.

The idea of community detection principally emphasizes the importance of links *within* communities rather than those *between* them. CPM uses a resolution parameter γ, i.e., the "*constant*" in the name, leading to communities such that the link density between the communities (external density) is lower than γ and the link density within communities (internal density) is higher than γ (Traag et al, 2011). Note that γ is the *resolution parameter* helping CPM to be a *resolution-limit-free* method. This allowed us to detect communities with a particular density and size. After checking, we set the resolution (i.e., γ, the density of communities in CPM) to 2×10^{-4}. This particular configuration gave us the five *largest* communities in the giant component with density equal to chosen γ. As discussed in Traag et al. (2019), compared to other community detection algorithms, e.g., Louvain, the Leiden algorithm is more robust to detect communities while ensuring high internal connectivity in those communities. It works efficiently in detecting communities of small sizes. Note that we used the earlier specified edge weights in the community detection.

Visualization techniques: substantive focus evaluation with VOS term maps

In order to study the substantive focus of publications, we used VOSviewer, a software tool developed by Van Eck and Waltman (2010). This allowed us to parse corpora of text, detect terms (i.e., *noun-phrases*) using natural language processing and obtain a term map visualization based on the VOS layout algorithm (van Eck et al. 2010). The distance between terms in this map reflects co-occurrence of such terms in documents: more frequently co-occurring terms tend to appear closer to each other. Additionally, this tool clusters terms in groups. We projected author level characteristics in the term map (e.g., coauthorship communities' membership, country of affiliation, first and last publication dates as proxy of academic career trajectories) to understand the substantive focus of research. For example, we overlaid the coauthorship communities found by the community detection method on top of the substantive term maps to see if communities corresponded to research specialization.

Exponential random graph model

We used Exponential Random Graph Models (ERGMs) to simulate networks based on the giant component of Italian sociologists and their coauthors network, i.e., our

[4] https://leidenalg.readthedocs.io/en/latest/intro.html.

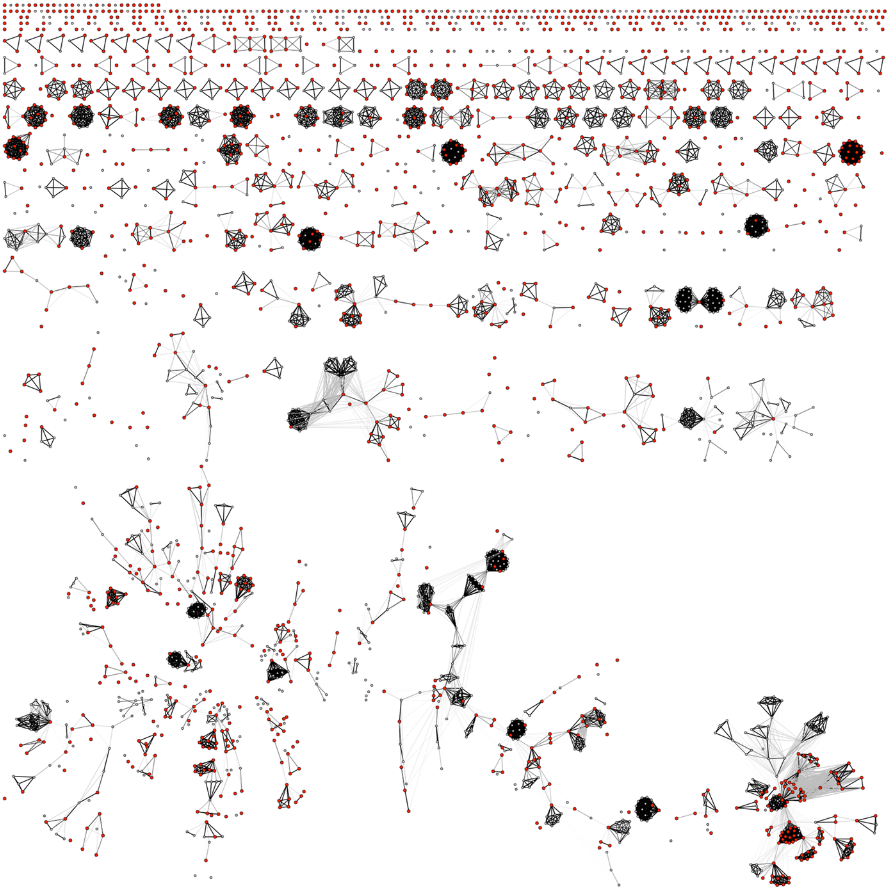

Fig. 2 The coauthorship network of all Italian sociologists and their coauthors (Colors: Affiliated to Italy = Red, Affiliated elsewhere = Gray, Ties width show the backbone of the network) (Scopus data). (Color figure online)

observed network (Lusher et al. 2013). This provided a baseline to estimate if our coauthorship network reflected unique characteristics compared to what we would expect from a distribution of random networks generated with the same size and density as the observed network.

Furthermore, ERGMs allowed us to consider different types of attributes while modelling the probability of tie existence in the network. For instance, we considered *node* attributes, including author's academic seniority, gender and continental region of affiliation. *Covariate* attributes allowed us to control and compare two connected nodes for similarity or differences of node attributes. This allowed us to control for differential homophily effects in our network (e.g., Morris et al. 2008, Bianchi et al. 2018). Furthermore, ERGMs enabled us to check *structural effects* such as preferential attachment (by degree distribution), thereby considering possible Matthew effects (Merton 1968), i.e., cumulative advantage from collaborations. This mix of nodal and structural attributes in one integrated model is important to understand tie existence more effectively.

Table 1 The main characteristics of the coauthorship network of Italian sociologists and their collaborators and its giant component (rows with “G-comp” in title)

Metric	Value
Number of nodes	2747
Number of ties	7618
Mean degree	5.55
Number of communities	512
Community size (mean)	5.37
Community size (SD)	32.72
Number of nodes (G-comp)	712
Number of ties (G-comp)	2221
% nodes in (G-comp)	25.92%
% ties in (G-comp)	29.15%
Number of female authors (G-comp)	314
Number of male authors (G-comp)	383
Density (G-comp)	0.0088
Diameter (G-comp)	32

Results

Figure 2 shows the coauthorship network of all Italian sociologists and their coauthors with affiliation of authors indicated by node colors. Ties are colored based on being influential in building the backbone[5] of the graph (Nick et al. 2013). We found that the majority of Italian sociologists (1641 out of 2747 total) have coauthors affiliated elsewhere (see gray nodes in Fig. 2), while the homophily of connections *within* Italians seems to be high (see the inter-connectivity of red nodes in most parts of the graph). Table 1 shows the main features of this network. The sparse coauthorship relations (Average degree = 5.5) with relatively high number of connected components (512) indicate the level of disconnectedness of the network. The connected components greatly vary in size: the largest connected component had 712 members, while the second largest component had 184 members (Mean = 5.37, SD = 32.72). The large number of small connected components indicate there are authors who published either alone, with few coauthors, or coauthoring only with authors not included in these networks.

As indicated in Table 1 (Rows indicated by *(G-comp)*), the giant component only contains 26% of the nodes in the full network, with 29% of the ties. This is a relatively low percentage compared to random networks simulated with similar degree distribution as the observed one. Note that we used scale-free, preferential attachment and Erdos–Renyi random networks to have a baseline of comparison. Our observed network showed lower density and percentage of nodes in giant component. It is worth noting that this percentage is lower than the observed rate in the sociological community in Slovenia (90.8% of nodes in giant component) discussed in Kronegger et al. (2011), as well as in the general case of Slovenian scientists (88% of nodes in giant component) discussed in Kastrin et al. (2017). The percentage of nodes is also smaller than the size of giant component of international

[5] This is a visualization technique described in detail by Nick et al. (2013) to emphasize the influential ties by colors. It positions influential ties in front, while less influential ties are colored with less dark colors and positioned behind to highlight the backbone.

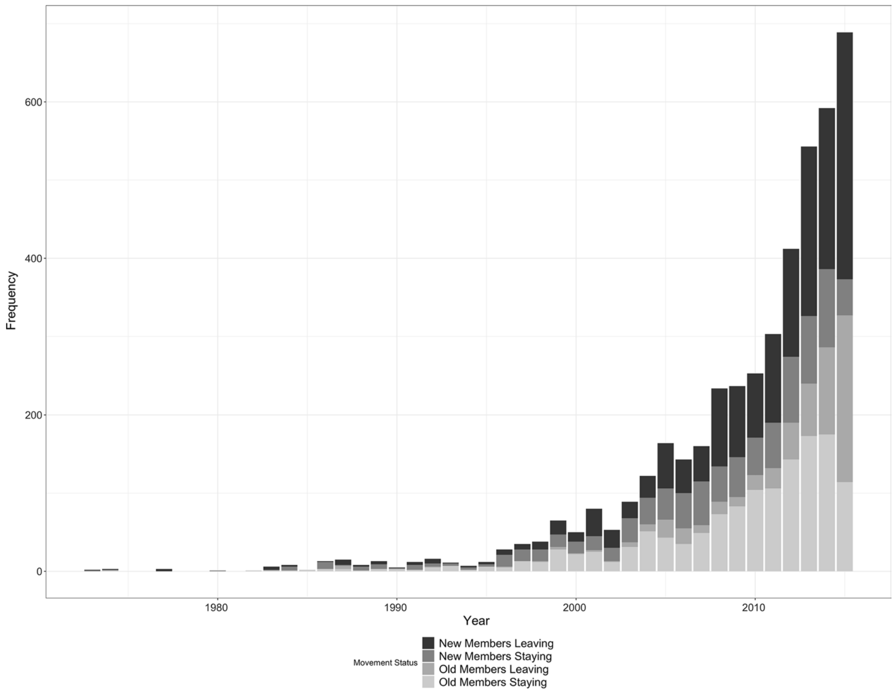

Fig. 3 The temporal evolution of all authors in sample the x-axis denotes the years, 1973–2015, the y-axis denotes the frequency of individual authors (Scopus data)

sociologists (40% of nodes in giant component) discussed in Moody (2004). The average degree of the giant component was relatively low (6.24, SD = 6.61) and only slightly higher than the average degree of the full network (5.55, SD = 6.74).

Figure 3 provides a different visualization of the coauthorship network and shows the temporal evolution of the arrival or leave of groups of authors in the network. Following Palla et al. (2007), who used this type of visualization to examine the movement and turnover of individuals in network, we distinguished four groups of authors: (1) Those who published at least 2 years before a given year and continued to publish for at least 2 years later (*old members/staying*, see lightest color, bottom stack of bars in plot), (2) Those who published at least 2 years before a given year with the last publication in the given year (*old members/leaving*, see darker color than first group, second stack of bars from bottom in plot), (3) Those who first published in the given year, and published for at least 2 more years (*new members/staying*, darker than the two first groups, third stack of bars from bottom in plot), and (4) Those who first published in the given year and did not publish anything later (*new members/leaving*, darkest colors, fourth stack of bars on plot). Note that most authors in the sample were *newcomers* who immediately left and disappeared from Scopus the following year (darkest stack of bars on Fig. 3). However, some newcomers joined the core of more senior authors of the sample (second stack of bars from the top). The systematic turnover of newcomers in each of the detected five communities of the giant component follows the general trend observed in the whole network. Note also that we removed the two last years in the sample from Fig. 3 to prevent a distorted decreasing picture.

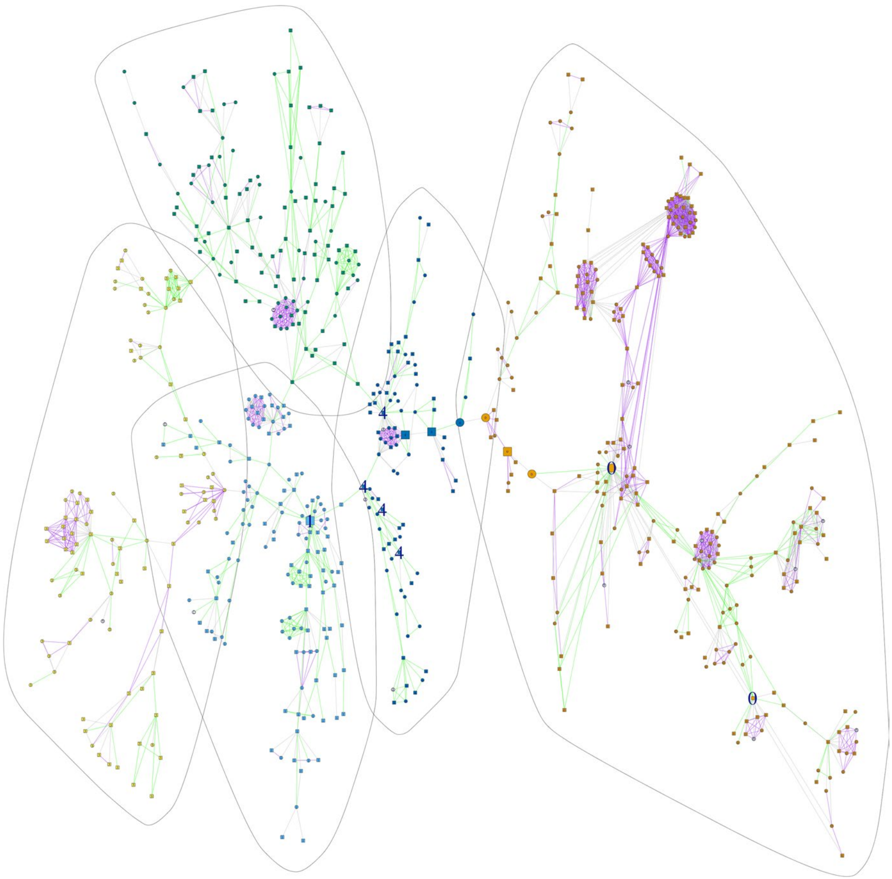

Fig. 4 The giant component of Italian sociologists and their coauthors network with five communities (Node colors and labels: community membership, Tie colors: Within Italians = green, Within non-Italians = purple, Between Italians and non-Italians = gray, Node shapes: Square = Male, Circle = Female, Pie = Missing gender, Node size = Betweenness centrality, Node label size is enlarged in case of 7 authors with higher than 25 publications, Ties width are kept equal throughout the visualization, borders show approximate location of communities) (Scopus data). (Color figure online)

Figure 4 shows the *five* communities detected from the giant component (designated by colors and identification numbers inside each node, while borders indicate approximately where each community is located), with a total of 712 authors. We will refer to these communities as 0, 1, 2, 3, 4 which have 254, 142, 122, 103, 91 members, respectively.

We found many isolated authors (244 in the whole network, see Fig. 2 for a visualization of the full graph) or connected components (total of 512) formed between Italian sociologists working with their own group of contacts. Figure 4 shows the prevalence of homophilous ties *within* Italians (green ties) and non-Italians (purple ties) and the infrequency of ties *between* Italians and non-Italians (gray ties). Note that four out of seven most prolific authors with more than 25 publications (designated with enlarged node labels) are members of *community 4* (dark blue nodes in center of graph) while only two of them have relatively high *betweenness centrality* in the coauthorship ties (designated with enlarged node sizes).

Table 2 Gender and sectors composition and internationality of members of the communities detected from the giant component (Percentages are calculated by rows for each community separately for gender, country and sectors)

		Gender			Country				Scientific disciplinary sectors (SPS)						
ID	# member	F (%)	M (%)	Missing G (%)	EU (%)	IT (%)	Other (%)	Missing C (%)	07 (%)	08 (%)	09 (%)	10 (%)	11 (%)	Postdoc (%)	Missing S (%)
0	254	43	54	3	54	29	11	5	1	5	0	0	0	2	91
1	142	50	49	1	36	55	6	3	6	3	8	1	1	2	78
2	122	38	61	1	37	56	3	4	10	1	7	0	1	5	76
3	103	45	54	1	41	44	5	11	4	2	12	1	0	2	80
4	91	47	49	3	32	57	9	2	7	7	0	1	2	1	82

Tables 2 shows the share of these five communities considering author attributes (i.e., gender, country of affiliation and scientific disciplinary sectors). It is worth noting that most of these communities (with community 1 being an exception) have a majority of *male* members and the gender difference of 2%, 9% and 11% in case of communities 4, 3 and 0 reaches its highest in case of community 2 (23%), which also has the highest share of *postdocs*. Regarding the country of affiliation, we found that communities 1, 2, 3 and 4 were mainly composed of sociologists working in *Italy*, whereas community 0 had higher than half of its members from *international* authors, either from *Europe* or *other* countries (65%). The share of *scientific disciplinary sectors* reveals certain interesting trends. Community 0 is composed mainly of *Sociology of culture and communication (SPS/08)*, community 1 of *Economic sociology (SPS/09)* and *General sociology (SPS/07)*, community 2 of *General sociology (SPS/07)* and *Economic sociology (SPS/09)*, community 3 of *Economic sociology (SPS/09)* and community 4 of *General sociology (SPS/07)* and *Sociology of culture and communication (SPS/08)*. It is worth noting that we could *only* assign scholars to *sectors* for giant component members who had a sector assigned to them in the MIUR list (120 members, 17% of total) while this does not cover all scientists affiliated to Italian universities and institutions (317 scientists) because we used the MIUR list of sociologists. It was impossible to generalize this assignment to all members of each community, including international collaborators. Despite this limitation in sectors coverage, the composition of members in each community is highly reflected in the substantive focus of research that characterized the members of each community (see detail in the Substantive focus of research Section).

The highest percentage of ties between authors of the giant component (43%) were *cross-gender collaborations*, while 34% of all ties formed in the giant component were within *male authors* and only 19% were within *female authors*. In line with previous findings, *female-to-female* coauthorship ties were rare (e.g., Teele and Thelen (2017)). However, considering that the total number of females (44%) were lower than male authors (54%), these findings could simply reflect the lower number of potential female collaborators to choose among.

We then looked at the percentage of ties *within* and *between* two specific groups of authors: (1) Those who were currently hired Italian sociologists and (2) those who could be affiliated either in an Italian institution or abroad, either sociologist or not, either active or retired, but in any case, not included in the administrative list of MIUR. Note that the highest percentage of ties (46%) were *within* those not currently employed in an Italian institution. Only 25% of ties were *within* Italian sociologists and 28% were *between* Italians and non-Italians. Note that our dataset included all articles published by authors in the MIUR list, while in the case of their coauthors who were not present in the MIUR list, we did not collect all their articles and their presence in our sample is limited to articles coauthored with Italian sociologists included in the MIUR list.

Table 3 shows the most popular and prolific authors in the giant component and their gender, community membership and measured network characteristics. Confirming previous findings (Cole and Zuckerman 1984; Leahey 2006; Abramo et al. 2009), *male* authors dominate the picture, being the majority among prolific authors, those with highest betweenness, triangle counts and degree. There are some exceptions though, the most prolific author, author with the highest triangle count and highest degree were all *females* who were members of community 0. Community 0 presents a higher rate of internal connectivity. This community dominates the picture in *triangle* counts and degree. Three of the ten most prolific authors were members of this community. Community 4 which is the smallest in size (see Table 2) is highly represented in most of

Table 3 Comparing ranking of top 10 authors last name, gender and community membership in some of the main network characteristics (gender, community membership and network characteristic in parentheses, respectively)

<table>
<tr><th>Betweenness</th><th>Most prolific</th><th>Triangles</th><th>Degree</th></tr>
<tr><td>Bosi (Male | 4 | 0.473)</td><td>Fortunati (Female | 0 | 69)</td><td>D’Ambrosi (Female | 0 | 391)</td><td>Fortunati (Female | 0 | 53)</td></tr>
<tr><td>Mattoni (Female | 0 | 0.468)</td><td>Pavolini (Male | 1 | 35)</td><td>Splendore (Male | 0 | 360)</td><td>Neresini (Male | 0 | 39)</td></tr>
<tr><td>Della Porta (Female | 4 | 0.465)</td><td>Diani (Male | 4 | 35)</td><td>Harro-Loit (Female | 0 | 355)</td><td>D’Ambrosi (Female | 0 | 38)</td></tr>
<tr><td>Treré (Male | 0 | 0.371)</td><td>Boccagni (Male | 4 | 28)</td><td>Eberwein (Male | 0 | 332)</td><td>Splendore (Male | 0 | 37)</td></tr>
<tr><td>Farinosi (Female | 0 | 0.362)</td><td>Bucchi (Male | 0 | 28)</td><td>Groenhart (Male | 0 | 332)</td><td>Harro-Loit (Female | 0 | 35)</td></tr>
<tr><td>Pavolini (Male | 1 | 0.36)</td><td>Brighenti (Male | 4 | 27)</td><td>Porlezza (Male | 0 | 332)</td><td>Diani (Male | 4 | 30)</td></tr>
<tr><td>Giugni (Male | 4 | 0.358)</td><td>Ambrosini (Male | 4 | 26)</td><td>Fengler (Female | 0 | 325)</td><td>Eberwein (Male | 0 | 28)</td></tr>
<tr><td>Fortunati (Female | 0 | 0.342)</td><td>Ballarino (Male | 2 | 24)</td><td>Alsius (Male | 0 | 325)</td><td>Groenhart (Male | 0 | 28)</td></tr>
<tr><td>Boccagni (Male | 4 | 0.327)</td><td>Ruzza (Male | 1 | 23)</td><td>Baisnée (Male | 0 | 325)</td><td>Porlezza (Male | 0 | 28)</td></tr>
<tr><td>Pilati (Female | 4 | 0.299)</td><td>Mazzoleni (Male | 0 | 21)</td><td>Bichler (Male | 0 | 325)</td><td>Pavolini (Male | 1 | 26)</td></tr>
</table>

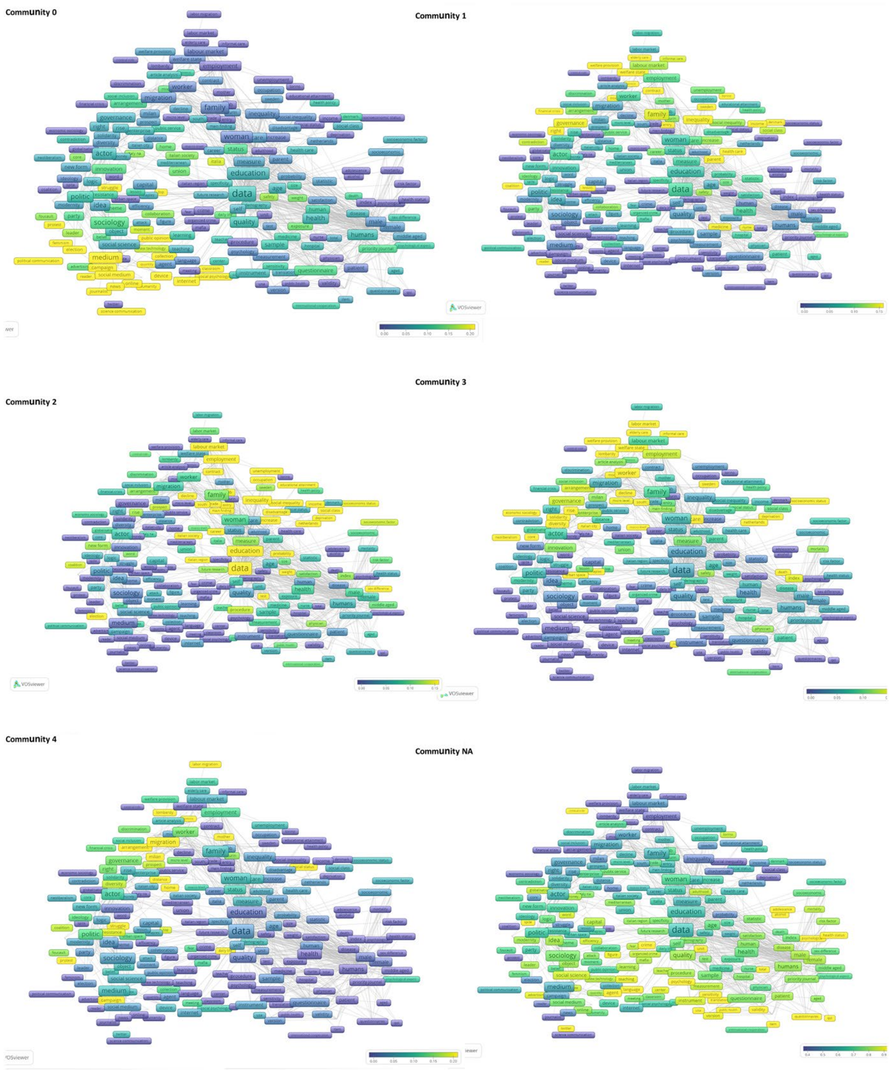

Fig. 5 The substantive focus of members of five communities of coauthorship network overlaid on terms extracted from all publications visualized with VOS viewer (Yellow parts on the plot show a higher substantive focus, that is, a higher frequency of those terms in publications of authors from communities. Community NA in bottom right are those not member of the giant component). (Color figure online)

these network characteristics, except *triangle count* which is completely dominated by community 0.

Substantive focus of research

In order to understand the dividing border between these communities of the giant component better, we considered the type of research performed by all sociologists included in our dataset. We developed term maps based on titles, keywords and abstracts of all papers included in the sample. By overlaying the community membership (i.e., communities 0, 1, 2, 3, 4 of the giant component) on top of the substantive term maps, we explored whether membership in communities was based on similarity in research focus between members.

As shown in Table 2, community 0 mainly consists of *Sociology of culture and communication (SPS/08)* with the highest share of members from Europe and other countries (65%), with only 29% of members coming from Italy. Scholars from community 0 are doing research on “medium”, “science communication”, “social medium”, “internet”, “political communication” and “public opinion” (see Fig. 5 top left). Community 1 has a slightly higher share of female sociologists and mainly consists of *Economic sociology (SPS/09)* and *General sociology (SPS/07)*. As shown in Fig. 5 top right, their research focuses on “family”, “inequality”, “welfare provision”, “elderly care”, “health care”, “parent”, “medicine”, “financial crisis” and “governance”. Community 2 has a majority of members from *General sociology (SPS/07)* and *Economic sociology (SPS/09)* from Italy (56%), while it has the highest gender difference between the members (61% vs. 38%) and the highest share of postdocs (5%). The research of these sociologists concentrated on “data”, “probability”, “weight”, “test”, “education”, “labour market”, “employment”, “unemployment”, to name a few (see Fig. 5 middle left). Community 3 has the highest share of *Economic sociology (SPS/09)* and 54% male members with a relatively high share of researchers from Italy (44%) and Europe and other countries (46%). These sociologists are doing research mainly on topics related to “governance”, “employment”, “worker”, “welfare state”, “neoliberalism” (see Fig. 5 middle right). The last community of the giant component is community 4 with a slightly higher share of male members (49% vs. 47%) and the highest share of Italian members among all communities (57%). It is mainly composed of *General sociology (SPS/07)* and *Sociology of culture and communication (SPS/08)* and it has 4 of the 7 most prolific authors. In these cases, research revolves around a niche set of subjects related to “migration”, “home”, “labor migration”, “daily life” and “social status” (see Fig. 5 bottom left). Figure 5 bottom right shows the substantive focus of authors who were not members of the giant component. While their focus is clearly different from communities 0, 1, 2, 3, 4, there is still an overlap between thematic areas.

In order to explore mechanisms that can account for these patterns, we built four ERGMs, including structural and individual factors. Table 4 shows results of the four ERGMs. Model 1 includes only structural effects, such as ties and preferential attachment. Results showed that there is a strong effect of *preferential attachment* in increasing probability of coauthorship ties existence. This indicates that authors who were already famous scholars with higher number of collaborations were also the ones with higher probability of forming coauthorship ties.

Note that the coauthorship network is a one-mode projection of the bipartite paper-author network. Hence, higher rates of cliquish structures can be simply due to articles with high number of authors. This can result in a high preferential attachment effect. To control for this, we included coauthorship edge *weights* in community detection although a better treatment would be to perform community detection on the bipartite network. However, unlike in the hard sciences, multiple coauthorship ties are rarer and relatively recent among sociologists, which is confirmed by the fact that 52% of articles in our sample were written by solo authors (see fractional count on Fig. 1).

We then added *differential homophily effects* and *absolute differences* based on *nominal/categorical* and *quantitative* author attributes (Morris et al. 2008, p. 6) to see if the preferential attachment effect could have been confounded by other factors. In Model 2, we included homophily effects based on author attributes: gender, continental region of

Table 4 ERGMs results explaining effect of author attributes and structural variables on coauthorship tie existence

	The giant component of Italian sociologists and their coauthors			
	ERGM models			
	(1)	(2)	(3)	(4)
Ties	−4.551*** (0.023)	−3.654*** (0.187)	−13.140*** (2.058)	−13.367*** (2.522)
Preferential attachment	15.224*** (4.697)			4.069*** (1.152)
Within females ties		−0.250 (0.169)	−0.042 (0.176)	−0.097 (0.259)
Within males ties		0.547*** (0.177)	0.357* (0.185)	0.385 (0.286)
Males main effect		−0.257 (0.164)	−0.086 (0.171)	−0.125 (0.267)
Within community 0			10.472*** (2.049)	10.771*** (2.487)
Within community 1			6.448*** (1.295)	6.513*** (1.470)
Within community 2			6.413*** (1.296)	6.064*** (1.567)
Within community 3			7.228*** (1.502)	7.484*** (1.747)
Within community 4			6.001*** (1.297)	5.986*** (1.430)
Community 1 main effect			2.034 (1.255)	2.153 (1.488)
Community 2 main effect			2.126* (1.254)	2.451 (1.582)
Community 3 main effect			1.809 (1.304)	1.837 (1.611)
Community 4 main effect			2.509** (1.252)	2.675* (1.407)
Within Europe		0.770*** (0.102)	0.840*** (0.105)	0.849*** (0.148)
Within Italy		1.054*** (0.125)	0.877*** (0.129)	0.872*** (0.178)
Within other countries		1.956*** (0.233)	1.842*** (0.241)	1.857*** (0.356)
Italy main effect		−0.456*** (0.099)	−0.194* (0.103)	−0.186 (0.138)
Other countries main effect		−0.438*** (0.102)	−0.465*** (0.105)	−0.482*** (0.151)
Difference in total pubs		0.060*** (0.002)	0.064*** (0.002)	0.065*** (0.004)
Difference in first pub		−0.090*** (0.006)	−0.095*** (0.007)	−0.096*** (0.009)
Difference in last pub		−0.372*** (0.014)	−0.367*** (0.014)	−0.378*** (0.020)
Akaike Inf. Crit.	25,238.680	22,710.210	16,466.530	16,344.340
Bayesian Inf. Crit.	25,270.010	22,835.510	16,685.800	16,584.500

$* = p < 0.1$; $** = p < 0.05$; $*** = p < 0.01$

affiliation, first and last publication dates of coauthors and the total number of publications assigned to each author as a node attribute. For the categorical variables of gender and continental region of affiliation, we considered the overall tie probability for each category (e.g., were males more likely to establish ties?), which we called the *main effect*, and the specific probability within each category (e.g., were males more likely to connect to other males?), which we called the *within* effect. For the continuous variables of first and last publication dates and the total number of publications, we considered the tie probability as a function of the absolute difference between two authors. For example, we observed a higher tie probability for authors that had large difference in number of publication, corresponding to a positive coefficient. In other words: a positive coefficient shows that authors with similar attributes are *less* likely to form ties, while a negative coefficient shows that authors with similar attributes are *more* likely to form ties.

In Model 3, we added the membership of the five detected communities as node attributes to see if, after controlling other attributes, coauthorship ties were more probable within same community members. Since our community detection configuration rewarded ties within a community, this could lead our ERGM to weight dis-proportionally the effect of community membership on tie existence and so adding endogeneity effects, whereas our intention was to have comparative models and see which homophily effect was more prevalent across different model specifications. We kept the main effects of Model 2 also in Model 3, while it had better BIC and AIC.

Results indicate that being a member of one of these communities (i.e., having similar substantive research focus) had the highest effect in increasing the probability of tie existence. The probability of tie existence was also higher between scholars in other countries (i.e., North America, Australia and New Zealand, South America and Asia), between scholars in Italy and between scholars in European countries, respectively. Having a similar date of the latest publication (e.g., publishing until recently) or earliest publication dates increased the probability of tie existence, while having a similar number of total publications decreased this probability. This suggests that ties are more likely between authors of the same seniority, and between more junior and more senior authors, most likely between PhDs and their supervisors. This means that highly prolific authors tend to collaborate with low prolific authors who are not necessarily younger because they have significant homophily in earliest and latest date of publications, i.e., academic age and seniority. This can be due to the fact that although the number of publications was increasing over time (see Fig. 1), the total number of publications for each author in the sample was highly skewed (Skewness = 5.93 which is in line with previous research Akbaritabar et al. (2018)) and the pool of potential collaborators for each author potentially included highly prolific and low prolific authors. Moreover, this can be the effect of the data gathering process. While we have looked at all articles throughout the scientific career of *Italian sociologists* and their collaborators, we could not include all articles of *these collaborators* in a one by one basis. This means that a person, either senior or junior, could appear in the sample only when s/he collaborated with an *Italian sociologist* in a Scopus tracked publication, so having a lower number of publications. While we did not find any trace of female homophily, we found some cues of male homophily.

Finally, in Model 4, we (re)included the preferential attachment effect. We found that including author attributes and community detection substantially decreased the structural effect of preferential attachment on tie existence (from a coefficient equal to 15.224 in Model 1 down to 4.069 in Model 4). Note that in order to compare different parameters' effect in ERGM results, it is essential to calculate and compare the *odds ratio*. However, it is possible to compare the change in the coefficient of the *same parameter* over *different* models to judge the change in the importance of the parameter among different models. Including the preferential attachment effect decreased the effect of gender and homophily of male authors disappeared. The effect of other variables had the same order and quite similar rates of Models 2 and 3.

To sum up, our results show that having a similar substantive research focus and sharing a similar geographical location had a positive effect on author collaboration. Collaboration was more likely between authors of the same seniority and between more junior and more senior authors. The general rule of the “rich get richer” was highly affected by the interplay between these variables. Note that the mixture of node attributes and community level effects model specification in Model 4 ensured a better fit than Models 1, 2 and 3 (see AIC and BIC measures in the last rows of the table, lower is better).

Conclusions and discussion

Our study provided an empirical overview of collaboration between sociologists in Italy and their international collaborators. We constructed a coauthorship network based on the publications indexed in Scopus and used community detection to detect communities in the giant component. Communities are relatively gender balanced (with one exception) and the highest gender difference in composition of members observed in a community was 23%. The communities differ in their research focus and international exposure. We ran ERGMs to control the effect of certain author attributes (i.e., gender, country of affiliation and scientific career), local structural configurations and community memberships (i.e., substantive focus) on coauthorship tie.

In general, we found that Italian sociology is a collection of isolated islands. The giant connected component could be split into five main communities. These communities showed a mixture of three main disciplinary sectors, i.e., sociology of culture and communication (present in two groups), economic sociology (present in three groups) and general sociology (present in three groups). The coauthorship pattern was mainly driven by preferential attachment and research focus. When we considered other author attributes, the effect of preferential attachment was reduced and gender differences were not significant anymore. Our findings confirmed that adopting a multi-level approach, while considering temporal dimensions in the study of scientific collaboration, can help to understand the interplay between factors of different levels (e.g., individual, community, covariate attributes and network structure levels).

Our results revealed five communities of sociologists that are relatively well connected among each other, though the composition of their members and the sectors represented in each community differ. This difference between the communities is reflected in the substantive focus of research of each community. Community 2 showed the highest gender difference and had the highest share of postdocs. The substantive focus of this community reflected the mix of *general sociology (SPS/07)* and *economic sociology (SPS/09)*. Community 0 had the highest continental and geographical diversity of members and showed substantive focus around *sociology of culture and communication (SPS/08)*.

The most interesting case among these five communities was community 4, the smallest in size, and relatively better gender balanced. This community is composed by Italians doing research across *general sociology (SPS/07)* and *sociology of culture and communication (SPS/08)* focused on topics such as *migration*, *home*, *labor migration*, *daily life* and *social status*. While being the smallest, this community had 4 out of the 7 most prolific authors among its members and once a layout based on influential ties constructing the backbone of network (Nick et al. 2013) was applied (see Fig. 4), it settles in center between all other communities as a small group of inter-mediators with three of its members among 8 nodes with highest *betweenness centrality*.

These five communities form the giant component of the coauthorship network of sociologists in Italy. We also found many other connected components of authors (511) publishing either alone or with few authors (Babchuk et al. 1999). Note that the size of all communities have increased in recent years, similarly to the overall trend found by Wuchty et al. (2007).

Returning to Moody (2004)'s study, our results would confirm the first hypothesis, i.e., a small world of distant communities with specific substantive research focus with shortest paths between them. Indeed, while we found many distant communities, the five main communities in the giant component shared specific research focuses with lowest possible ties with the other communities. This would indicate a high level of internal cohesion in each community.

Whether these communities were influenced by the specific university/department environments in which scientists are embedded could be subject to further inquiry. This would also require a more extensive sample coverage. Indeed, despite the fact that Scopus has the highest coverage among bibliographic data sources (especially in Italian language as shown in Mongeon and Paul-Hus (2016)), Italian sociologists regularly publish many articles, book chapters and monographs that are not indexed in Scopus. Previous research showed that only 63.81% of Italian sociologists had at least one publication record indexed in Scopus, which could be due to lower coverage of Italian language or local publication outlets (Akbaritabar et al. 2018). This could have limited the completeness of the coauthorship networks. Further research using other more comprehensive sources, such as Google Scholar, which includes more Italian publishers, could help to complete our analysis.

Furthermore, our analysis did not provide a robust explanation of the underlying mechanisms that account for these observed patterns. For instance, strategic decisions about collaboration could be constrained by certain factors, such as joint collaboration in research proposals, PhD programs, and scientific associations or academic mobility across institutions, which we could not consider here. Collaboration could also be inhibited by the institutional separation between different disciplines in Italy, which has a strong influence on grants, hiring and promotions and is reflected in our analysis of disciplinary sectors. Finally, coauthorship patterns could also reflect the capacity of certain scientists to forge international ties. Not only do international collaborations increase recognition and prestige of the most productive scientists; they in turn tend to stimulate network expansion leading to self-reinforcing processes (Leydesdorff et al. 2014). Studying this type of individual trajectories requires a mixed methods research design, capable of disentangling the motives and personal strategies of scientists. The formation of these communities cannot be explained as the mere effect of their specific substantive themes. Social network formation and evolution is a complex phenomenon that can be driven by different motives.

In addition, scientific cooperation can take many different shapes and formats (Katz and Martin 1997). Here, we have only focused on *coauthorship* with many limitations and underlying assumptions (e.g., see a discussion in Subramanyam (1983)) while considering also other forms of scientific interaction, such as *citations*, *funding proposal writing*, *conference* and *scientific events co-participation*, could provide a more comprehensive picture.

Appendix

Figure 6 presents the goodness of fit analysis for our ERGM model (It is only shown for the most extensive model, Model 4 in Table 4). It considers the model specification and estimates to what extent our model was able to detect the observed network's behavior. Considering that in order to control for preferential attachment our attention was on the degree distribution, the first panel in top left side of Fig. 6 indicates that our model predicted considerably well our observed network. However, it must be said that in other goodness of fit measures, which are based on edgewise shared partners (that we did not include in the model due to degeneracy issues) on top right, and the minimum geodesic distance (on which normally most ERGMs are not good) left bottom, our model did relatively bad. On the Goodness of fit evaluation based on covariates (bottom right of Fig. 6) our model did well.

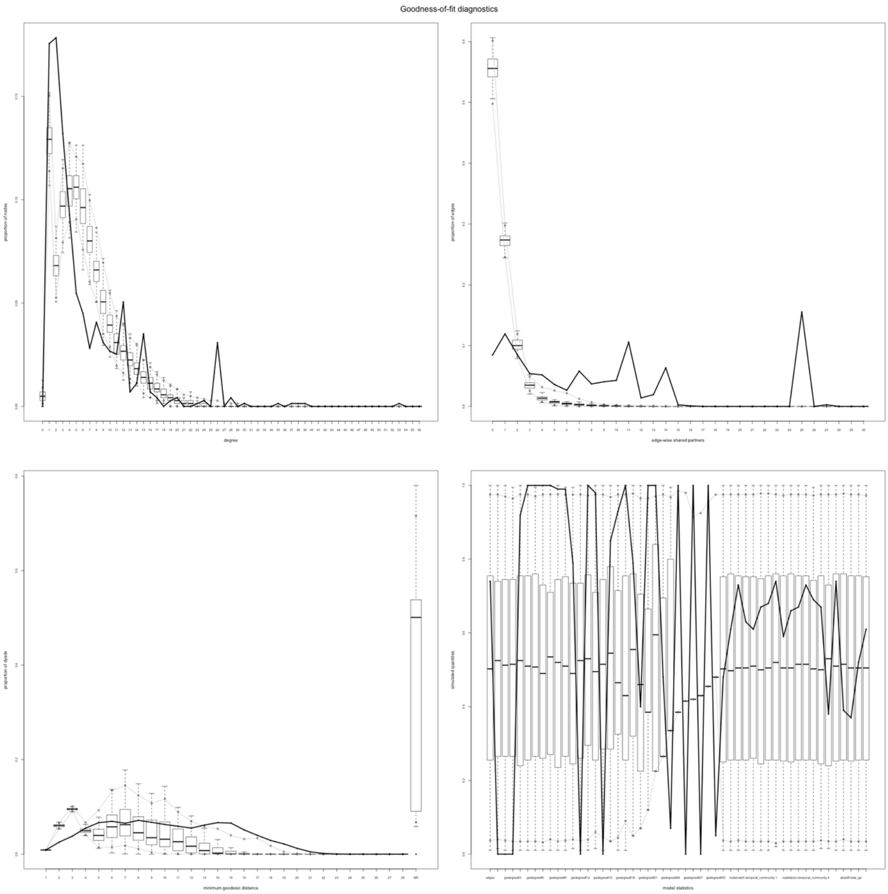

Fig. 6 Goodness of fit analysis of ERGM results (black solid line represents the observed network)

References

Abbott, A. (2000). Reflections on the future of sociology. *Contemporary Sociology, 29*(2), 296–300.

Abbott, A. (2001). *Chaos of disciplines*. Chicago: University of Chicago Press.

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Gender differences in research productivity: A bibliometric analysis of the italian academic system. *Scientometrics, 79*(3), 517–539.

Akbaritabar, A., Casnici, N., & Squazzoni, F. (2018). The conundrum of research productivity: A study on sociologists in italy. *Scientometrics, 114*(3), 859–882.

Babchuk, N., Keith, B., & Peters, G. (1999). Collaboration in sociology and other scientific disciplines: A comparative trend analysis of scholarship in the social, physical, and mathematical sciences. *The American Sociologist, 30*(3), 5–21.

Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: A network analysis of research on peer review. *Scientometrics, 113*(1), 503–532.

Bellotti, E., Kronegger, L., & Guadalupi, L. (2016). The evolution of research collaboration within and across disciplines in Italian Academia. *Scientometrics, 109*(2), 783–811. https://doi.org/10.1007/s11192-016-2068-1.

Bianchi, F., Casnici, N., & Squazzoni, F. (2018). Solidarity as a byproduct of professional collaboration: Social support and trust in a coworking space. *Social Networks, 54,* 61–72.

Cole, J. R., & Zuckerman, H. (1984). The productivity puzzle. In M. Steinkamp & M. L. Maehr (Eds.), *Advances in motivation and achievement. Women in science*. Greenwich: JAI Press.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695,* 1–9.

De Stefano, D., Fuccella, V., Vitale, M. P., & Zaccarin, S. (2013). The use of different data sources in the analysis of co-authorship networks and scientific performance. *Social Networks, 35*(3), 370–381.

Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science, 34*(1), 51–61.

Garvey, W. (1979). *Communication, the essence of science: Facilitating information exchange among librarians, scientists, engineers and students*. Amsterdam: Elsevier.

Grant, L., & Ward, K. B. (1991). Gender and publishing in sociology. *Gender and Society, 5*(2), 207–223.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., Bender-deMoll, S., & Morris, M. (2016). *Statnet: Software tools for the statistical analysis of network data*. The Statnet Project. Retrieved from CRAN.R-project.org/package = statnet. http://www.statnet.org. Accessed 12 Feb 2020.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). Statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software, 24*(1), 1–11. Accessed 12 Feb 2020.

Hargens, L. L. (2004). What is mertonian sociology of science? *Scientometrics, 60*(1), 63–70.

Hlavac, M. (2015). *Stargazer: Well-formatted regression and summary statistics tables*. Cambridge, USA: Harvard University. Retrieved from http://CRAN.R-project.org/package=stargazer. Accessed 12 Feb 2020.

Kastrin, A., Klisara, J., Lužar, B., & Povh, J. (2017). Analysis of slovenian research community through bibliographic networks. *Scientometrics, 110*(2), 791–813.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.

Kronegger, L., Ferligoj, A., & Doreian, P. (2011). On the dynamics of national scientific systems. *Quality & Quantity, 45*(5), 989–1015.

Lazega, E., Jourda, M.-T., Mounier, L., & Stofer, R. (2008). Catching up with big fish in the big pond? Multi-level network analysis through linked design. *Social Networks, 30*(2), 159–176.

Leahey, E. (2006). Gender differences in productivity: Research specialization as a missing link. *Gender and Society, 20*(6), 754–780.

Leahey, E., Keith, B., & Crockett, J. (2010). Specialization and promotion in an academic discipline. *Research in Social Stratification and Mobility, 28*(2), 135–155.

Leydesdorff, L., Park, H. W., & Wagner, C. (2014). International coauthorship relations in the social sciences citation index: Is internationalization leading the network? *Journal of the Association for Information Science and Technology, 65*(10), 2111–2126.

Long, J. S. (1992). Measures of sex differences in scientific productivity. *Social Forces, 71*(1), 159–178.

Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.

Marsden, P. V. (2002). Egocentric and sociocentric measures of network centrality. *Social Networks, 24*(4), 407–422.

Merton, R. K. (1968). The matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56–63.
Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of web of science and scopus: A comparative analysis. *Scientometrics, 106*(1), 213–228.
Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review, 69*(2), 213–238.
Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software, 24*(4), 1548.
Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics, 66*(1), 81–100.
Newman, M. E. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*(1), 016131.
Newman, M. E. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences, 98*(2), 404–409.
Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 026113.
Nick, B., Lee, C., Cunningham, P., & Brandes, U. (2013). Simmelian backbones: Amplifying hidden homophily in Facebook networks. In *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 525–532). ACM.
Ooms, J. (2014). *The jsonlite package: A practical and consistent mapping between json data and r objects* [stat.CO]. Retrieved from https://arxiv.org/abs/1403.2805. Accessed 12 Feb 2020.
Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*(7136), 664.
Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics, 10*(4), 1178–1195.
R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/. Accessed 12 Feb 2020.
Reichardt, J., & Bornholdt, S. (2004). Detecting fuzzy community structures in complex networks with a potts model. *Physical Review Letters, 93*(21), 218701.
Sciabolazza, V. L., Vacca, R., Okraku, T. K., & McCarty, C. (2017). Detecting and analyzing research communities in longitudinal scientific networks. *PLoS ONE, 12*(8), e0182516.
Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Journal of Information Science, 6*(1), 33–38.
Teele, D. L., & Thelen, K. (2017). Gender in the journals: Publication patterns in political science. *PS: Political Science and Politics, 50*(2), 433–447.
Traag, V. (2014). *Algorithms and dynamical models for communities and reputation in social networks*. Berlin: Springer.
Traag, V. A., Van Dooren, P., & Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E, 84*(1), 016114.
Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports, 9,* 1–12.
Turner, J. H. (2006). American sociology in chaos: Differentiation without integration. *The American Sociologist, 37*(2), 15–29.
van der Loo, M. P. J. (2014). The stringdist package for approximate string matching. *The R Journal, 6*(1), 111–122.
Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538.
van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and vos. *Journal of the Association for Information Science and Technology, 61*(12), 2405–2416.
Wallerstein, I. (2000). Where should sociologists be heading? *Contemporary Sociology, 29*(2), 306–308.
Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. New York: Springer. Retrieved from http://ggplot2.org. Accessed 12 Feb 2020.
Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse. Accessed 12 Feb 2020.
Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr. Accessed 12 Feb 2020.
Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039.

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology, 69*(1), 72–86.