

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2022-03-07

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Horta, H., Feng, S. & Santos, J. M. (2022). Homophily in higher education research: A perspective based on co-authorships. *Scientometrics*. 127 (1), 523-543

Further information on publisher's website:

[10.1007/s11192-021-04227-z](https://doi.org/10.1007/s11192-021-04227-z)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Horta, H., Feng, S. & Santos, J. M. (2022). Homophily in higher education research: A perspective based on co-authorships. *Scientometrics*. 127 (1), 523-543, which has been published in final form at <https://dx.doi.org/10.1007/s11192-021-04227-z>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Homophily in higher education research: A perspective based on co-authorships

Hugo Horta¹, Shihui Feng², and João M. Santos^{3,*}

1 Social Contexts and Policies of Education, Faculty of Education, The University of Hong Kong, Hong Kong SAR, China

2 Human Communication, Development, and Information Sciences, Faculty of Hong Kong, The University of Hong Kong, Hong Kong SAR, China

3 Centro de Investigação e Estudos de Sociologia (CIES-Iscte), Iscte—Instituto Universitário de Lisboa, Portugal

* Corresponding author: joao_marques_santos@iscte-iul.pt

Abstract

Research collaborations are the norm in science today, and are usually evaluated using co-authorships as the unit of analysis. Research collaborations have been typically analyzed using a mapping perspective that focuses on countries, institutions, or individuals, or by assessments of the determinants of research collaboration, i.e., who engages in collaborations and who collaborates the most. One analytical perspective that has been used less frequently is the homophily perspective, which attempts to understand the likelihood of research collaborations based on the similarity of collaborators' preferences and attributes. In addition, compared to studies focused on the fields of the natural and exact sciences, engineering, and the health sciences, research collaborations in the social sciences have been underexamined in the literature, despite the growing numbers of social scientists who engage in such collaborations. This study assessed homophily with respect to geographical, ascribed, acquired and career-related attributes in co-authorships in the social sciences, based on a co-authorship matrix of 913 higher education researchers. The findings showed that geographic and institutional attributes were by far the most powerful homophilic drivers of collaborations, suggesting the importance of physical proximity, national incentives, and shared culture, language, and identity. Another driver was the similarity of acquired attributes, particularly certain preferences regarding research agendas; these absorbed the residual explanatory power that ascribed attributes such as gender or age had in co-authorship preferences. The study is novel in its analysis of the extent to which similarities in the research agendas of researchers predicted co-authorship. The findings indicate the need for further co-authorship homophily analyses around a broader set of acquired attributes and the trajectories that lead to them.

Keywords: Homophily; Research collaborations; Co-authorship; Higher education research; Social sciences.

Introduction

Collaboration has become the norm in science research. Research collaborations have always been an essential part of scholarly work in the natural and exact sciences, the medical sciences, and technology and engineering, because most of the research in these fields is centered around the social cosmos of the laboratory, where senior researchers, junior researchers, and students engage in team-based projects; collaboration is also necessary in these fields for researchers to gain access to instrumentation and equipment that only a few research groups possess (Lauto & Valentin, 2013). Research collaborations in the social sciences (and in the humanities, to a lesser extent) are relatively more recent as a *habitus* of these fields (Henriksen, 2018). Increased collaboration in the social sciences has been driven largely by the global evolutive dynamics of science and higher education, increasing levels of competition and performativity, and the incentives and rewards offered by national research funding agencies for co-authored work (Xu, 2020). It can be argued that in most countries, social scientists are still in the process of adapting to a more international and collaborative working environment; constraints such as those related to English language ability remain significant, but an increasing number of social scientists are publishing and collaborating internationally (Yemini, 2021).

Past quantitative studies about research collaborations have tended to map them individually, institutionally, or geographically (e.g., Belli et al., 2020); they have assessed factors or specific characteristics that influence researchers' propensity to collaborate, and the attributes, incentives and environments that lead some researchers to collaborate more than others (Jeong et al., 2011; Moody, 2004). Many past studies focused on specific groups of researchers, such as Nobel prize winners or researchers from specific countries (e.g., Kwiek, 2020). These studies focused on the determinants of research collaborations and were useful for discerning the extent to which individual attributes were relevant in determining the propensity, type, and number of collaborations, whether these determinants were ascribed (e.g., gender; Kwiek & Roszka, 2020), or the result of knowledge accumulation and educational and professional experiences (e.g., doing a PhD abroad; Edwan, 2019). Past research also investigated the organizational determinants of research collaborations, such as the characteristics of the organization and the working environment (e.g., reward system; Kim & Bak, 2017), and the relationship between research collaboration and productivity (Abramo et al., 2017).

Comparatively, an approach based on homophily theory has been neglected in the study of research collaborations. Homophily is a term introduced in the literature by Lazarsfeld and Merton (1957) and is defined as individuals' tendency to engage or be associated with others based on similar attributes or specific shared values. This preference does not mean that individuals do not engage in social activities with others who are different from them, and there are cases of individuals who prefer to engage in activities with those who are different from themselves (a process termed as heterophily). Despite the expectation of heterophily in interdisciplinary research collaborations, Feng and Kirkley (2020) found that homophily remained strong, with researchers preferring to work with others who had the same interdisciplinary research experience. Homophily relates to preferences: individuals may prefer to develop activities with others like themselves, localizing quality within specific sociodemographic and geographical spaces, and reinforcing specific positions, preferences, and attitudes (McPherson et al., 2001). This resulting reinforcement of preferences sometimes leads to the creation of exclusive cliques, and this explains why homophily theory has been used frequently as an analytical concept in gender and race studies (e.g., Wang et al., 2019). Homophily theory has also been used in studies that focus on research collaborations, and mostly centered on ascribed attributes or geographical indicators (e.g., Ma et al., 2020; Medina, 2018; Zhang et al., 2018). Our study contributes to the advancement of knowledge concerning homophily in science by focusing on a field

of the social sciences (higher education research), by analyzing in combination ascribed attributes, career resources, and geographical and institutional proximity, and by assessing acquired attributes that the literature so far has not explored. The study has two main research questions: 1) What homophily preferences determine research collaborations among higher education researchers? and 2) What are the most powerful homophily indicators in research collaborations in higher education research?

To respond to these questions, a statistical analysis of the co-authorships of 913 higher education researchers was undertaken based on their career publications, associating the co-authorship matrix with researchers' ascribed and acquired attributes, and their research preferences. Like most studies on the topic, this study defined research collaboration by co-authorship. With co-authorships, the fact of publication indicates the success of the collaborative endeavor and makes it an easy focus of study for analytical purposes. This explains its frequent use in scientometric and bibliometric analyses, whereas collaborations of a more informal nature, despite their relevance, are not analyzed as often (see Laudel, 2002)¹.

As a field of the social sciences, higher education research provides an interesting case to analyze. The field has seen fast growth in the numbers of internationally peer-reviewed publications and collaborations (Jung & Horta, 2013). The number of emergent journals in the field has been growing, and many long-standing ones have come to be recognized as core journals of the social sciences (Tight, 2018). The journals in the field have also become more diverse in content, receiving contributions from many geographical areas, in contrast to the previous dominance of authors based in North America and the United Kingdom (Kwiek, 2021). This increase in global diversity is likely related to the way that human capital and endogenous growth theories have become central in an accelerated, global, uncertain, and technology-dependent world. As higher education systems have become massified and academic research has become valued for social and economic development, themes such as the governance of higher education systems and institutions, supranational and national policies, the evaluation of scholarly work, learning experiences and assessments, and evolving pedagogies have become central to the work of an increasing number of researchers, practitioners, and policymakers. This has led the field of higher education to be characterized as a concurrently broad and specialized field in the social sciences (Daenekindt & Huisman, 2020). In this context, it is unsurprising that the field is participated in by researchers from all disciplines of the social sciences, including sociology, anthropology, political science, economy, management, education, and even occasionally science, technology, engineering, and mathematics (STEM) and the humanities (Santos & Horta, 2018). This diversity, and the fact that both practitioners and academics (with recent greater participation from the latter) engage in the field, makes it strongly multidisciplinary and applied. These characteristics may be critically important to better prepare researchers in this field to study the challenges pertaining to the role of higher education in society, particularly amid crisis situations that demand societal resilience and change (Coates et al., 2021), and to foster the importance of the social sciences amid advancements in science and technology (Wooley et al., 2015).

¹ Co-authorship of a scientific publication is a visible form of research collaboration, but does not represent the entire spectrum of research collaborations (Laudel, 2002). This spectrum is broad and involves diverse forms of collaboration, including informal and casual contributions to the research process. Co-authorship is usually based on the authors who contributed the most to a research study, but in some cases it may include honorary and ghost co-authors (Kumar, 2018). The order of authors is supposed to identify those who contributed the most to the research process, but it is often contextual to disciplinary and sub-disciplinary fields (Marusic et al., 2011). Specifically, some contributors making similar contributions to a research project can be co-authors in some disciplinary and sub-disciplinary fields, but not in others (Whetstone & Moulaison-Sandy, 2020).

The remainder of this article is organized as follows. In the next section, a brief overview of the concept of homophily is presented, including a review of studies that have explored homophily in science. Then in the methods section, the data and the analytical methodology are presented. The results are included in the section after that, and the main insights are presented and discussed in the conclusion.

Literature review

Homophily is at the core of human relationships because it relates to people's sense of belonging. Humans tend to cluster into cultural, social, and economic tribes that share similar characteristics and interests, and homophily helps structure the often multiple social systems in which people participate (Lawrence & Shah, 2020). Conceptually, homophily is related to the willingness to interact or actual interaction with others who have similar characteristics, and also to the formation of group identities based on the association with others who share similar attributes (Dahlander & McFarland, 2013; Currarini & Mengel, 2016). In their seminal article on homophily, Lazarsfeld and Merton (1957) considered two main types of attributes. Ascribed attributes are those that individuals possess, such as gender, age, or race, whereas acquired attributes are those that an individual accumulates throughout life, resulting from educational, social and professional experiences. The latter shape an individuals' values and ideals concerning specific events, social circumstances, work ethic, and job performance. Similarity in ascribed and acquired attributes does not necessarily lead to homophily in both. For example, Alstott et al. (2014) found that when individuals sought to engage others' participation to achieve a specific goal through their own efforts, acquired traits were homophilic, whereas ascribed traits were heterophilic. Another study (McPherson et al., 2001), however, showed opposite findings, suggesting that the context and nature of the activity may influence the extent to which attributes have homophilic or heterophilic effects.

In the context of the present study, the type of association between individuals matters. In co-authorships, the associations tend to be mostly instrumental in nature, because they are organized around specific tasks and activities. Different aspects of the research work may be accomplished by different collaborators because of their individual expertise: for example, one may be an expert in theories, whereas another may be an expert in methods. This dynamic suggests the importance of knowledge specialization, but also of having shared values and perspectives about research thinking and research work. It is likely that researchers who have similar acquired attributes collaborate with each other, whereas researchers with different acquired attributes do not. For example, researchers who are vying for a global scholarly reputation are likely to publish in international peer-reviewed journals, whereas others who are not interested in that, most likely do not (Kwiek, 2020). In the same way, researchers who have strategic research agendas oriented towards multidisciplinary, discovery, collaboration, and the expansion of research into other fields of knowledge may want to collaborate with similar researchers, whereas researchers with research agendas focused on knowledge mastery, specialization, and disciplinary orientation may prefer to collaborate with others who share similar research agendas (Santos & Horta, 2018). Considering that research collaborations are relationships that sometimes combine professional and social dimensions (i.e., friendship), it is likely that personality homophily also matters. Most researchers in academia have the autonomy to engage in research with whom they want, and given the choice, people usually tend to relate with and work with others who have similar personalities (Oh & Kilduff, 2008; Melin 2000). Past research on social communication found that extraverted, agreeable, and open people were more likely to engage with similar others than with those of other personality types, whereas people who had neurotic and conscientious personality traits were less likely to do so (Noë et al., 2016; Solomon et al., 2019; Balmaceda et al, 2013).

Ascribed attributes may also be homophilic in co-authorship. Men and women have been found to have different work strategies and preferences in collaborative research work (Santos et al., 2021; Kwiek & Roska, 2021; González Ramos et al., 2015). Specifically, men are more likely to prefer collaborating with other men, and women generally prefer to collaborate with other women when engaging in collaborative research (Zhang et al., 2018; Abramo et al., 2013; Boschini & Sjögren, 2007). In this sense, co-authorships may also assume both identity associations, which involve self-perceptions of group membership, and knowing associations, which involve the awareness and knowledge of others and by others. This means that the likelihood of collaborations between researchers might not necessarily be founded on an instrumental association, such as the skills each has in relation to the development of the research project. Rather, collaborations may be more likely to occur between researchers who are known to each other, or because both researchers assume a common field of knowledge identity. Indeed, sociologists have been found to be likely to collaborate with other sociologists because they know each other and identify themselves as sociologists, have similar views of the world and their field, and share similar knowledge of sociological theories and methods (Hunter & Leahey, 2008). With reference to the present study, co-authorships in higher education are also likely to be informed by identity associations in the sense that researchers publish on higher education themes in higher education journals. The field of higher education research has gained increasing legitimacy through regional and national organizations, and the group identity of higher education researchers is strengthening (Teichler, 2013). Therefore, one can assume that all of the observed co-authorships in this study represent identity associations, because they pertain to membership of a particular field of knowledge (Lawrence & Shah, 2020), which represents the context in which these co-authorships took place.

A collaboration may also occur because researchers share a societal and/or an institutional identity. For example, academically inbred researchers tend to collaborate mostly with those at their own university because they share and value the institutional identity of their *alma mater* (Tavares et al., 2021). In relation to this, geographical proximity may also be critical: it is likely that researchers collaborate with scholars that they know and may be close to. Past studies have found that researchers tend to collaborate with those working at the same institution as them, but also that when they collaborate with others outside their institution, those working in close geographical proximity are privileged as potential collaborators (Ma et al., 2020). This suggests that geographical attributes may represent an important homophily factor, because they relate to geographical proximity (the same institution, city, or country), making collaboration more convenient and with potentially fewer transaction costs (Evans et al., 2011). Thus, although developments in information and communication technologies may have facilitated some international collaborations, they may not yet have fostered them to a large degree, at least in some fields of the social sciences, including higher education research. Kosmutzky and Krucken (2014) found that international comparative higher education research was in a relatively stable state, indicating that despite the increase in research collaborations, including between authors from different countries, the number of studies that compared two countries remained relatively stagnant. Geographical attributes may also be related to specific values, norms, and taken-for-granted attitudes and behaviors that are rooted in local and national identities, leading researchers from a given city and country to prefer to collaborate with others located there. In this regard, Shahjahan and Kezar (2013) argued that despite occasional efforts to foster holistic global perspectives within higher education research, most of the dominant perspectives are methodologically nationalistic, shaped by a view that the conditions in the researcher's nation-state are equivalent to those in societies in general. In relation to this, higher education research (as in most of the fields in the social sciences) may be characterized by a geographical compartmentalization of

theories, methods, and understandings about the social phenomena under analysis (Tight, 2014), emphasizing the relevance of geographical attributes to research collaboration homophily.

Finally, positional goods (e.g., prestige and access to resources) are important in research, and it is likely that they are also relevant to homophily-related collaborative dynamics. It was found in one previous study that when researchers looked for collaborators, they preferred to collaborate with researchers with the same scientific standing (Evans et al., 2011). However, this tendency may no longer be as strong as it was, for two reasons. First, the complexity of the social and technological phenomena that researchers investigate requires increasingly complicated theoretical and methodological approaches, possibly encouraging more utilitarian associations in research collaborations rather than associations based on prestige and reputation (Feng & Kirkley, 2020). Second, co-authorships with postdoctoral fellows and Ph.D. students are increasing in number because of the evident benefits of such collaborations to universities, scientific fields, and the career development of both mentors and students (Horta & Santos, 2016; Ahmed et al., 2015; Pinheiro et al., 2014; Larivière, 2012; Black & Stephan, 2010). Attributes related to access to career resources are also possibly characterized by mixed homophilic dynamics. It is possible that researchers who obtain research funding are more likely to collaborate with those who do not have access to these resources in a process of heterophily; the latter's willingness to engage in collaborations may be driven by their own lack of access to resources, whereas the former's willingness to allocate funds to the latter may be driven by the desire to meet the research goals that condition the funding grants (Bammer, 2008). Finally, researchers with lighter teaching loads are more likely to engage in collaborations because of the greater amount of time they can allocate to research compared to those with heavier teaching loads (Muriithi et al., 2018; Bozeman & Gaughan, 2011). Teaching is known to constrain research productivity, and time allocated to research is essential for collaborations to take place, leading to an expectation of a homophilic trend concerning the allocation of time to teaching and collaboration (Kwiek, 2018; 2016; Postiglione & Jisun, 2013).

Overall, ascribed attributes (such as gender and age), acquired attributes (such as personality and strategic research agenda preferences), geographical and cultural attributes (such as proximity and the commonality of the home institution, city, or country), and career prestige and resource attributes (such as the number of career publications, citation proclivity, access to research funding, and task allocation) are all bound to have an effect on research collaboration homophily.

Method

Participants

The sample was collected in two stages. We began the first stage by identifying all corresponding authors of articles in Scopus-indexed higher education journals that were published between the years 2004 and 2014². The study was limited to corresponding authors, as only their email addresses are associated with their publications, and given their status, they are likely to be contacted about their work. We identified 6086 authors with this method. These corresponding authors were invited to participate in an online survey, which was administered between May and November 2015. Invitations were sent out in seven waves approximately one month apart. Among all of the email addresses, 643 were inactive and 168 had opted out of receiving online surveys, leaving 5275 valid email addresses.

² The Boolean search string used was as follows: "(SRCTITLE ("higher education") OR SRCTITLE ("tertiary education")) AND DOCTYPE (ar) AND PUBYEAR >2003 AND PUBYEAR <2015." This step yielded 40 journals related to higher education, two of which were excluded: "Chronicle of Higher Education" because of characteristics that distinguish its articles from other journals (see Horta, 2017) and "Art Design Communication In Higher Education," because the journal published only two articles during the period of interest.

The survey contained questions relating to sociodemographic data, the Multi-Dimensional Research Agendas Inventory (MDRAI) (Horta & Santos, 2016), and the 10-item version of the Big-Five Inventory (BFI-10) (Rammstedt & John, 2007). Of the 5275 invitations sent, 1348 were accepted. Of these, 73 were duplicate entries likely caused by participants opening the link at different times and from different computers. After removing all duplicate data, we had 1275 valid participants, representing a response rate of 24.17%, which is excellent for online surveys (Han et al., 2019). Of these participants, 10 were excluded because they did not have a valid Scopus Author ID, which made the subsequent analysis impossible. In addition, 362 participants were excluded from the analysis for failing to complete the survey at any point up to and including the end of the MDRAI block, leaving large portions of the survey missing. A possible explanation for this occurrence was survey fatigue, as the survey was rather long. To ascertain whether the excluded participants differed from the non-excluded participants, we conducted a series of comparisons on their demographic variables. We found that the subgroups did not differ in terms of age ($t(1207) = -0.546, p = 0.585$), gender ($\chi^2(1) = 1.286, p = 0.284$), field of science (FOS) ($\chi^2(5) = 9.282, p = 0.103$), or country ($\chi^2(20) = 12.112, p = 0.912$), suggesting that there were no distinct differences between excluded and non-excluded participants.

This filtering resulted in a working sample of 913 participants. Of these, 488 (53.5%) were females and 425 (46.5%) were males; the participants' ages were between 24 and 84 years ($M = 50.96, SD = 11.17$). About a quarter of the participants were from the United States ($N = 122; 24.8%$); the next two most represented countries were Australia ($N = 142; 15.6%$), and the United Kingdom ($N = 122; 13.4%$). The remaining participants were distributed across a variety of other countries and jurisdictions, and were overall in alignment with the expected geographical distribution of higher education researchers (Kuzhabekova et al., 2015).

The second stage was conducted in 2020. Bibliometric data was collected from the working sample of participants; these data included the country, university, and city information of the participants, their *h*-index numbers, and the list of papers they had published up to the end of 2019. The lag between survey collection (2015) and bibliometric collection was intended to account for possible recency effects between their previously stated agendas and their immediate work, as we had no way of knowing whether agendas reflected only current, or historical preferences. The bibliometric data was extracted from Scopus, which has been reported to have a good coverage of social sciences journals (Norris & Oppenheim, 2007). The co-authorship matrix between the 913 authors was extracted based on the co-authorship information in the historical publications of the participants. The original dataset contained a link to each participants' Scopus ID, ensuring the reliability of bibliometric extraction and matching. This removed the need for author disambiguation.

Variables

In this section, we describe the base variables for each participant that were available to us. Note that these were not used as-is; in the section further below about data processing we describe how these were transformed into measures of dissimilarity to permit our intended analysis.

In terms of ascribed attributes, data on *age* and *gender* were included. Regarding geographical attributes, data on the participant's *country*, *university* and *city* were also included. In terms of career prestige and resource attributes, *h*-index was used, as well as the *percentage of career with research funding* and *percentage of career teaching*. *Percentage of career with research funding* referred to the share of the participant's career in which he or she had access to and benefited from research funding.

Percentage of career teaching indicated the share of time spent on teaching duties after the researcher had concluded the Ph.D.

Finally, two sets of variables were related to acquired attributes. First, one set of variables related to the researcher’s personality, following the well-established Big Five framework as measured by the BFI-10 inventory (Rammstedt & John, 2007). The personality types in this framework are *extraversion*, a measure of how outgoing the individual is; *agreeableness*, the individual’s propensity for cooperation; *conscientiousness*, how meticulous and organized one is; *neuroticism*, which reflects the degree of emotional stability; and *openness to experience*, which refers to a preference for doing new things and experiences. The second set of variables relative to acquired attributes was measured using constructs in the MDRAI (Horta & Santos, 2016). These dimensions are *discovery*, the preference for innovative and breakthrough research topics; *branching out*, the preference for working in multiple topics and fields; *multidisciplinarity*, the preference for work of a multidisciplinary nature; *mastery*, the preference for attaining mastery in a single field of knowledge; *stability*, the preference for stable endeavors and avoidance of shifting interests; *tolerance for low funding*, a measure of risk tolerance regarding conducting research in topics with limited funding; *prestige*, the desire to be recognized among one’s academic peers; *drive to publish*, representing the motivation and willingness to publish academic articles; *willingness to collaborate*, indicating the degree to which the individual is willing to participate in collaborative ventures; *opportunity to collaborate*, the perceived amount of opportunities the researcher has to effectively collaborate; *mentor influence*, measuring the degree of influence that the individual’s mentor (Ph.D. or otherwise) has over his or her work; and *conservative*, which is a preference for doing research in stable fields and topics. The information mapping the sources of data collection to the studied variables is provided in Table 1 below:

Table 1. Data collection and research variables

Data collection source		Variables	
Stage I	Online survey	Ascribed attributes	Gender, age
		Acquired attributes	Personality (Big-Five Inventory 10 item version [BFI-10], Rammstedt & John, 2007)
			Research agenda (Multi-Dimensional Research Agendas Inventory [MDRAI], Horta & Santos, 2016)
		Career prestige and resource attributes	Percentage of career with research funding Percentage of career teaching
Stage II	Scopus API	Geographic attributes	Country, city, university information
		Career prestige and resource attributes	<i>h</i> -index
		Co-authorship information	Binary co-authorship data

Data processing

As the goal of this analysis was to measure similarity (homophily) and dissimilarity (heterophily), substantial database transformations were required. A matrix representing the co-authorship relationship between the 913 researchers was arranged for each possible participant pair. The redundant lower diagonal of the matrix and the missing cases were removed, resulting in a set of 416,328 entries in total in the co-authorship matrix ($913 \times 912/2$). These 416,328 entries, each representing a pair of authors, formed the unit of analysis/cases in this study. For each pair, and for

each of the aforementioned variables, similarity and dissimilarity measures were computed. For quantitative variables, these measures were computed as the absolute value of the difference, and were denoted in the analysis as “deltas.” For qualitative variables, these measures were computed using dummies that assumed the value of 1 if the responses were identical, and 0 if they were not.

Finally, for the dependent variable, we created a *collaboration* variable that assumed the value of 1 if the members of the pair had co-authored a publication at least once, and 0 if they had not. This yielded 321 collaborating pairs and 416,007 non-collaborating pairs. In the next section, we describe how we handled this skewness in our analysis.

Procedure

The nature of the dependent variable was binary; therefore, the most appropriate analytical option was to employ a logistic regression (Hair et al., 2014). However, the collaboration matrix was quite sparse, with only 321 actual collaborations out of the possible maximum 416,328 pairs, indicating a low density in research collaboration between the 913 participants. This was expected, because although international collaborations are increasing in the fields of the social sciences, co-authorships tend to consist of only about 2 to 3 authors per publication (Kwiek, 2020). The scarcity of co-authorships could have potentially led to a rare event bias in estimation (King & Zeng, 2001). To address this, in lieu of a conventional logistic regression, a penalized likelihood method was used, also known as a Firth regression (Firth, 1993). McFadden’s Pseudo R-squared (Smith & McKenna, 2013) was manually computed for each model, using the following formula:

$$pseudoR^2 = 1 - \frac{lpLL(M_{full})}{lpLL(M_{intercept})}$$

where $lpLL(M_{full})$ is the penalized log likelihood value for the fitted model, and $lpLL(M_{intercept})$ is the penalized log likelihood value for the intercept-only model.

Categorical variables were inserted into the models as fixed factors, where the reference category was “different.” Four models were specified in a hierarchical manner: Model I referred to ascribed attributes; Model II included geographical similarity; Model III included career attributes; and Model IV added acquired attributes. The analysis was conducted in R, using the *logitf* library for model estimation and *ggplot2* for the visualizations.

Results

Model I—Ascribed attributes

The first model included only ascribed attributes, which are gender and age. Being of the same gender increased the odds of collaboration by 29.9% ($B = 0.262$, $p < 0.05$, $OR = 1.299$, 95% $CI = [1.043; 1.622]$); inversely, as the age differences increased, the odds of collaboration decreased ($B = -0.003$, $p < 0.05$, $OR = 0.996$, 95% $CI = [0.989; 1.003]$).

Model II—Geographical attributes

In the second model, we included variables relating to geographical proximity. Being in the same country increased the odds of collaboration nearly five-fold ($B = 1.651$, $p < 0.01$, $OR = 5.224$, 95% $CI = [4.017; 6.761]$). Being in the same university increased the odds of collaboration by roughly three times ($B = 1.082$, $p < 0.01$, $OR = 2.951$, 95% $CI = [1.836; 4.594]$), and being in the same city led to an 18-fold increase in odds ($B = 2.933$, $p < 0.01$, $OR = 18.796$, 95% $CI = [13.363; 25.781]$).

Model III—Career attributes

In the third model, we included career features. In this category, only the differences in percentage of career with research funding seemed to matter; increased asymmetries in funding availability led to increased odds of collaboration ($B = 0.007$, $p < 0.01$, $OR = 1.007$, $95\% CI = [1.003; 1.010]$).

Model IV—Acquired attributes

In this model, acquired attributes were introduced in addition to all of the aforementioned variables. Regarding agenda features, increased differences in *multidisciplinary* ($B = -0.153$, $p < 0.05$, $OR = 0.858$, $95\% CI = [0.749; 0.977]$) and *discovery* ($B = -0.242$, $p < 0.01$, $OR = 0.789$, $95\% CI = [0.676; 0.915]$) led to a decreased likelihood of collaboration. Regarding agenda characteristics that related directly to collaboration, asymmetric scores in *invited to collaborate*, predictably led to reduced odds of collaborating ($B = -0.284$, $p < 0.01$, $OR = 0.755$, $95\% CI = [0.644; 0.882]$). In terms of personality, only agreeableness seemed to matter; specifically, asymmetric levels of agreeableness led to reduced odds of collaboration ($B = -0.303$, $p < 0.01$, $OR = 0.741$, $95\% CI = [0.600; 0.908]$). Gender ($B = 0.150$, $p = 0.215$, $OR = 1.116$, $95\% CI = [0.914; 1.490]$) and age ($B = -0.003$, $p = 0.342$, $OR = 0.996$, $95\% CI = [0.988; 1.004]$) both ceased to be significant in this model. This suggested that the homophily of ascribed attributes were explained by acquired attributes, because the variance that explained the significance in the previous three models was absorbed by the variables pertaining to research agendas and personality.

Figure 1 illustrates the model-predicted probabilities of collaboration for the significant variables in Model IV, whereas Table 2 summarizes all four models.

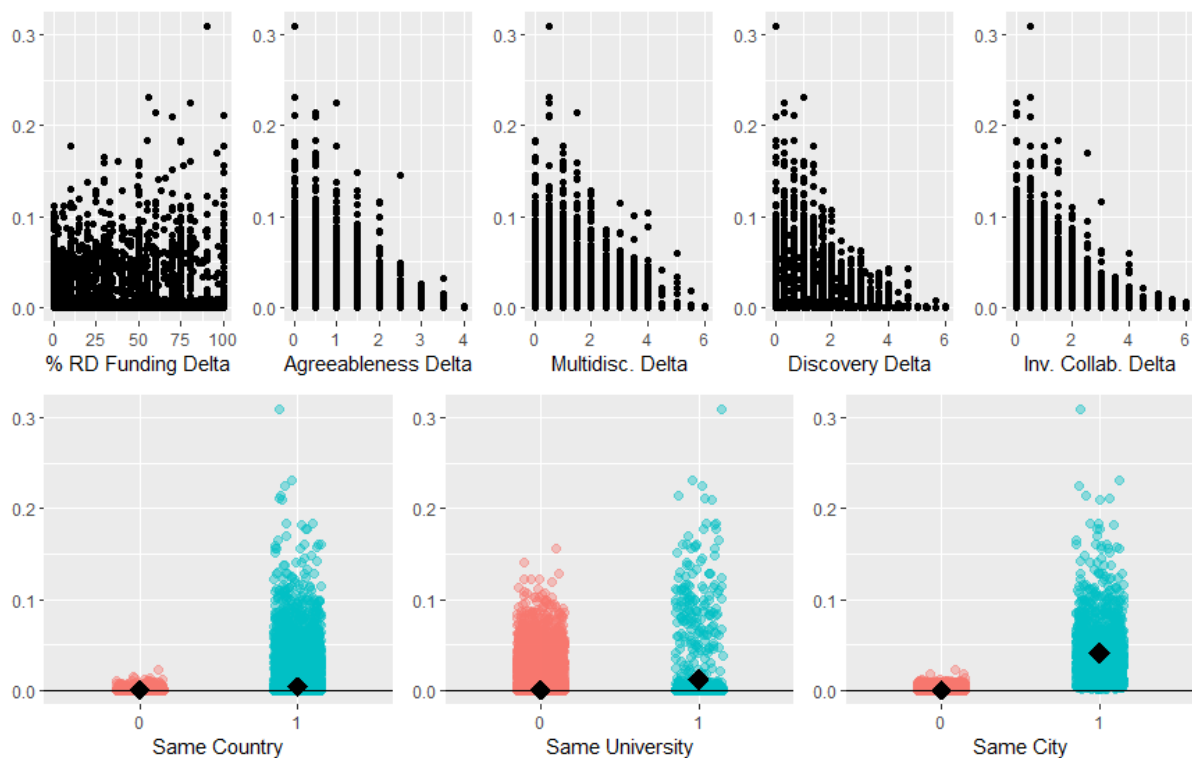


Figure 1. Model-predicted probabilities of collaboration based on the predictor variables. Notes: Only significant effects are shown. Dots indicate predicted probabilities for a given datapoint. The top half

plots non-discrete predictors. The bottom half plots discrete predictors. Diamonds indicate the mean predicted probability.

Table 2. Logistic regression with penalized likelihood

	Variables	Model I	Model II	Model III	Model IV
Ascribed attributes	same gender	0.262**	0.237**	0.234**	0.154
		(0.112)	(0.113)	(0.112)	(0.122)
	Age delta	-0.003**	-0.001*	-0.001*	-0.003
		(0.003)	(0.003)	(0.003)	(0.003)
Geographical attributes	Same country		1.653***	1.672***	1.752***
			(0.132)	(0.132)	(0.142)
	Same university		1.082***	1.035***	0.829***
			(0.232)	(0.234)	(0.271)
	Same city		2.933***	2.969***	2.959***
			(0.161)	(0.162)	(0.173)
Career attributes	<i>h</i> -index delta			-0.013	-0.004
				(0.008)	(0.008)
	Percentage of career with RD funding delta			0.007***	0.007***
				(0.001)	(0.002)
	Percent of career teaching delta			0.002	0.003
				(0.001)	(0.002)
Acquired attributes	Extraversion delta				0.066
					(0.066)
	Agreeableness delta				-0.299***
					(0.103)
	Conscientiousness delta				-0.022
					(0.092)
	Neuroticism delta				-0.081
					(0.075)
	Openness delta				-0.134
					(0.086)
	Prestige delta				-0.001
					(0.069)
	Drive to publish delta				-0.034
					(0.064)
	Mastery delta				0.063
					(0.067)
Stability delta				0.035	
				(0.075)	
Branching out delta				0.088	
				(0.069)	
Multidisciplinarity delta				-0.153**	
				(0.067)	
Discovery delta				-0.237***	
				(0.075)	
Conservative delta				-0.041	
				(0.072)	
Tolerance to low funding delta				-0.071	
				(0.060)	
Mentor influence delta				0.044	
				(0.056)	

Willingness to collaborate delta				-0.061 (0.088)
Invited to collaborate delta				-0.279*** (0.078)
Pseudo R ²	0.001	0.135	0.140	0.157
Observations	416,328	416,328	416,328	335,790

Notes: Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Discussion and conclusion

This study has important findings that contribute to the advancement of knowledge concerning co-authorship homophily in the field of higher education research. The first is that the most powerful attributes influencing homophily were geographical attributes ($\Delta R^2 = 0.134$). Geographical proximity seemed to be a major homophilic drive for co-authorship, in that researchers based in the same institution, city, and country, preferred to collaborate with one another rather than with those located elsewhere.

There can be many explanations for the explanatory power of this attribute, and it is likely that a combination of these rather than a single explanation is the key to understanding the homophilic drive of geographical attributes. One possible explanation is that geographical proximity reduces transaction costs in the research collaboration process; although it is undeniable that international research collaborations are facilitated by low-cost or no-cost information and communication technologies, social scientists still prefer to work with those who are in their national communities and in close physical proximity (González-Brambila & Olivares-Vásquez, 2020). International collaborations are already common practice in STEM fields, and the disparity with the findings of this study likely reflects the continuing adaptation process that higher education researchers are undergoing as they adjust to working on international projects and using online networking platforms (Hoffman et al, 2014). Another possible explanation is the existence of incentives for researchers in the same locality or country to work with one another. Most of the research funding comes from public sources that are of national origin (Chen, 2015). Funding provided by taxpayers is expected to help national communities thrive and improve, and researchers who tap into this funding are expected to use their evolving expertise not only to collaborate but also to compete with researchers elsewhere. Such funding often also comes with restrictions that demand that the funding is mostly spent nationally (Cuntz & Peuckert, 2015). It is also possible that researchers face institutional and peer pressure to participate in, contribute to, and maintain a standing within national associations. Finally, higher education research in particular is strongly influenced by national policies, which may drive researchers to collaborate more nationally rather than internationally (Teichler, 2014).

A further possible explanation for this finding has to do with language. Even though an increasing number of social scientists are publishing internationally, it is possible that their command of the English language (the current *lingua franca* of science) is not at a level at which they can comfortably communicate with others. Researchers find it more convenient to discuss research matters with others in their national languages (Yonezawa, 2015). Another possible related explanation has to do with culture and identity. Sharing a common culture and identity with collaborators can facilitate not just communication, but also the smoothness of social relationships, freeing researchers of the burden of managing intercultural relationships, a task which is not always easy (Wildemeersch & Masschelein, 2018).

Another possible explanation is that higher education researchers (and other social scientists) still tend to focus on mostly national or regional research issues. The nature of research in higher education tends to be highly contextual because the specific characteristics of national higher education systems tend to be unique, localised, and related to national and local cultural, social, and political structures and behaviours. As a result, higher education studies are more likely to focus on the context of specific countries rather than adopt a more universalistic perspective (Kosmützky, 2015; Reale, 2014). As such, higher education researchers may not wish to collaborate with researchers outside their locality or country because, compared with local researchers, outside researchers may be perceived as less likely to understand the local culture and society and the contextuality of the issues being studied, to have similar interests, and to possess the expertise necessary to conduct the focal research. These factors may also explain the stable state of national comparative analyses in higher education research, as mentioned above (Kosmützky & Krucken, 2014).

Finally, researchers in some institutions face organizational pressures to form research teams or groups and collaborate internally. In other institutions, academic inbreeding and a strong institutional identity are prevalent, creating a preference for researchers to collaborate and exchange information mostly with researchers at the same institutions (Tavares et al., 2021).

The second most powerful homophily attributes in this study were the acquired attributes ($\Delta R^2 = 0.017$), but with a much lower explanatory power compared to geographical attributes. The findings pertaining to personality were only statistically significant with regard to agreeableness, but not with regard to other personality traits. This is consistent with the findings of Balmaceda and colleagues (2013) that agreeable people tend to prefer working with similarly agreeable people. The failure in this study to find any other personality trait having a homophily effect in co-authorship preferences can be explained by specificity of the context of research collaborations.

One of the more interesting findings in this study concerns research agendas, an intellectual acquired attribute that had not previously been tested by studies that use the homophily framework. The findings show that only three sub-dimensions of research agendas were relevant to co-authorship preference: multidisciplinary, discovery, and having been invited to collaborate. Researchers tended not to collaborate with each other when one had multidisciplinary preferences and the other preferred single-discipline endeavors. The same held true when one preferred to focus on breakthrough research and the other preferred to explore more established topics that led mostly to incremental research findings. Researchers that received frequent invitations to collaborate in research projects also tended not to co-author publications with researchers who were not frequent recipients of such invitations. From a social network perspective, a possible explanation is that researchers who receive frequent invitations more likely occupy a central place in the network (Biancani and McFarland, 2013), whereas researchers who receive fewer invitations are more peripherally located, reducing the probability of the two crossing academic paths.

Homophily attributes related to career prestige and resources had four times weaker explanatory power than did acquired attributes. Similarities in *h*-index were not statistically significant, suggesting that researchers most likely balanced their co-authorship between junior and senior colleagues, and no preference was identified. Similar levels of involvement in teaching throughout one's career (regardless of whether this was high or low) also did not influence co-authorship, but the percentage of the career in which research funding had been received did lead to heterophily. Researchers who had received funding for greater parts of their career tended to collaborate with those who had not received funding during their careers. This remits to a function of resource dependency in which those with less funding need to collaborate with those with more funding to be able to access resources (the funding drive), whereas those who have funding collaborate with those without funding to benefit

from their expertise and availability (the human resource expertise drive) (Ebadi & Schiffauerova, 2015a). Therefore, it is expected that this collaborative relationship is complementary, in that both parties are bound to benefit, despite the likelihood that the power dynamics of the collaboration are shaped by the resource that is most needed (Ebadi & Schiffauerova, 2015b).

Ascribed attributes mostly had residual explanatory power ($R^2 = 0.001$). The findings show that male researchers preferred to co-publish with other male researchers, and female researchers preferred to collaborate with other females. This homophilic trend is consistent with previous findings, and is one issue that has been identified in relation to the gender gap in science (Wang et al., 2019). The results concerning age also suggest a homophilic trend: researchers of approximately similar ages tended to collaborate more with each other. The homophilic effects of ascribed attributes remained consistent when geographical and career prestige and resource attributes were included in the models. However, when the acquired attributes were included, the effects of the ascribed attributes became non-statistically significant, suggesting that the acquired attributes explained the homophily effects of the ascribed attributes. This indicates that the homophilic effects of the ascribed attributes in higher education research collaborations can be annulled by acquired attributes, implying that the latter are more important than the former.

In sum, this study on co-authorship homophily in higher education research underscores the importance of geographical, cultural, and institutional proximity as the strongest predictors of collaboration. Given that social science research in general tends to be country focused and often uses case study approaches, these findings are to some extent expected. The unexpected finding is the comparably weaker explanatory power of other homophily attributes. Of particular note is the relatively weak explanatory power of ascribed attributes, which were strongly emphasized in many previous studies on research collaborations. Our analysis shows that in the social sciences, acquired attributes take precedence over ascribed attributes; this suggests that more attention needs to be paid to the former. Based on our findings, one can argue that some issues related to known gender gaps in science and academia can be potentially mitigated by policies that address organizational cultures and incentive frameworks that meet the scientific aspirations and intellectual and personal styles of researchers; this is a departure from the usual approach of focusing on ascribed attributes. Future studies on research collaboration homophily should aim to better understand acquired attributes, broaden their scope by including other potential measurements of relevance, and comprehend how they are built upon and related to research career trajectories.

Limitations

Like all research, this study has several limitations that should be acknowledged. First, the target population was restricted to corresponding authors in at least one paper in Scopus. This sample selection criterion left out published authors who were not corresponding authors. As explained above, this was done out of necessity, for two reasons: (1) only corresponding authors have an email address readily available in Scopus, and (2) only these authors are likely to be contacted about their work. Obtaining the contact details of non-corresponding authors would require offsite web scraping. It would also raise ethical issues, as sending invitations to non-corresponding authors could be construed as unsolicited emails. As such, we believe that this limitation was unfortunately unavoidable. Consequently, we acknowledge that the results may not be generalisable to authors who have never been corresponding authors.

Second, our research may suffer from author disambiguation, which is a pressing issue in most bibliometric studies. As noted by Aman (2018), Scopus Author IDs are not truly unique, in the sense that some authors have so-called "split identities," duplicate accounts that contain part of their work,

indexed separately by the algorithm, typically due to the lack of email address. However, although institutional disambiguation is much more challenging, other studies have used author disambiguation based only on Scopus Author IDs (Akbaritabar & Barbato, 2021), which is still reasonable as an author's Scopus Author ID captures on average 97.14% of their publication records (Aman, 2018). Although we verified each Scopus Author ID manually to make sure that a participant matched the Author ID, this did not prevent indexing errors on the Scopus database itself, such as misattributed or missing articles (see De Stefano et al., 2013). As such, the degree of robustness of the co-authorship matrix is solid but not perfect.

Third, there were no data available to allow us to include a control variable for the level of trust when choosing co-authors or engaging in collaborations. Trust is important in research collaborations as a social mechanism that enables repeated collaborations and therefore the establishment of long-lasting collaborative research agendas that many difficult topics require (Bossio et al., 2016; Sargent & Waters, 2004). Indeed, if one collaborator does not trust the other, future collaborations are unlikely to take place. Trust is also important because a research career is based to a large extent on positional goods (i.e., reputation); as a result, a researcher's career can be blemished or tainted by working in collaboration with careless or unethical researchers, such as researchers who engage in fabrication, data manipulation, or plagiarism (Parker & Kingori, 2016; Hanawalt, 2006). Thus, well-established researchers are likely to be cautious in choosing collaborators, not only in terms of expertise but also in terms of research integrity, which means that trustworthiness in choosing collaborators may affect co-authoring decisions. For these reasons, we will include controls for trust in future studies on research collaborations and homophily.

Acknowledgements

This work was partially funded by the HKU Seed Fund for Basic Research for New Staff for new faculty (SF). This study was also partially funded by the Research Grants Council (Hong Kong) through a project entitled 'Characterizing researchers' research agenda-setting: An international perspective across fields of knowledge' (project number: 27608516).

References

- Abramo, G., D'Angelo, A.C., & Murgia, G. (2016) The relationship among research productivity, research collaboration, and their determinants. *Journal of Informetrics* 11: 1016–1030.
- Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics* 7(4), 811-822
- Ahmed, M.Z., Plotkin, D., Qiu, B.-L., & Kawahara, A.Y. (2015) Postdocs in science: a comparison between China and United States. *BioScience* 65(11): 1088-1095.
- Alstott, J., Madnick, S., & Chander Velu (2014) Homophily and the speed of social mobilization: the effect of acquired and ascribed traits. *Plos One* 9(4): doi: 10.1371/journal.pone.0095140.
- Akbaritabar, A., & Barbato, G. (2021). An internationalised Europe and regionally focused Americas: A network analysis of higher education studies. *European Journal of Education*, 56(2), 219-234.
- Aman, V. (2018). Does the Scopus author ID suffice to track scientific international mobility? A case study based on Leibniz laureates. *Scientometrics*, 117(2), 705-720.

- Balmaceda, J.M., Schiaffino, S., & Godoy, D. (2013) How do personality traits affect communication among users in online social networks? *Online Information Review* 38: 136–153.
- Bammer, G. (2008) Enhancing research collaborations: three key management challenges. *Research Policy* 37(5): 875–887.
- Belli, S., Mugnaini, R., Baltà, J., & Abadal, E. (2020) Coronavirus mapping in scientific publications: when science advances rapidly and collectively, is access to this knowledge open to society? *Scientometrics* 124: 1661–2685.
- Biancani, S., & McFarland, D. A. (2013). *Social networks research in higher education*. In Higher education: Handbook of theory and research (pp. 151-215). Springer, Dordrecht.
- Black, G.C. & Stephan, P.E. (2010) The economies of university science and the role of foreign graduate students and postgraduate scholars. In Clotfelter, C.T. (Ed.) *American Universities in a global market*. Chicago: The University of Chicago Press. (pp. 129-162)
- Boschini, A., & Sjögren, A. (2007). Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics* 25(2), 325-365.
- Bossio, D., Loch, B., Schier, M., & Mazzolini, A. (2014) A roadmap for forming successful interdisciplinary education research collaborations: a reflective approach. *Higher Education Research & Development* 33(2): 198-211.
- Bozeman, B. & Gaughan, M. (2011) How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy* 40(10): 1393-1402.
- Chen, S. (2015) Boundary objects and boundary brokering to make the research-policy-practice nexus possible: the case of the Chinese higher education field. *Higher Education Policy* 28(4): 441–457.
- Coates, H., Xie, Z., & Hong, X. (2021) Engaging transformed fundamentals to design global hybrid higher education. *Studies in Higher Education* 46(1): 166–176.
- Cuntz, A. & Peuckert, J. (2015) Openness determinants of national research funding programmes in EU27. *Science and Public Policy* 42(4): 474–486.
- Currarini, S., & Mengel, F. (2016) Identity, homophily and in-group bias. *European Economic Review* 90: 40–55.
- Daenekindt, S. & Huisman, J. (2020) Mapping the scattered field of research on higher education. A correlated topic model of 17,000 articles, 1991-2018. *Higher Education* 80: 571–587.
- 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.
- 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.
- Ebadi, A., & Schiffauerova, A. (2015a) How to receive more funding for your research? Get connected to the right people! *PloS One*, 10(7), e0133061.
- Ebadi, A., & Schiffauerova, A. (2015b) How to become an important player in scientific collaboration networks? *Journal of Informetrics*, 9(4), 809–825.

- Eduan, W. (2019) Influence of study abroad factors on international research collaboration: evidence from higher education academics in sub-Saharan Africa. *Studies in Higher Education* 44(4): 774–785.
- Evans, T.S., Lambiotte, R., & Panzarasa, P. (2011) Community structure and patterns of scientific collaboration in Business and Management. *Scientometrics* 89: 381–396.
- Feng, S., & Kirkley, A. (2020) Mixing patterns in interdisciplinary co-authorship networks at multiple scales. *Scientific Reports* 7731. <https://doi.org/10.1038/s41598-020-64351-3>
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80(1), 27–38.
- González Brambila, C.N., & Olivares-Vázquez, J.L. (2020) Patterns and evolution of publication and co-authorship in social sciences in Mexico. *Scientometrics*. Doi: 10.1007/s11192-020-03644-w
- González Ramos, A. M., F. Fernández Palacín, & M. Muñoz Márquez (2015) Do men and women perform academic work differently? *Tertiary Education and Management* 21 (4): 263–276.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014) *Multivariate Data Analysis*. Pearson Education Limited.
- Han, J., Fang, M., Ye, S., Chen, C., Wan, Q., & Qian, X. (2019) Using decision tree to predict response rates of consumer satisfaction, attitude, and loyalty surveys. *Sustainability* 11(8): 2306. <https://doi.org/10.3390/su11082306>
- Hanawalt, P.C. (2006) Research collaborations: trial, trust, and truth. *Cell* 126(5): 823-825.
- Henriksen, D. (2018) What factors are associated with increasing co-authorship in the social sciences? A case study of Danish Economics and Political Science. *Scientometrics* 114(3): 1395–1421.
- Hoffman, D.M., Blasi, B., Culum, B., Dragsic, Z., Ewen, A., Horta, H., Nokkala, T. & Rios-Aguilar, C. (2014) The methodological illumination of a blind spot: information and communication technology and international research team dynamics in a higher education research program. *Higher Education* 67(4): 473–495.
- Horta, H., & Santos, J. M. (2016) An instrument to measure individuals' research agenda setting: The multi-dimensional research agendas inventory. *Scientometrics* 108(3), 1243–1265.
- Horta, H., & Santos, J.M. (2016) The impact of publishing during PhD studies on career research publication, visibility, and collaborations. *Research in Higher Education* 57(1): 28–50.
- Hunter, L., & Leahey, E. (2008) Collaborative research in sociology: trends and contributing factors. *The American Sociologist* 39(4): 290–306.
- Jeong, S., Choi, J.Y., & Kim, J. (2011) The determinants of research collaboration modes: exploring the effects of research and researcher characteristics on authorship. *Scientometrics* 89(3): 967–983.
- Jung, J., & Horta, H. (2013) Higher education research in Asia: a publication and co-publication analysis. *Higher Education Quarterly* 67(4): 398–419.
- Kim, D.H., & Bak, H.-J. (2017) Incentivizing research collaboration using performance-based reward systems. *Science and Public Policy* 44(2): 186–198.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis* 9(2): 137–163.

Kosmützky, A. (2015) In defence of international comparative studies. On the analytical and exploratory power of the nation state in international comparative higher education research. *European Journal of Higher Education* 5(4): 354-370.

Kosmützky, A., & Krucken, G. (2014) Growth or steady state? A bibliometric focus on international comparative higher education research. *Higher Education* 67: 457–472.

Kumar, S. (2018) Ethical concerns in the rise of co-authorship and its role as a proxy of research collaborations. *Publications* 6(3): 37. <https://doi.org/10.3390/publications6030037>

Kuzhabekova, A., Hendel, D. D., & Chapman, D. W. (2015). Mapping global research on international higher education. *Research in Higher Education* 56(8), 861–882. <https://doi.org/10.1007/s11162-015-9371-1>

Kwiek, M. (2021) The prestige economy of higher education journals: a quantitative approach. *Higher Education* 81(3): 493–519.

Kwiek, M. (2020) Internationalists and locals: international research collaboration in a resource-poor system. *Scientometrics* 124: 57–105.

Kwiek, M. (2018) High research productivity in vertically undifferentiated higher education systems: who are the top performers? *Scientometrics* 115: 415-462.

Kwiek, M. (2016) The European research elite: a cross-national study of highly productive academics in 11 countries. *Higher Education* 71: 379-397.

Kwiek, M., & Roszka, W. (2020) Gender disparities in international research collaboration: a study of 25,000 university professors. *Journal of Economic Surveys*. <https://doi.org/10.1111/joes.12395>

Kwiek, M. & Roszka, W. (2021) Gender-based homophily in research: a large scale study of man-woman collaboration. *Journal of Informetrics* 15(3): 101171.

Larivière, V. (2012) On the shoulders of students? The contribution of PhD students to the advancement of knowledge. *Scientometrics* 90(2): 463-481.

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation* 11(1): 3–15.

Lauto, G., & Valentin, F. (2013) How large scale research facilities connect to global research. *Review of Policy Research* 30(4): 381–408.

Lawrence, B.S., & Shah, N.P. (2020) Homophily: measures and meaning. *Academy of Management Annals* 14(2): <https://doi.org/10.5465/annals.2018.0147>

Lazarsfeld, P.F., & Merton, R.K. (1954) Friendship as a social process: a substantive and methodological analysis. In Berger, M., Abel, T., and Page, C.H. (Eds.) *Freedom and Control in Modern Society* (pp. 18–66). New York: Van Nostrand.

Ma, D., Narayanan, V.K., Liu, C.R., & Fakharizadi, E. (2020) Boundary salience: the interactive effect of organizational status distance and geographical proximity on co-authorship tie formation. *Social Networks* 63: 162–173.

Marusick, A., Bosnjak, L., & Jeroncic, A. (2011) A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *Plos One*: <https://doi.org/10.1371/journal.pone.0023477>

- McPherson, J.M., Smith-Lovin, L., & Cook, K.M. (2001) Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27:415–444.
- Medina, A.M. (2018) Why do ecologists search for co-authorships? Patterns of co-authorship networks in ecology (1977-2016). *Scientometrics* 116: 1853–1865.
- Melin, G. (2000) Pragmatism and self-organization: research collaboration on the individual level. *Research Policy* 29(1): 31-40.
- Muriithi, P., Horner, D., Pemberton, L., & Wao, H. (2018) Factors influencing research collaborations in Kenyan universities. *Research Policy* 47(1): 88-97.
- Noë, N., Whitaker, R. M., & Allen, S. M. (2016, August). Personality homophily and the local network characteristics of Facebook. In *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)* (pp. 386-393). IEEE.
- Norris, M., & Oppenheim, C. (2007) Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics* 1(2): 161–169.
- Oh H, Kilduff M. (2008). The ripple effect of personality on social structure: self-monitoring origins of network brokerage. *Journal of Applied Psychology* 93(5): 1155–1164
- Parker, M. & Kingori, P. (2016) Good and bad research collaborations: researchers' views on science and ethics in global health research. *Plos One*: <https://doi.org/10.1371/journal.pone.0163579>
- Pinheiro, D., Melkers, J., & Youtie, J. (2014) Learning to play the game: student publishing as an indicator of future scholarly success. *Technological Forecasting & Social Change* 81: 56-66.
- Postiglione, G. & Jisun, J. (2013). World class university and Asia's top tier researchers. In Wang, Q., Cheng, Y., & Liu, N. C. (Eds.), *Building world-class universities. Different approaches to a shared goal* Sense: Rotterdam. (pp. 161–180).
- Rammstedt, B., & John, O. P. (2007) Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41(1), 203–212.
- Reale, E. (2014) Challenges in higher education research: the use of quantitative tools in comparative analyses. *Higher Education* 67: 409-422.
- Santos, J. M., & Horta, H. (2018) The research agenda setting of higher education researchers. *Higher Education* 76(4), 649–668
- Santos, J.M., Horta, H., & Amâncio, L. (2021) Research agendas of female and male academics: a new perspective on gender disparities in academia. *Gender and Education* 33(5), 625-643
- Sargent, L.D. & Waters, L.E. (2004) Careers and academic research collaborations: an inductive process framework for understanding successful collaborations. *Journal of Vocational Behavior* 64(2): 308-319.
- Shahjajan, R.A., & Kezar, A.J. (2013) Beyond the “national container”: addressing methodological nationalism in higher education research. *Educational Researcher* 42(1), 20–29.
- Smith, T. J., & McKenna, C. M. (2013). A comparison of logistic regression pseudo R2 indices. *Multiple Linear Regression Viewpoints* 39(2), 17–26.

- Solomon, R. S., Srinivas, P. Y. K. L., Das, A., Gamback, B., & Chakraborty, T. (2019). Understanding the psycho-sociological facets of homophily in social network communities. *IEEE Computational Intelligence Magazine*, 14(2), 28-40
- Tavares, O., Sin, C., Sá, C., Bugla, S., & Amaral, A. (2021) Inbreeding and research collaborations in Portuguese higher education. *Higher Education Quarterly*. <https://doi-org.eproxy.lib.hku.hk/10.1111/hequ.12301>
- Teichler, U. (2014) Opportunities and problems of comparative higher education research: the daily life of research. *Higher Education* 67: 393–408.
- Teichler, U. (2013) Academically ambitious and relevant higher education research: the legacy of the consortium of higher education researchers. *European Journal of Higher Education* 3(3): 242-254.
- Tight, M. (2014) Working in separate silos? What citation patterns reveal about higher education internationally. *Higher Education* 68(3): 379–395.
- Tight, M. (2018) Higher education journals: their characteristics. *Higher Education Research and Development* 37(3): 607–619.
- Wang, Y.S., Lee, C.J., West, J.D., Bergstrom, C.T., & Erosheva, E.A. (2019) Gender-based homophily in collaborations across a heterogenous scholarly landscape. *arXiv:1909.01284v1*
- Whetstone, D. & Moulaison-Sandy, H. (2020) Quantifying authorship: a comparison of authorship rubrics from five disciplines. *Proceedings of the Association for Information Science and Technology* 57(1): e277; <https://doi.org/10.1002/pra2.277>
- Wildemeersch, D., & Masschelein, J. (2018) Lessons from the south: research collaboration as an educational practice. *Social Sciences* 7(11): <https://doi.org/10.3390/socsci7110235>
- Wooley, R., Sánchez-Barrioluengo, M., Turpin, T., & Marceau, J. (2015) Research collaboration in the social sciences: what factors are associated with disciplinary and interdisciplinary collaboration? *Science and Public Policy* 42: 567–582.
- Xu, X. (2020) China “goes out” in a centre-periphery world: incentivizing international publications in the humanities and social sciences. *Higher Education* 80(1): 157–172.
- Yemini, Y. (2021) International research collaboration as perceived by top-performing scholars. *Journal of Studies in International Education* 25(1): 3–18.
- Yonezawa, A. (2015) Connecting higher education research in Japan with the international academic community. *Higher Education Policy* 28: 477–493.
- 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.