



Research collaboration networks in maturing academic environments

Luís Filipe de Miranda Grochocki¹ · Andrea Felipe Cabello²

Received: 5 August 2022 / Accepted: 16 February 2023 / Published online: 7 March 2023
© Akadémiai Kiadó, Budapest, Hungary 2023

Abstract

We use data on research collaboration among 5,230 scholars in the University of São Paulo between 2000 and 2019 to understand how a network with high academic endogamy is structured, to identify if academic collaboration is more commonly found among those who share endogamy status, and to analyze if the likelihood of tie formation is distinct among inbred and non-inbred scholars. Results show growth of collaborations over time. However, ties between scholars are more likely to occur when endogamy status is shared by both inbred and non-inbred ones. Furthermore, such homophily effect seems to gradually be more influential on non-inbred scholars, suggesting this institution could be missing out on opportunities of exploring non-redundant information from within its own faculty members.

Keywords Academic inbreeding · Endogamy · Homophily · Social networks · Higher education · Brazil

Introduction

Research collaboration has been continuously growing in academia to increase scientific productivity, to share research costs, and to achieve new knowledge and interdisciplinary skills. Despite being predominantly found in the fields of science, technology, engineering, and math (STEM), it has been gaining relevance even in areas which have historically been less cooperative, such as the humanities and social sciences (Dahlander & McFarland, 2013; Wuchty et al., 2007).

As research collaboration becomes the norm, the study of the social networks of scientific communities has gained importance. Structural and relational studies analyze how individuals, communities and institutions interact and influence one another

✉ Andrea Felipe Cabello
andreafc@gmail.com; andreafc@unb.br

Luís Filipe de Miranda Grochocki
grochocki@stanford.edu

¹ Graduate School of Education, Stanford University, Stanford, CA, USA

² Faculdade de Economia, Administração e Contabilidade, Universidade de Brasília, Brasília, DF, Brazil

(Blau, 2017; S. P. Borgatti & Cross, 2003; Burt, 1992; Granovetter, 1973; Marsden, 1990; M. McPherson et al., 2001; Uzzi, 1997). Furthermore, the advancement of network analysis methods has given rise to several studies that explored the behaviors of collaborative academic communities (Barabási et al., 2002; S. Borgatti et al., 2009; De Montjoye, Stopczynski, Shmueli, Pentland, & Lehmann, 2014; Ding, 2011; Katz, 1994; Lee & Bozeman, 2005; Newman, 2000; Newman & Park, 2003; Zhang et al., 2018). However, a lot is yet to be learned as research collaborations are constantly evolving and researchers have not yet been able to unveil all the complexities of these networks, such as addressing how academic endogamy impacts research collaboration networks.

How are scientific collaboration groups structured in universities where so many scholars share the same alma mater? Endogamy is not limited to a few countries, it is found both in developed and developing countries. It affects both established and new institutions, which are often structurally constrained by their pool of applicants due to induced homophily, or occasionally even nepotism-choice homophily (Kossinets & Watts, 2009). Uncovering the type of research relationships that are built in an environment with high endogamy can be vital to understand the underlying causes of success and failure of scientific productivity in an array of institutions worldwide.

At first glance, a group in which members have similar characteristics might seem positive, given that they share local boundaries and have viewpoints that are more likely to converge. However, scientific work is not always performed by a homogeneous group of researchers. On the contrary, nowadays the process of coming up with innovative ideas requires crossing scientific boundaries such that a diverse range of actors and fields of knowledge can intersect (Cummings & Kiesler, 2005; Star & Griesemer, 1989). Networks with high homophily are characterized by strong provincial ties and vast numbers of links to redundant contacts, which result in a flow of repetitive information. Besides, clustered network structures, with their resulting lack of opportunities to contact external actors, may invariably limit the construction of new ideas (Burt, 1992, 2004; Granovetter, 1973; M. McPherson et al., 2001; Michelfelder & Kratzer, 2013). Despite the dilemma of local cohesion, it is in weak ties that members of homogeneous groups find some of their greatest opportunities for building and trading new ideas. Also known as structural holes, these network opportunities allow for well-positioned players to build bridges connecting actors from distinct clusters, thereby providing faster and more direct access to unique information (S. Borgatti et al., 2009; Burt, 1992, 2004; Granovetter, 1973, 1983; Hansen, 1999).

Elite research universities in Brazil hire a significant number of their own alumni as scholars among their faculty members. Brazil's most extreme case, the University of São Paulo (USP) has 70% of its faculty members hired from its own alumni pool. USP is not only the country's most prestigious and affluent university, but it is also commonly recognized among the top three universities in Latin America (Times Higher Education, 2021; QS World University Ranking, 2022). Moreover, it boasts the largest student enrollment of all Brazil's public universities and is considered the main birthplace of Brazilian professors. These characteristics make the USP an interesting case to analyze from a network analysis angle.

This study aims to shed light on how scientific collaborative communities are structured in elite research universities with high levels of academic endogamy. It further aspires to understand the dynamics of local scientific networks and their changes over time. Is such a homogeneous academic setting open to newcomers? Are scholars in those settings more responsive to further homophily?

Literature review

Social connections are more likely to occur among individuals who are alike, that is, who share physical attributes or have a similar educational level and socioeconomic background (Dahlander & McFarland, 2013; J. M. McPherson & Smith-Lovin, 1987; M. McPherson et al., 2001; Ruef, Aldrich, & Carter, 2003; Smith et al., 2016). The literature has focused on academic endogamy, a type of homophily that takes place in higher education institutions, with the hiring of faculty members that are also alumni to that school (Blau, 1973; Dutton, 1980; Hargens & Farr, 1973; McGee, 1960; Smyth & Mishra, 2014). What means to be an alumni may vary from author to author, although many focus on the most recent degree, usually the PhD degree (Delamont & Atkinson, 2001).

The literature is interested in these types of connections usually due to their possible impact on scientific productivity (Eisenberg & Wells, 2000; Horta, 2013; Horta et al., 2010; Inanc & Tuncer, 2011; Yudkevich & Sivak, 2012). However, it considers endogamy a byproduct of both individual choice and structural constraints (Kossinets & Watts, 2009; M. McPherson et al., 2001).

This provides an interesting case to be investigated with tools based on network analysis, especially Granovetter's Strength of Weak Ties theory. According to Granovetter (1983), weak ties allow for a wider diffusion of information as they can reach individuals connected to other social networks, whereas strong ties are more likely to bond similar individuals, thereby limiting the spread of information to its own cluster. This means that outsiders or non-inbred scholars could be a local source of non-redundant connections and leading to a more effective information diffusion in an environment where so many scholars share the same contacts.

This discussion is far from being consensual, however. Many have advocated for the relevance of strong ties, associating them with team excellence, stronger information diffusion patterns and the likelihood of change due to the familiarity these members already have with each other (Brown & Reingen, 1987; De Montjoye et al., 2014; Krackhardt, D., Nohria, N., & Eccles, 2003; (Rawlings et al., 2015, p.1717). In a more appealing tone, others suggest that a balance between strong and weak ties can be optimal for information exchange and creativity (Michelfelder & Kratzer, 2013; Zhou et al., 2009).

Burt (1992, 2004) also defends the role of weak ties in social networks. He argues that the similarity of ideas and attitudes within the group, which occurs with redundant contacts may reduce opportunities for the exchange of new knowledge. This highlights the role of bridge-builders, and their ability of obtaining new ideas that can be shared with other members in their own cluster. In other words, they are a point of access for new information.

The literature has given its attention to scientific collaboration and co-authorship. There are studies that focus on comparisons between fields, concluding that STEM areas tend to show more cooperative efforts (Dahlander & McFarland, 2013; Wuchty et al., 2007), but also studies that focus on research impact and productivity (Li et al., 2013, Bordons et al., 2015) or even the effects of the Covid-19 pandemic on co-authorship networks (Sachini et al., 2021).

Data sources differ, however. Kossinets, G., & Watts, D. J (2009) used e-mail interactions and course registrations to identify relationships in a large US University. Zhang et al. (2018) used papers extracted from Web of Science for a coauthoring analysis. Hâncean, M. G., & Perc, M. (2016) used a similar strategy restricted their analysis to sociology and Eastern Europe countries. The main conclusion of this literature is that when homophily

is present, highly productive authors tend to work together, increasing output inequalities, which may be bad for the system output as whole.

Research questions

This study explores the structure and characteristics of research collaboration in a homogeneous scientific community and, therefore, contributes to the relevant literature on endogamy, homophily and social networks. Its main hypothesis is that scientific cooperation in an environment with high academic endogamy is largely influenced by homophily, which could lead to collaboration clusters of inbred scholars, making it hard for non-inbred scholars to form local ties. Thus, this paper addresses the following questions:

- (1) How are local research collaboration networks structured in elite research universities with high levels of academic endogamy? Do they change over time?
- (2) Does homophily influence the formation of ties between faculty members? Are scholars with identical academic endogamy status more likely to work together? Are these preferences maintained over time?

The case of professors in the University of São Paulo could lead to important findings helpful not only to the Brazilian higher education system, but also to the evaluation and planning of educational policies in other countries, mainly ones with the same level of system maturity as Brazil, where academic endogamy also occurs.

This study contributes to this literature in four ways: (i) by using an official dataset, in which affiliation and publications are self-reported but submitted for regulatory review, it differs from most studies in which data is obtained from the same sources (usually Web of Science) and usually only from publications in English—here we include publications in other languages as well; (ii) by analyzing the case of one single large ego network, which also happens to be the largest and most elitist university network in Brazil; (iii) a network with 70% level of endogamy; (iv) an analysis of co-authoring pattern in a developing country. For all these reasons, we believe this study is relevant and contributes greatly to the literature.

Data

The University of São Paulo (USP) is the best ranked Brazilian University on Times Higher Education rankings and QS University rankings. It was selected as a case study for this study due to its status in Brazil and to its high levels of endogamy¹—70% in 2016 (Grochocki, 2020). USP is a public university with an enrollment of 97,982 students and 5631 faculty members. Because it is such an important institution for higher education in Brazil, it is also responsible for the PhD degrees of 24.4% of active scholars in the whole system.

The dataset used was collected by CAPES (Coordination for the Improvement of Higher Education Personnel), the Brazilian government agency responsible for the establishment,

¹ For this study, endogamy is identified when former students are employed as scholars by their alma mater after they obtained their final degree (PhD).

evaluation, and financing of graduate programs in Brazil. Scholars in Brazil maintain an official profile with their *curriculum vitae* at “Plataforma Lattes”, an online academic resume database managed by CNPq (Brazilian National Council for Scientific and Technological Development), with information on researchers’ education, language skills, current and past employment, scientific publications, awards, and grants, among others. Information on these resumes is self-reported, but used for official purposes of funding and regulation.

An open-source Python program called “ScriptLattes” (Mena-Chalco & Cesar Junior, 2009) was used to extract the online data available on the scientific production of 5,230 unique ID numbers identified as USP professors from the years 2000 to 2019.

The total sample of scientific production of the 5,230 scholars represents 93% of the university’s faculty population, in all fields of knowledge. We excluded those scholars who were not linked to any graduate program. The final sample of 5,230 scholars led us to 196,941 journal papers, 71,239 conference papers and 18,992 books. More information is given in the appendix.

Collaboration information was split into five groups composed of four years of aggregated data each: 2000–2003, 2004–2007, 2008–2011, 2012–2015, and 2016–2019. An adjacency matrix on coauthorship data was created to run different methods of collaboration network analyses. Tables 1, 2 summarize descriptive information on the values of variables on this sample, while Table 3 shows how variables are correlated.

Finally, the dataset was restructured from individual to dyadic format. In such a layout, every row corresponds to a potential or actual tie formed between two scholars within a determined period (4 years in the case of this study). Thus, rows describe not only tie characteristics, but also attributes of both individuals. The first listed scholar is referred to as an “ego” and its immediate contact as an “alter”. Equal pairs of scholars were given a unique “dyad ID” to match the same ties in distinct time periods. Following, a binary variable “tie” was generated as 1 for all 53,526 ties which were identified as having taken place in the last 20 years based on the coauthoring of journal and conference papers, and books. All other potential ties received a 0.

Considering the extensive number of potential ties (close to 137 million), a sample of 5 million of those was randomly selected while keeping every actual tie (Kleinbaum et al., 2013). This method was chosen as ties among older professors (former advisors/teachers) and young faculty members (former advisees/students) are expected in an environment with high endogamy. Furthermore, this study aims to illustrate cross-disciplinary collaboration. Therefore, the adoption of other methods, such as selecting potential ties based on the absolute difference in hiring year (Dahlander & McFarland, 2013) or limiting ties within fields, would not allow to describe those relationships. As expected, no significant differences were found when comparing the variables of the full and the randomly selected tie sample.

Method

This study uses descriptive social network analysis (SNA) methods, as well as multilevel modeling (MLM). At the network level, methodologies and tools were adopted to replicate and measure characteristics of complete networks. Images were produced using the program Gephi, version 0.9.2., layout method Force Atlas2 and shape method Polygon. Edge weights were rescaled to a normalized range.

Table 1 Summary of USP's Academic Collaboration Networks from 2000 to 2019

Period	Edges	Weakly connected scholars or that do not form any ties	Average path length	Average clustering coefficient	Observations
2000–2003	6,466	2,542	6.8	0.446	There is a big collaboration cluster in the medical and health sciences while interdisciplinary fields scholars are spread collaborating with members from distinct areas. Collaboration in the humanities was limited and can hardly be noticed in this larger network
2004–2007	9,359	1,930	6.1	0.43	Collaboration expanded considerably; The number of weakly connected scholars dropped to 1,930 which indicates scholars intensified their collaborations. Medical and health sciences remain a large collaboration cluster, now with many ties to the natural sciences. Engineering and technology share smaller clusters perhaps due to their interdisciplinary collaboration with other fields
2008–2011	12,173	1,648	6.2	0.4	The number of weakly and unconnected nodes drops again to 1,648, suggesting another increase in research collaboration over time. Enhanced importance of interdisciplinary academic collaborations. It is likely the period with the highest interdisciplinary academic collaborations
2012–2015	13,518	1,621	5.7	0.39	In this period, research collaboration networks reached 13,518 edges. Its peak in connectivity is also confirmed by the shortest average path length of 5.7 and lowest number of connected components. The average clustering coefficient dropped due to an increase in field diversity. Faculty groups in the social sciences were located on the edges of this larger network while the hard sciences remained in its core
2016–2019	12,010	1,859	5.6	0.38	The number of ties among USP scholars declined, but this would be expected as the number of published books, papers in journals, and conferences also dropped in 2019. This could be due to under reporting of recent scientific productivity in faculty member's curriculum vitae. Medical, health sciences and agricultural sciences remained very clustered in their academic collaboration while scholars in more interdisciplinary fields were spread. Social sciences and humanities were mostly positioned on the edges of the large network. Scholars in engineering, social sciences, natural sciences, and medical and health sciences seemed to also be connected to a variety of faculty members from other fields

Table 2 Logistic regression model of predictors of tie formation (2000–2019)

Variables ^a	(1) MLM logit	(2) MLM–Pref homophily	(3) MLM–over time
j is non-inbred scholar	0.035 (0.022)	0.177*** (0.027)	0.091* (0.054)
j is female	0.098*** (0.017)	0.098*** (0.017)	0.099*** (0.018)
j had intern'l academic mobility	-0.114*** (0.029)	-0.113*** (0.029)	-0.104*** (0.029)
j shares same field with i	2.696*** (0.029)	2.695*** (0.029)	2.695*** (0.029)
j has same gender as i	0.282*** (0.016)	0.283*** (0.016)	0.282*** (0.016)
j is non-inbred or inbred scholar and has same endogamy status as i	0.244*** (0.013)	0.340*** (0.017)	0.409*** (0.038)
j is non-inbred scholar and has same endogamy status as i		-0.296*** (0.031)	-0.447*** (0.067)
j is non-inbred scholar in 2004–2007			0.012 (0.058)
j is non-inbred scholar in 2008–2011			0.041 (0.063)
j is non-inbred scholar in 2012–2015			0.179*** (0.062)
j is non-inbred scholar in 2016–2019			0.143** (0.061)
j shares same endogamy status with i in 2004–2007			-0.040 (0.044)

Table 2 (continued)

Variables ^a	(1) MLM logit	(2) MLM–Pref homophily	(3) MLM–over time
j shares same endogamy status with i in 2008–2011			-0.093** (0.045)
j shares same endogamy status with i in 2012–2015			-0.050 (0.046)
j shares same endogamy status with i in 2016–2019			-0.128*** (0.048)
j is non-inbred scholar and has same end. status with i in 2004–2007			0.139* (0.077)
j is non-inbred scholar and has same end. status with i in 2008–2011			0.171** (0.080)
j is non-inbred scholar and has same end. status with i in 2012–2005			0.141* (0.081)
j is non-inbred scholar and has same end. status with i in 2016–2019			0.238*** (0.082)
Constant	-6.140*** (0.042)	-6.206*** (0.042)	-6.213*** (0.050)
Observations	4,912,929	4,912,929	4,912,929
j's individual characteristics	YES	YES	YES
j's individual scientific production information	YES	YES	YES
Year dummies	YES	YES	YES
Number of groups	5,219	5,219	5,219
EgoID var(_cons)	0.215 (0.013)	0.214 (0.013)	0.214 (0.013)
Robust standard errors in parentheses			

Table 2 (continued)

Variables ^a	(1) MLM logit	(2) MLM–Pref homophily	(3) MLM–over time
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$			

^a i_1 and j denote two different members of the dyad

Relational data challenges the assumption that observations are independent of one another. Consequently, multilevel modeling (MLM) has been widely adopted to address this limitation of Ordinary Least Squares regressions when analyzing ego networks. Furthermore, MLM avoids both ecological and atomistic fallacy, allowing for cross-level inferences. Multilevel modeling simultaneously estimates the variance within and between groups for an outcome variable, and its association with individual and group independent variables (Crossley et al., 2015; Peugh, 2010; Rabe-Hesketh & Skrondal, 2012; Snijders & Bosker, 2012; Snijders, Spree, & Zwaagstra, 1995). Among others, MLM holds that Level 1 residual variance is assumed to be constant within and between Level 2 units and that Level 1 and level 2 residuals are assumed to be uncorrelated (Perry, Pescosolido, & Borgatti, 2018).

To correct for heteroscedasticity, standard errors were clustered robust at the ego level. Besides, an unstructured covariance matrix was adopted to maintain the assumption of uncorrelated residuals.

The main model for this study is described in the equation:

$$\text{logit}(Y_{ij}) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2j} + \beta_{3j}x_{3ij} + \beta_{4j}x_{2j}x_{3ij} + U_{0j}$$

where Y_{ij} is the outcome variable of interest “tie” between j (ego) and i (alter). β_{ij} represents random differences between groups, where β_{0j} equals the average intercept plus group-dependent deviation U_{0j} . X_{1ij} serves as characteristics of an individual (level 1) in an ego network j (level 2). Likewise, x_{2j} exhibits characteristics of group ego j . Following, x_{3ij} represents the association of homophily of alter and ego shared traits. Finally, $x_{2j}x_{3ij}$ depicts homophily by ego interaction terms. U_{0j} is an ego-level (level 2) residual (error) term. Thus, $\sigma^2_{U_{0j}}$ represents the magnitude of variation found among the average tie-values. Clustered robust standard errors were computed for all models at the ego level.

While non-inbred (dummy) is the treatment variable, other individual characteristics of egos and alters are female (dummy), age (continuous), years of experience (continuous), academic experience abroad (dummy), quantity of Postdoctoral researchers (continuous), PhD (continuous) and Masters’ students (continuous), and undergraduates (continuous), number of published papers in journals (continuous) and conferences (continuous), number of published books (continuous) and book chapters (continuous), and fields of study (categorical). Tie characteristics are degree (continuous), same academic endogamy origin (dummy), same gender (dummy), same field (dummy), and year period (categorical). The estimates of all models can be found on Table 4 in the results section.

Results and discussion

As discussed in the previous sections, Collaboration information was split into five groups composed of four years of aggregated data each: 2000–2003, 2004–2007, 2008–2011, 2012–2015, and 2016–2019. Inbred professors are square shaped while non-inbred ones are represented by circles. Scholars’ fields are represented by colors: medical and health sciences (pink), social sciences (light green), natural sciences (blue), humanities (orange), engineering and technology (brown), interdisciplinary (red) and agricultural sciences (dark green).

Although a large university wide connected network was identified for every four-year period, a significant number of nodes neither connected to the broad university network nor formed any local collaboration (image with all nodes included at the left corner).



Fig. 1 USP’s academic collaboration between 2000–2003

Figures 1, 2, 3, 4, and 5 represent the academic community of professors in the University of São Paulo and their local scientific collaborations between years 2000 and 2019, grouped into four-year periods while Table 5 summarizes the information in those figures.

Figures 1, 2, 3, 4, and 5 and Table 5 show collaboration clusters within the University of São Paulo have increased over time. However, the core of these networks remain to be the STEM fields, with groups in the social sciences positioned on the edges, and collaboration in the humanities rather limited and barely showing in our figures. Interdisciplinary academic collaboration remained scattered during the whole period considered in our sample, which is expected, given its possible ties with several other fields. In other words, our data shows that, in the case of USP, collaboration happens within fields or, at the most, to loosely connected fields, as the strong clusters that emerged show.

Medical and health sciences were, and still are, the most important feature of these networks throughout the whole period considered. However, these clusters expanded and developed ties with researcher in other fields, such as the natural sciences, engineering and social sciences, which gained importance in these networks over the years.

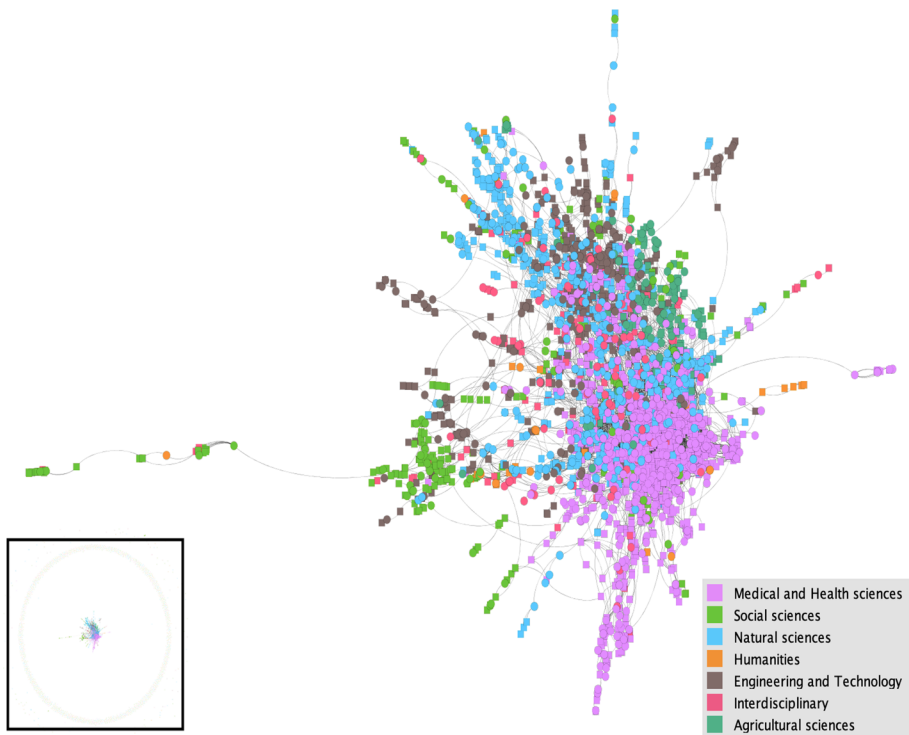


Fig. 2 USP's academic collaboration between 2004–2007

Despite representing nodes with squares for inbred scholars and with circles for non-inbred ones, the visual analyses based on Figs. 1–5 would be limited. Thus, to further analyze the issue, we now turn to our multilevel logit model, which focus on evaluating how collaboration is affected by endogamy status. Since, in our data, positive events represent around 1% of the total sample and, therefore, the dataset is a sparse matrix, rare event bias could be a concern. However, this is not the case in our sample.

The first model is a multilevel logit model with fixed effects (level 1) and random effects for ego (level 2). This model includes individual characteristics of ego (j) and homophily tie traits with alter (i) as control variables. Standard errors were clustered at the ego level. Following, interaction term effects were added to Model 2 to compare the likelihood of tie formation of four groups: non-alumni inbred (j) to non-inbreds (i), inbreds (j) to inbreds (i), non-inbreds (j) to inbreds (i), and inbreds (j) to non-inbreds (i). Furthermore, Model 3 adds new interaction terms of mutual endogamy status trait with the five distinct year periods. This model contributes to the analysis of homophily effects changes on the likelihood of ties over time.

Like previous studies (Dahlander & McFarland, 2013; M. McPherson et al., 2001), results indicate that the hypothesis that homophily influences the establishment of academic collaboration among faculty members seems to be true in the case of those who share the same academic endogamy status. Data on joint academic publications show inbred scholars are more likely to hold ties with inbreds, as well as non-inbreds with their non-inbred colleagues. This means that those with endogamy ties to the university seem to

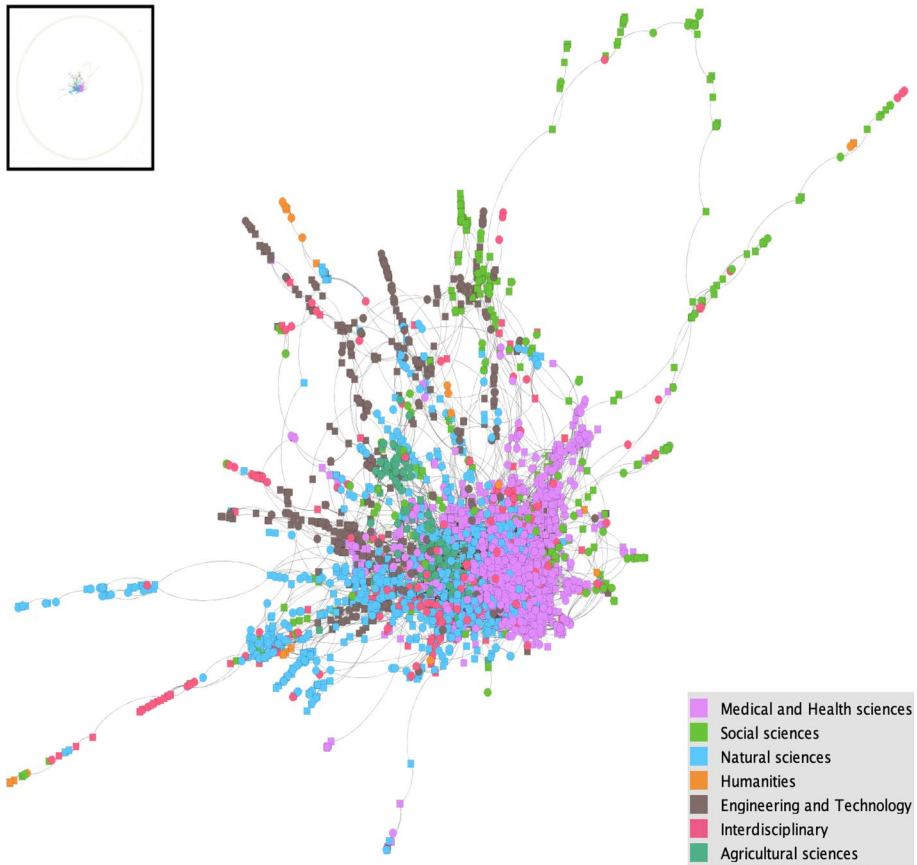


Fig. 3 USP’s academic collaboration between 2008–2011

collaborate more among themselves. Likewise, those who received their PhD degrees elsewhere seem to collaborate more among themselves, which suggest that research networks within the university may not be as integrated and diverse as they could be.

Ties among faculty members of distinct endogamy status occur, but they seem to be less likely than among those who share endogamy status homophily. This trend could be the result of endogamy status homophily influencing the formation of local ties among inbred scholars which in turn only gives non-inbred scholars the option of collaborating with each other or with scholars outside the University of São Paulo. Considering non-inbred scholars are a minority of around 30% of the faculty body in the USP, a higher probability was expected of ties among non-inbreds and inbreds based on numbers alone. However, our models show collaboration is more likely to be found amongst those who share the same endogamy status.

Outcomes suggest that the slope for these homophily effects seem to become steeper every four-year period among non-inbreds. Therefore, it is likely that endogamy status preference is more influential on non-inbred faculty members over the years.

Besides these findings on the effects of shared endogamy status, other homophily characteristics also seem to impact the likelihood of academic collaborations. Collaboration

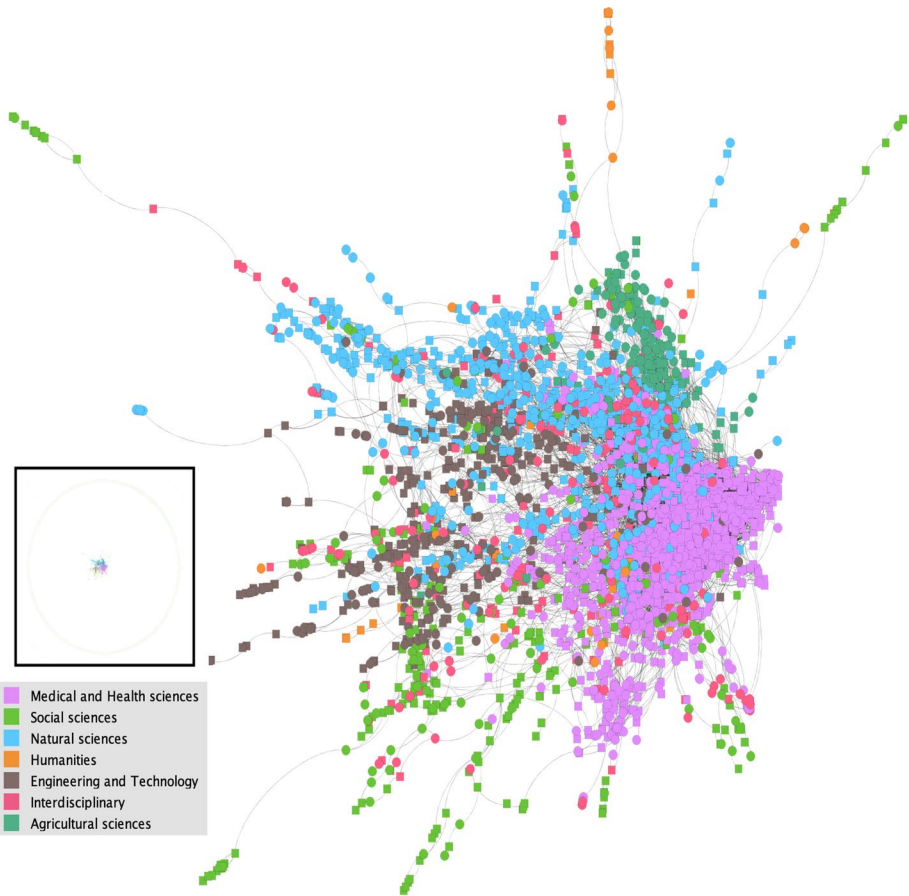


Fig. 4 USP's academic collaboration between 2012–2015

networks are more likely to be found among those scholars of same gender and field. Females are also more likely to contribute to academic collaboration. On the other hand, if scholars were subjected to international academic mobility, they are less likely to collaborate within the University, suggesting that scholars with these kinds of experiences abroad may prefer collaborating academically with their external networks, to do research by themselves or to pursue other types of collaboration within their own university.

Conclusion

Results show that local research collaboration has been growing among University of São Paulo faculty members. Both inbred and non-inbred scholars have been benefitting from the opportunity of cooperating with their university colleagues. Notwithstanding, there is still a significant number of them who is weakly or not at all connected with their co-workers.

The hypothesis that a homogeneous setting is prone to an increased likelihood of ties being formed among faculty members who share mutual characteristics is confirmed.

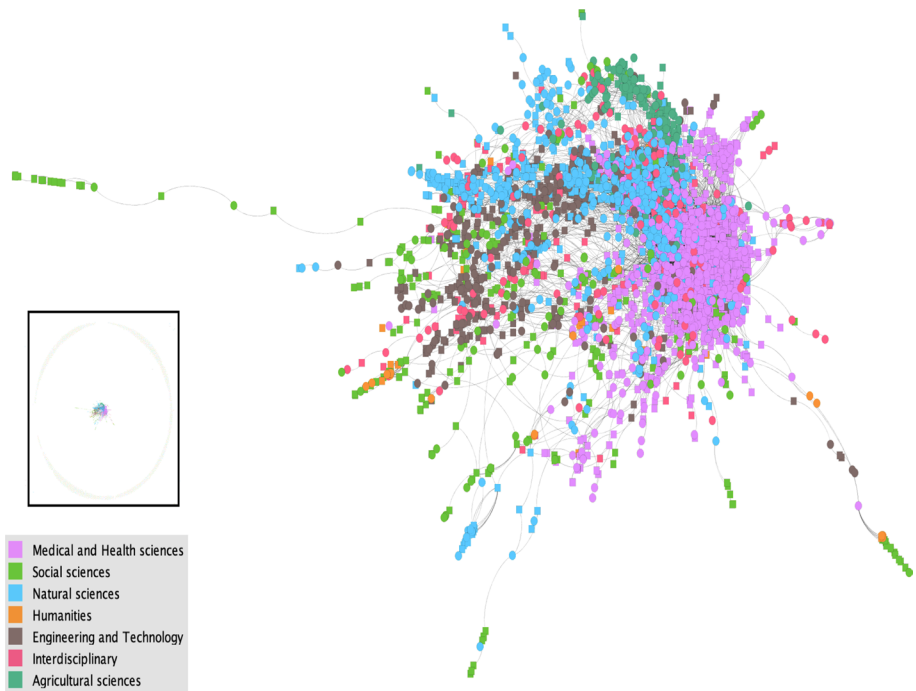


Fig. 5 USP’s academic collaboration between 2016–2019

Same academic endogamy status is a highly statistically significant predictor of local research collaboration. That would be expected of inbred scholars for being a majority group which already knows the local culture and shares mutual contacts. However, non-inbred faculty members were also more likely to build collaboration ties among themselves. This could mean that research networks are not as connected and integrated as they could be. Furthermore, outcomes suggest that over time endogamy status preferences get stronger for non-inbred scholars. In other words, non-inbred scholars are more likely to increase their collaboration among themselves over the years. We would have expected that years of work at the university would have allowed for these scholars to integrate and to establish new and more intense collaborations with local inbred scholars and their established research clusters, but that does not seem to be the case in the University of São Paulo.

High academic endogamy may further promote the bond of similar individuals in local research collaboration networks. Consequently, these communities might be isolating their faculty members who were trained at other academic institutions, leading to segregated networks. Over time, such practice might discourage newcomers to integrate already established research clusters, pushing these non-inbred scholars to create their own local ties.

Universities with high endogamy could be neglecting a high valued local resource of non-redundant contacts and their connections and the possibilities it brings for increased more diverse collaboration, internationalization, and research productivity. Such behavior could limit opportunities for new information to be exchanged and for knowledge to be jointly produced. Non-inbred faculty members have the potential to form bridges

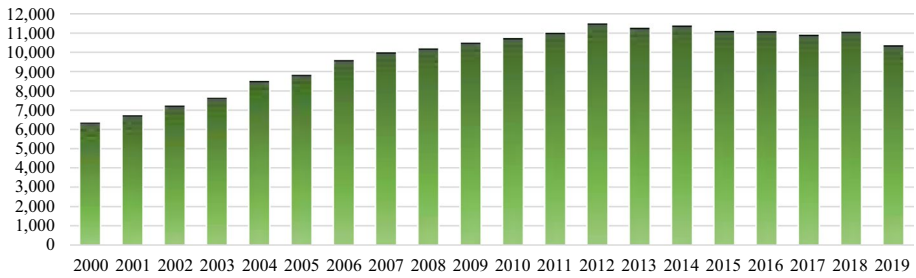


Fig. 6 Total number of published papers per year

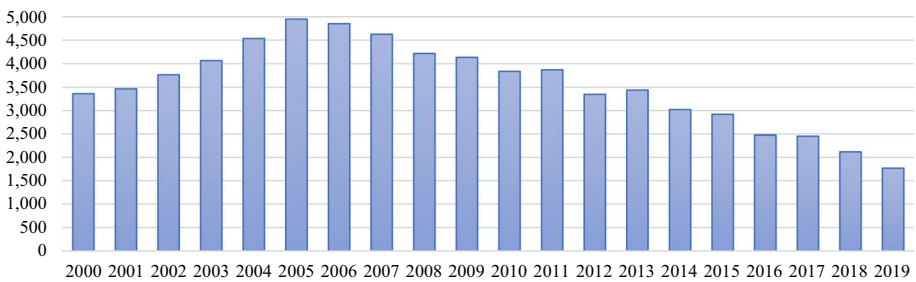


Fig. 7 Total number of conference papers per year

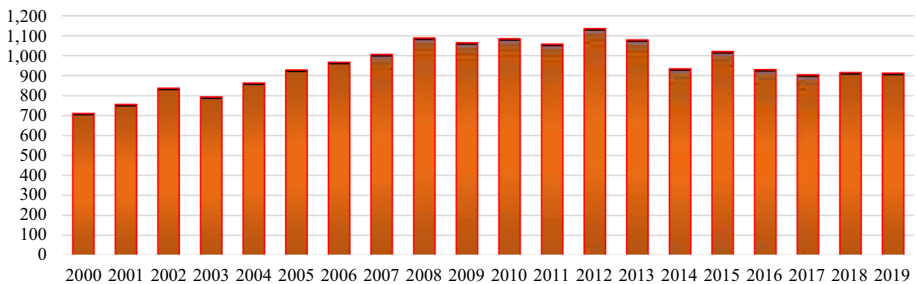


Fig. 8 Total number of published books per year

connecting inbred scholars to scientific network contacts outside their own departments and university. Perhaps, a more balanced environment is the optimal format for information exchange and creativity to flourish within and outside universities.

Appendix

Summary statistics

Group of years

See Tables 3, 4, 5, 6, 7, 8 and Figs. 6, 7, 8.

Table 3 Summary statistics 2000–2003

	N	Mean	Sd	Min	Max
Tie	977,895	.006	.08	0	1
j is Non-Inbred scholar	977,895	.349	.477	0	1
j is female	977,895	.392	.488	0	1
j's age	977,895	43.588	10.641	25	75
j's years of experience	977,895	8.338	8.881	0	49
j had intern'l academic mobility	977,895	.178	.383	0	1
j's n. pub. papers	977,895	7.116	9.931	0	144
j's n. pub. books	977,895	.658	1.954	0	79
j's n. book chapters	977,895	2.159	5.099	0	127
j's n. conference papers	977,895	3.334	7.873	0	234
j's n. Postdoc researchers	977,895	.07	.372	0	6
j's n. PhD advisees	977,895	.948	1.87	0	20
j's n. Master's advisees	977,895	1.587	2.752	0	64
j's n. Undergrad advisees	977,895	1.153	2.577	0	47
j's degree centrality	977,895	2.634	4.914	0	67
Same endogamy status	977,895	.546	.498	0	1
Same gender	977,895	.524	.499	0	1
Same field	977,895	.184	.388	0	1

Table 4 Summary statistics-2004–2007

	N	Mean	Sd	Min	Max
Tie	982,539	.009	.097	0	1
j is Non-Inbred scholar	982,539	.352	.477	0	1
j is female	982,539	.391	.488	0	1
j's age	982,539	46.591	11.044	25	75
j's years of experience	982,539	11.112	9.453	0	48
j had intern'l academic mobility	982,539	.183	.387	0	1
j's n. pub. papers	982,539	9.511	12.129	0	220
j's n. pub. books	982,539	.777	2.066	0	71
j's n. book chapters	982,539	2.951	6.118	0	134
j's n. conference papers	982,539	4.127	8.538	0	108
j's n. Postdoc researchers	982,539	.179	.649	0	10
j's n. PhD advisees	982,539	1.21	1.954	0	20
j's n. Master's advisees	982,539	1.866	2.747	0	71
j's n. Undergrad advisees	982,539	1.789	3.317	0	74
j's degree centrality	982,539	3.692	5.721	0	79
Same endogamy status	982,539	.545	.498	0	1
Same gender	982,539	.524	.499	0	1
Same field	982,539	.185	.388	0	1

Table 5 Summary statistics-2008–2011

Tie	985,134	.012	.11	0	1
j is Non-Inbred scholar	985,134	.347	.476	0	1
j is female	985,134	.395	.489	0	1
j's age	985,134	50.1	10.943	25	75
j's years of experience	985,134	14.509	9.615	0	49
j had intern'l academic mobility	985,134	.186	.389	0	1
j's n. pub. papers	985,134	11.44	13.94	0	164
j's n. pub. books	985,134	.886	2.27	0	78
j's n. book chapters	985,134	3.349	5.884	0	109
j's n. conference papers	985,134	3.535	7.643	0	97
j's n. Postdoc researchers	985,134	.336	.919	0	13
j's n. PhD advisees	985,134	1.427	2.028	0	19
j's n. Master's advisees	985,134	2.17	2.56	0	29
j's n. Undergrad advisees	985,134	2.187	3.551	0	68
j's degree centrality	985,134	4.814	7.118	0	80
Same endogamy status	985,134	.548	.498	0	1
Same gender	985,134	.523	.499	0	1
Same field	985,134	.187	.39	0	1

Table 6 Summary statistics-2012–2015

Tie	985,611	.013	.115	0	1
j is Non-Inbred scholar	985,611	.335	.472	0	1
j is female	985,611	.425	.494	0	1
j's age	985,611	53.59	10.555	28	75
j's years of experience	985,611	17.752	9.138	0	53
j had intern'l academic mobility	985,611	.171	.376	0	1
j's n. pub. papers	985,611	12.5	15.788	0	267
j's n. pub. books	985,611	.886	2.342	0	88
j's n. book chapters	985,611	3.189	5.747	0	110
j's n. conference papers	985,611	2.779	6.764	0	128
j's n. Postdoc researchers	985,611	.527	1.205	0	15
j's n. PhD advisees	985,611	1.717	2.039	0	17
j's n. Master's advisees	985,611	2.285	2.394	0	29
j's n. Undergrad advisees	985,611	2.187	3.278	0	41
j's degree centrality	985,611	5.43	7.871	0	80
Same endogamy status	985,611	.556	.497	0	1
Same gender	985,611	.513	.5	0	1
Same field	985,611	.188	.391	0	1

Table 7 Summary statistics-2016–2019

Tie	981,750	.012	.108	0	1
j is Non-Inbred scholar	981,750	.334	.471	0	1
j is female	981,750	.426	.494	0	1
j’s age	981,750	56.653	9.904	32	75
j’s years of experience	981,750	21.008	8.458	3	50
j had intern’l academic mobility	981,750	.176	.381	0	1
j’s n. pub. papers	981,750	12.112	15.805	0	177
j’s n. pub. books	981,750	.786	2.056	0	71
j’s n. book chapters	981,750	2.809	5.234	0	78
j’s n. conference papers	981,750	2.015	5.641	0	108
j’s n. Postdoc researchers	981,750	.449	1.04	0	17
j’s n. PhD advisees	981,750	1.436	1.839	0	21
j’s n. Master’s advisees	981,750	1.917	2.197	0	33
j’s n. Undergrad advisees	981,750	1.581	3.315	0	103
j’s degree centrality	981,750	4.981	7.371	0	71
Same endogamy status	981,750	.556	.497	0	1
Same gender	981,750	.511	.5	0	1
Same field	981,750	.187	.39	0	1

Table 8 Matrix of correlations

Tie	985,611	.013	.115	0	1
j is Non-Inbred scholar	985,611	.335	.472	0	1
j is female	985,611	.425	.494	0	1
j’s age	985,611	53.59	10.555	28	75
j’s years of experience	985,611	17.752	9.138	0	53
j had intern’l academic mobility	985,611	.171	.376	0	1
j’s n. pub. papers	985,611	12.5	15.788	0	267
j’s n. pub. books	985,611	.886	2.342	0	88
j’s n. book chapters	985,611	3.189	5.747	0	110
j’s n. conference papers	985,611	2.779	6.764	0	128
j’s n. Postdoc researchers	985,611	.527	1.205	0	15
j’s n. PhD advisees	985,611	1.717	2.039	0	17
j’s n. Master’s advisees	985,611	2.285	2.394	0	29
j’s n. Undergrad advisees	985,611	2.187	3.278	0	41
j’s degree centrality	985,611	5.43	7.871	0	80
Same endogamy status	985,611	.556	.497	0	1
Same gender	985,611	.513	.5	0	1
Same field	985,611	.188	.391	0	1

Acknowledgements A previous version of this research is featured in the fourth chapter of Grochocki's PhD dissertation, "More of the Same? The Structure of Research Collaboration Networks in Homogeneous Academic Environments" (Grochocki, 2020).

Funding Capes, 99999.000225/2016-09, Luis Filipe Grochocki.

Declarations

Conflict of interest Luis Grochocki has employment ties and has received funding from Capes. Andrea Cabello has no affiliations nor other types of conflict of interest.

References

- Barabási, A., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica a: Statistical Mechanics and Its Applications*, 311(3–4), 590–614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)
- Berelson, B. (1960). Graduate education in the United States. *Carnegie series in American education*. McGraw-Hill.
- Blau, P. M. (1973). *The Organization of Academic Work* (Vol. 11). Transaction Publishers.
- Blau, P. M. (2017). Exchange and power in social life. In *Exchange and Power in Social Life*. <https://doi.org/10.4324/9780203792643>
- Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, 49(4), 432–445. <https://doi.org/10.1287/mnsc.49.4.432.14428>
- Borgatti, S., Mehra, A., Brass, D., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, 323(April), 892–896. <https://doi.org/10.1126/science.1165821>
- Brown, J. J., & Reingen, P. H. (1987). Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research*, 14(3), 350. <https://doi.org/10.1086/209118>
- Burt. (1992). *Structural Holes: The Social Structure of Competition*. Harvard University Press.
- Burt. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2), 349–399. <https://doi.org/10.1086/421787>
- Crossley, N., Bellotti, E., Edwards, G., Everett, M. G., Koskinen, J., & Tranmer, M. (2015). *Multilevel Models For Cross-Sectional Ego-Nets*. Social Network Analysis for Ego-Nets. <https://doi.org/10.4135/9781473911871>
- Cruz-Castro, L., & Sanz-Menéndez, L. (2010). Mobility versus job stability: Assessing tenure and productivity outcomes. *Research Policy*, 39(1), 27–38. <https://doi.org/10.1016/j.respol.2009.11.008>
- Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, 35(5), 703–722. <https://doi.org/10.1177/0306312705055535>
- Dahlander, L., & McFarland, D. A. (2013). Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly*, 58(1), 69–110. <https://doi.org/10.1177/0001839212474272>
- De Montjoye, Y. A., Stopczynski, A., Shmueli, E., Pentland, A., & Lehmann, S. (2014). The strength of the strongest ties in collaborative problem solving. *Scientific Reports*. <https://doi.org/10.1038/srep05277>
- Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5(1), 187–203. <https://doi.org/10.1016/j.joi.2010.10.008>
- Dutton, J. K. (1980). The impact of Inbreeding and Immobility on the Professional Role and Scholarship Performance of Academic Scientists. *Annual Meeting of the American Educational Research Association*, 33. Retrieved from <http://eric.ed.gov/?id=ED196714>
- Eisenberg, T., & Wells, M. T. (2000). Inbreeding in Law School Hiring: Assessing the Performance of Faculty Hired from within. *The Journal of Legal Studies*, 29(S1), 369–388. <https://doi.org/10.1086/468077>
- Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380. <https://doi.org/10.1086/225469>
- Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 1, 201. <https://doi.org/10.2307/202051>
- Grochocki, L. F. (2020). *Academic endogamy in Brazil and its influences on faculty productivity and collaboration*. Stanford University.
- Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, 44(1), 82. <https://doi.org/10.2307/2667032>

- Hargens, L. L., & Farr, G. M. (1973). An Examination of Recent Hypotheses About Institutional Inbreeding. *American Journal of Sociology*, 78(6), 1381–1402. Retrieved from <https://www.jstor.org/stable/2776393>
- Horta, H., Veloso, F. M., & Grediaga, R. (2010). Navel Gazing: Academic Inbreeding and Scientific Productivity. *Management Science*, 56(3), 414–429. <https://doi.org/10.1287/mnsc.1090.1109>
- Inanc, O., & Tuncer, O. (2011). The effect of academic inbreeding on scientific effectiveness. *Scientometrics*, 88(3), 885–898. <https://doi.org/10.1007/s11192-011-0415-9>
- Katz, J. S. (1994). Geographical Proximity and Scientific. *Scientometrics*, 31(1), 31–43.
- Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, 24(5), 1316–1336. <https://doi.org/10.1287/orsc.1120.0804>
- Kossinets, G., & Watts, D. J. (2009). Origins of Homophily in an Evolving Social Network. *American Journal of Sociology*, 115(2), 405–450. <https://doi.org/10.1086/599247>
- Krackhardt, D., Nohria, N., & Eccles, B. (2003). The strength of strong ties. In *Networks in the knowledge economy* (pp. 216–239). Retrieved from <https://www.jstor.org/stable/202051?origin=crossref>
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5), 673–702. <https://doi.org/10.1177/0306312705052359>
- Marsden, P. V. (1990). Network Data and Measurement. *Annual Review of Sociology*, 16(1), 435–463. <https://doi.org/10.1146/annurev.so.16.080190.002251>
- McGee, R. (1960). The Function of Institutional Inbreeding. *American Journal of Sociology*, 65(5), 483–488.
- McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review*. <https://doi.org/10.2307/2095356>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Mena-Chalco, J. P., & Cesar Junior, R. M. (2009). ScriptLattes: An open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, 15(4), 31–39. <https://doi.org/10.1590/s0104-65002009000400004>
- Michelfelder, I., & Kratzer, J. (2013). Why and How Combining Strong and Weak Ties within a Single Interorganizational R&D Collaboration Outperforms Other Collaboration Structures. *Journal of Product Innovation Management*, 30(6), 1159–1177. <https://doi.org/10.1111/jpim.12052>
- Newman, M. E. J. (2000). The structure of scientific collaboration networks. <https://doi.org/10.1073/pnas.98.2.404>
- Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 68(3), 8. <https://doi.org/10.1103/PhysRevE.68.036122>
- Perry, B. L., Pescosolido, B. A., & Borgatti, S. P. (2018). *Egocentric Network Analysis: Foundations, Methods, and Models*. Retrieved from <https://books.google.com/books?id=tjRNDwAAQBAJ>
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1), 85–112. <https://doi.org/10.1016/j.jsp.2009.09.002>
- Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata-Volume I: Continuous Responses*. New York: Stata Press.
- Rawlings, C. M., McFarland, D. A., Dahlander, L., & Wang, D. (2015). Streams of Thought: Knowledge Flows and Intellectual Cohesion in a Multidisciplinary Era. *Social Forces*, 93(4), 1687–1722. <https://doi.org/10.1093/sf/sov004>
- Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review*. <https://doi.org/10.2307/1519766>
- Smith, S., McFarland, D. A., Tubergen, F. V., & Maas, I. (2016). Ethnic composition and friendship segregation: Differential effects for adolescent inbred s and immigrants. *American Journal of Sociology*, 121(4), 1223–1272. <https://doi.org/10.1086/684032>
- Smyth, R., & Mishra, V. (2014). Academic inbreeding and research productivity and impact in Australian law schools. *Scientometrics*, 98(1), 583–618. <https://doi.org/10.1007/s11192-013-1052-2>
- Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publishers.
- Snijders, T., Spree, M., & Zwaagstra, R. (1995). The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *Journal of Quantitative Anthropology*, Vol. 5, pp. 85–105. Retrieved from <http://stat.gamma.rug.nl/JQA.pdf>

- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations, and boundary objects: {A}mateurs and professionals in {B}erkeley's {M}useum of {V}ertebrate {Z}oology, 1907–39. *Social Studies of Science*, 19(3), 387–420.
- USP. (2018). *Anuário Estatístico USP*.
- Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative Science Quarterly*, 42(1), 35. <https://doi.org/10.2307/2393808>
- Wells, R. A., Hassler, N., & Sellinger, E. (1979). Inbreeding in social work education: An empirical examination. *Journal of Education for Social Work*, 15(2), 23–29. <https://doi.org/10.1080/00220612.1979.10671562>
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039. <https://doi.org/10.1126/science.1136099>
- Wyer, J. C., & Conrad, C. F. (1984). Institutional Inbreeding Reexamined. *American Educational Research Journal*, 21(1), 213–225. <https://doi.org/10.3102/00028312021001213>
- 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. <https://doi.org/10.1002/asi.23916>
- Zhou, J., Shin, S. J., Brass, D. J., Choi, J., & Zhang, Z.-X. (2009). Social networks, personal values, and creativity: Evidence for curvilinear and interaction effects. *Journal of Applied Psychology*, 94(6), 1544–1552. <https://doi.org/10.1037/a0016285>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.