*Industrial and Corporate Change*, 2023, **32**, 1077–1108
DOI: https://doi.org/10.1093/icc/dtad043
Advance Access Publication Date: 4 September 2023
**Original Article**

# The role of local colleagues in establishing international scientific collaboration: Social capital in emerging science systems

**Moritz Müller[1], Robin Cowan[1,2,3,*] and Helena Barnard[4]**

[1]BETA, Université de Strasbourg, 61 Ave de la Foret Noire, France. e-mail: mueller@unistra.fr; e-mail: cowan@unistra.fr, [2]CREST, University of Stellenbosch, Krotoa Building, 52 Ryneveld Street, Stellenbosch 7600, South Africa., [3]UNU-MERIT, Maastricht University, Boschstraat 24, Maastricht, Netherlands and [4]GIBS, University of Pretoria, Private Bag x20, Johannesburg 0028, South Africa. e-mail: barnardh@gibs.co.za

*Main author for correspondence.

## Abstract

International collaborations are important for developing science systems. Using a dataset of South African university scientists, we ask whether social capital relevant to international collaboration held by one scientist spills over to local colleagues. Distinguishing between different ways of acquiring foreign ties, we find that 20% of our cases resemble the most-studied form of international collaborations, via the unique ties of an individual with specific characteristics, e.g., foreign research training. In all other cases, both personal and local peers' international social capital is relevant for foreign tie formation. Underlining the systemic functioning of science, international social capital is activated through scientific collaboration among local scientists. The mediating effect of local scientific collaboration is present across all scientific fields and holds for scientists trained locally or abroad. Our findings thus imply that local collaboration is a relevant mechanism to strengthen international collaboration and the formation of international social capital.

**JEL classification:** F63, H52, I2, O15, O20, O30

## 1. Introduction

This paper examines the formation of international research collaborations. The main premise is that social capital plays an important role in the formation of scientific collaborations, whether local or international. Thus, we ask whether social capital relevant to international collaboration held by one scientist spills over to local colleagues. That is, we examine how international collaborations are created, and whether the international social capital of a local scientist can facilitate the formation of international collaborations by his or her local colleagues.

In the modern knowledge economy, global scientific collaboration is central for the development of countries behind the technology frontier. Science and higher education systems in low- and middle-income countries often present handicaps to the scientists in them: human and physical capital resources can be weak, but teaching and administrative demands can be strong (Mazzoleni and Nelson, 2007; Nkomo, 2015; Ismail *et al.*, 2020). These challenges often reinforce a situation where few local[1] scholars are at the cutting edge of science. One common

[1] In what follows, "local" is defined by formal affiliation: an author is "local" if he or she has a formal affiliation to a South African university at the time of a paper being published.

approach to address these handicaps is through international research collaboration (IRC). Participating in global collaborations allows these countries to acquire (Nelson, 2004) and create (Barnard *et al.*, 2012) new knowledge, and global collaborations can be productive even if a science system (e.g. from a low- or middle-income country) is lacking (Ynalvez and Shrum, 2011). This insight has given rise to extensive policy and scholarly work to ensure that scientists in developing countries are connected to the global science system. Much of this work has focused on how to increase the human capital of individual scientists, for example, through PhD training and socialization abroad (Ynalvez and Shrum, 2011; Coey, 2018; Müller *et al.*, 2018). But stays abroad also provide international social capital: returning scientists benefit from retaining links, often in the form of collaboration or co-authorship, with colleagues in the distant location (Turpin *et al.*, 2008; Jonkers and Cruz-Castro, 2013; Gibson and McKenzie, 2014; Netz *et al.*, 2020), and migrant scientists tend to retain links to their home countries (Trippl, 2013). While there is some evidence suggesting that such connections could be at the expense of local social capital (Li and Tang, 2019; Yang *et al.*, 2022), whether, and if so, how the international social capital of a developing country scientist is relevant for local colleagues has received very little attention.

A notable exception is Fry (2023). She studies peer effects from African scientists returning from the United States on non-migrating colleagues in their African home institution. She finds that non-migrating colleagues benefit from returnees in terms of publication output and provides evidence of knowledge spillovers (measured through citations to the US host institution), as well as social capital spillovers (measured through co-authoring with members of the US training institution). Spillovers identified in that study take place by research design at the home institution. We differentiate from Fry (2023) first in the empirical setting: Fry observes returnees doing research in HIV from 15 African countries, whereas our sample is close to the population of sceince, technology, engineering and mathematics (STEM) scientists in South Africa. More importantly though, we do not investigate peer effects on scientific productivity. Our focus is on how international collaborations are created, and whether a local scientist with well-developed international social capital facilitates the formation of the international collaborations of his or her local colleagues.

We examine this by looking at the formation of international co-authorship ties of South African academics from 2005 to 2012. We first differentiate among the ways local scientists can acquire foreign research collaborators. The type best documented in the literature, of an individual developing unique ties to a foreign collaborator, occurs in about a fifth of the cases in our dataset. The bulk of the international scientific collaborations that we observe, though, results either from repeated collaborations with foreigners, from foreign ties acquired jointly by locals, or via referrals from local co-authors or colleagues.

For each type of IRC formation, we attempt to identify causal effects notably of (i) the focal scientist's international social capital, (ii) her local peers' international social capital, and (iii) the joint involvement of local peers in newly created IRCs.

As can be expected, for local scientists engaging in IRC without local collaborators, international social capital from own international experience is paramount. In all other types of IRC formation, that is 80% of IRC we observe, local scientists benefit considerably from the involvement of their local peers — local collaboration mediates the effect of international social capital (held by the focal and her peers) on the acquisition of international collaborators. The estimated effects are comparable for local scientists trained locally or abroad and hold across STEM fields.

We observe that social capital created by the past foreign collaboration works jointly with local social capital to increase the likelihood of subsequent international scientific collaborations. This underlines the importance of understanding the science system as a system, even when it operates behind the technology frontier. Much of the extant work on the internationalization of developing science systems is focused on the effects of global mobility on the mobile scientists as individuals. Our work draws attention to the important role of local collaborations, with important implications for both scholars and policymakers. Furthermore, though our empirical context is South Africa, the processes we describe seem likely to be relevant in most academic milieux.

## 2. Literature review

A challenge for emerging economies with their often small and underresourced science systems is how to access, diffuse, and (hopefully) contribute to frontier knowledge. As these societies and economies seek to "catch up" with the rich, industrialized countries and push their higher education and research toward the knowledge frontier, accessing frontier knowledge through the international scientific community is both extremely important and very difficult. The difficulties include issues of inadequate absorptive capacity (Mazzoleni and Nelson, 2007), limited finance (Nkomo, 2015), and greater geographical distance (Ismail *et al.*, 2020). A common response to these difficulties is to explicitly connect to the global community, either through international mobility or through international collaboration (or both).

It is of course inaccurate to suggest that only countries behind the technology frontier stand to benefit from being connected to the global scholarly community. In recent decades, science has become more collaborative and the incidence of international collaboration has increased. Although the evidence is not entirely unequivocal,[2] the review of Tahamtan *et al.* (2016) provides strong evidence that a greater number of co-authors, and especially international co-authors, increases citations. This effect is likely due both to social network effects (more people discussing their latest work with more colleagues who later cite it) and to increased paper quality. To the extent that the latter is true, international co-authorship is correlated with scientists' (quality corrected) productivity.

These observations have led to the idea of scientific and technical human capital (Bozeman *et al.*, 2001). Bozeman *et al.* argue that science advances through individuals not only employing their accumulated knowledge and skills but also accessing resources through their social networks. Thus, simple human capital (the assemblage of knowledge and skills) is only part of the picture, and contact with other scientists and institutions (the social capital part of a scientists' resources) is also central to knowledge creation. It follows that a researcher with a larger and/or more varied network of contacts likely has more opportunities to source interesting or valuable knowledge (for knowledge spillovers in science networks, see Azoulay *et al.*, 2010, and Mohnen, 2022).

To the extent that scientific and technical human capital is founded on contacts with scientists who can provide access to useful knowledge or resources, the link between scientific and technical human capital and international scientific networks is particularly apposite for developing countries with their typically small science systems (Fry, 2023).

One source of international collaboration, pursued by policymakers in many developing countries, is international mobility. A strategy also followed by South African policymakers: *"[…] The enhanced strategic focus is on building human resources through international mobility and strengthening and growing research infrastructure capabilities. […]"* (Mr. Daan Du Toit, DSI Deputy Director General of International Cooperation and Resources.[3])

In this policy perspective and in most scholarly work on international mobility, the individual scientist is the focus. But given that science is increasingly acknowledged to be a system, international scientific collaboration demands a more systemic analysis. While collaboration is generally understood to be valuable to the individual collaborating scientist, the collaboration may also have benefits for local colleagues not directly participating in it. This understanding of the social dimension of international scientific collaborations is missing in the current literature.

This motivates us to explore the relationship between local and international collaboration. Most empirical studies dealing with internationalization of science focus on how IRC relates to scientific productivity and excellence. These studies consistently report a positive association between local and international collaboration on all levels, i.e. at the national, institutional,

[2] Zhao and Guan (2011) find that co-authorship in nano-biopharmaceuticals increases citations especially when collaboration is international, but the latter only for China and Japan. Didegah and Thelwall (2013) find that international collaboration increases citations in some but not all fields. (Increasing in biology and biochemistry and Chemistry, but not in nanoscience and nanotechnology.) International papers receive fewer citations than local ones in Harvard Law Reviews (Ayres and Vars, 2000). Puuska *et al.* (2014) find that international co-authorship does not garner more citations per author than domestic co-authorship, among Finish papers.

[3] Quote from the committee on Internationalisation of Higher Education: DSI and DHET cooperation on the "2022 Science, Technology and Innovation (STI) Decadal Plan," https://pmg.org.za/committee-meeting/34430/.

research group, and individual levels (Bordons *et al.*, 1996; Martin-Sempere *et al.*, 2002; Kyvik and Reymert, 2017; Scarazzati and Wang, 2019; Sooryamoorthy, 2019; Kwiek, 2020a). Thus, international collaboration does in general not crowd out domestic collaboration. This also holds true for developing economies that rely particularly on IRC to access funding and other resources needed for research that are not available in the home country (Pouris and Ho, 2014). Barnard *et al.* (2012) analyze a dataset on South African scientists similar to the data we use and find that internationally renowned scientists are well connected both internationally and to local scientists through (direct and indirect) co-authorship links. If international collaboration is more beneficial than local collaboration, in particular in resource-poor environments, why then are international and local collaboration strongly positively correlated? Explanations, put forward but never tested, are that scientific excellence and productivity cause collaboration in general (domestic and international), that individuals have a general taste for collaboration (e.g. Kwiek, 2020a; Kwiek, 2020b), or that brokering global and local knowledge flows is valuable (Barnard *et al.*, 2012). Our findings suggest the alternative explanation that the net benefit of IRC increases with local collaboration. The particular mechanism for which we provide empirical evidence is the spillover of international social capital through local collaboration.

We address this question as an issue of link formation, as much scientific knowledge is communicated through various types of discussion and collaboration. Particularly in the context of an emerging economy, it can be very valuable if the discussion involves some participants at the knowledge frontier and some seeking to approach it.

Research on scientific collaboration tends to use co-authorship as an indication of collaboration and knowledge sharing. Recent work has successfully expanded beyond co-authorship to include other kinds of interaction and traced some of their effects on knowledge diffusion (Baruffaldi and Poege, 2020), future formal collaboration, and citation behavior (Lane *et al.*, 2021). Granting that the co-authorship measure has well-known deficiencies (see Scellato *et al.*, 2015, for a discussion), Fafchamps *et al.* (2010) argue that co-authorships still represent a sub-graph of the overall collaboration or interaction network of scientists and so offer a reasonable proxy for tie formation and interaction among researchers.

The general issue of how new links are formed constitutes a major theme in the social network analysis literature. Several factors recur in studies across a variety of domains: returning to old partners; complementarity of resources; homophily or assortativity; referrals by current partners; and generalizing, distance in the existing network.

Several studies have identified factors that support the creation of a co-authorship link between two researchers who had not previously collaborated, for example, the work by Fafchamps *et al.* (2010) on economics co-authorships, Dahlander and McFarland (2013) using data of Stanford University authors, Essers *et al.* (2022) using the International Monetary Fund working paper series, and Rivera-Leon (2021) using the publications of South African STEM scientists. One important explanation for co-authorship is disciplinary similarity: a common set of research interests is likely to result in people not only being concerned with the same problem but also better able to communication nuances regarding that issue. Differences in seniority or productivity often occur, that is, tie formation is dis-assortative. This may be driven by student–supervisor collaborations. Centrality also matters, and scientists with more co-authors are more likely to form new partnerships. This may be capturing an inherent taste of a researcher for collaboration, greater productivity (more papers imply more opportunities for collaboration), or simply the central author's desirability as a partner.

Finally, referrals are often seen as an important source of co-authorships. All four papers find a negative relationship between the probability of forming a co-authorship and distance in the existing co-authorship network. Having a common co-author permits a direct referral. Often these referrals take the form of a new person joining an existing and repeating co-authorship team, but it could also be that a researcher recommends a colleague who could be interested in a particular project. Fafchamps *et al.* (2010) go further, generalizing from indirect ties (distance 2) to network distance, and find effects up to a distance of 11. As they point out, it seems highly unlikely that there is a chain of referrals whereby A finally finds (by being referred by B to C, by C to D…) that scientist K is a good collaborator. Rather, they infer that the co-authorship network is much less dense than the network that defines the invisible college (Crane, 1972) and that these chains of long length in the co-authorship network would actually be much shorter in the colleagues' network.

An emerging question is the extent to which networks are the outcome of strategic intent and action or not (Carayol *et al.*, 2019). Cowan, Jonard and collaborators (Cowan and Jonard, 2003; Cowan *et al.*, 2007; Baum *et al.*, 2010) argue that it need not be. In their models, triadic closure takes place in the course of other actions of agents, that is, as a consequence of joint knowledge creation in a common knowledge space. By contrast, Coleman and others argue that having common partners improves social processes among which is joint knowledge creation (Granovetter, 1985; Coleman, 1988; Gulati, 1995).

Carayol *et al.* (2019) address these alternative explanations in a theoretical network formation model, which they apply to the longitudinal network of French inventors. Agents benefit from research collaboration in two ways: there is an immediate benefit from a joint research project and the aggregated network provides future access to knowledge. Their empirical analysis focusses on the latter, treating the former as a fixed effect. Parameter estimates suggest that network externalities arise from non-redundant ties, whereas transitive triples are explained by proximity of different kinds rather than by deliberate triadic closure. Thus, Carayol *et al.*, consistently with Azoulay *et al.* (2010) and Mohnen (2022), suggest that network value originates from having many diverse connections in the aggregated network (rather than from being embedded in closed triangles). Our results complement these. In our data, some foreign connections are made through referrals by local colleagues, thus involving triadic closure, but this is a relatively small proportion of all foreign connections. The largest group is made by two (or more) South African scientists jointly collaborating with a foreign scientist (who had no previous South African collaborations). Local collaboration seems to be an important means to access foreign collaborators, suggesting that local collaboration may positively affect the immediate benefit from international research projects.

That relates to the focus of this paper: when a researcher has managed to establish ties internationally, does this affect the scientific and technical human capital of his or her local colleagues? We investigate how international collaboration partners are acquired within a national science system. Our focus is on the social interactions of local scientists in the formation of foreign co-author ties. For example, multiple local scientists may jointly acquire and repeat foreign co-author ties or provide referrals of foreign co-authors to each other.

Conceptually this is an important step because we remove an inconsistency in virtually all prior related studies: while the explanandum, international scientific collaboration, is clearly acknowledged to be systemic, the focal agent contributing to this phenomenon, the explanans, remains seen as an isolated individual whose actions have no external effects at all. This runs against the idea of international collaborators being a form of *social* capital, as it ignores not only that capital itself is social but also that the processes through which this form of capital are formed and retained are also social.

In the context of development, such a perspective may reflect the belief that the local science system is dysfunctional to the point that meaningful engagement within the system is unlikely. We take as point of departure that development is systemic and investigate the (larger) social process of how international collaboration is fostered not only by a mechanism like the international mobility of a focal agent but by the social processes among that agent and his or her local peers. Thus, we see the establishment of foreign ties as a process involving not only the focal agent but also his or her peers; and international scientific collaborations as valuable not only because they develop the capital of the focal agent but also as a source of social capital for other agents in the system.

Our empirical analysis is of the science system of South Africa (SA), which is relatively small (about the size of Belgium) and emerging. Post-Apartheid, South Africa officially adopted an internationalization strategy for scientific upgrading as a way of fostering economic development.[4] Our analysis confirms the value of well-known policy measures for internationalization like the hiring of (renowned) foreign scientists and support of foreign stays. A novel finding is that maintaining an inclusive collaboration environment within the local science system is instrumental

[4] http://www.nrf.ac.za/division/irc/rating.

for foreign tie formation. Foreign ties are rarely kept or formed by a local scientist in isolation. Instead, foreign ties seem to flourish in local environments that are in themselves collaborative.

## 3. Data and methods

### 3.1 Sample

The following paragraphs present what we consider the main features of our sample. More detailed discussions on various aspects of the data are provided in Appendix 1.1.

Our focus is on how scientists active in the South African academic community establish co-author links to foreign scientists. The scientists rated by the South African National Research Foundation (NRF) form our sample. The NRF is mandated with the support of research within South Africa, in particular, but not only, within the university system. Since the mid-1980s, one of its activities has been the "rating" of researchers at universities and other public research institutions. The rating system has first been applied in Science, Engineering, and Technology (SET) where it widely diffused within 10 years, reaching saturation in the mid-1990s (NRF, 2005, p.18). From 2002 on, the Social Sciences and Humanities (SSH) were included in the NRF rating system, with a strong uptake within the first 2 years. Ratings are voluntary but encompass the vast majority of academics with a research career (they are estimated to cover roughly 30% of all academics who account for 90% of all South African publications; see Barnard *et al.*, 2012).

The rating process involves an examination by international referees of a researcher's research activities and output, and scholars must (re-)apply roughly every 4 years. Application for a rating demands submission of a complete record of education, work history, supervision, and publication (of all types), along with other demographic data. Scholars are assigned one of the six ratings: A, B, C are the standard categories, and L, Y and P are special categories.[5] NRF ratings involve a rigorous evaluation process and so can be used as an effective measure or indicator of the quality of a researcher's output. There are strong incentives for participation in the process both for the individual researcher and for her institution (Pouris, 2007).

Therefore, the vast majority of (academic) researchers in South Africa do apply for ratings, so the coverage of the dataset is very good if the population in question is research-active scientists. It seems likely however that our dataset covers SET research somewhat better than SSH research, not only due to the relatively late introduction of ratings in SSH but also due to relatively low funding needs in SSH compared to SET. Therefore, we focus in the following on SET scientists, generally referred to as "scientists" in the following.[6]

Our sample is based on digital files of the NRF established in the period from 2002 to 2015. During that period, 2762 SET scientists applied for at least one NRF rating.[7] Some scientists did not receive a valid rating for several reasons—mostly due to ineligibility or incomplete application or because the rating was in process at the time of the data snapshot. A valid rating decision, including an unsuccessful rating, was obtained by 2696 scientists.

Table 1 describes the basic features of our sample of scientists with at least one valid rating decision. Scientists are characterized by basic social variables (constant), scientific field (constant), higher education degrees (constant), employment (dynamic), rating (dynamic), and their scientific output in terms of papers (dynamic).

Co-author networks are measured on the scientific publications entered into the NRF database by South African scientists for various programs run by the NRF, most notably their rating applications. Removing entries with missing data and disambiguating names in the publication data leave 163,591 authors (individuals). Co-authors of NRF-rated scientists are identified as being local or foreign based on (i) additional information in the NRF dataset (employment in SA or corresponding e-mail in SA), (ii) a complementary dataset on publications of South African authors

[5] From the NRF website: A – Leading international researchers; B – Internationally acclaimed researchers; C – Established researchers; P – Prestigious Awards; Y – Promising young researchers; L – Latecomers to Academia.

[6] Descriptive statistics on SSH scientists can be found in the Appendix, and we include regressions on foreign tie formation in SSH in Section 4.3 in our comparison of scientific fields.

[7] We restrict attention in these estimations to the natural sciences, engineering and Mathematics (SET) because of the relative homogeneity of publishing practices there. Publishing practices in social science and humanities (SSH) are varied, and often differ from those in SET.

**Table 1.** Basic description of SET sample

| Variable | Value (min, 25%, 50% 75%, max) | NA's |
|---|---|---|
| *Social characteristics* | | 5 |
| Birth year | [1926, 1954, 1962, 1970, 1984] | 2 |
| Gender | Female 708, Male 1985, | 3 |
| Skin color | Black 406, Coloured 90, Indian 183, White 2014, | 3 |
| Citizenship by region | Africa (exc. SA) 313, Asia 73, Europe 311, North America 55, South Africa 1934, South America 9, | 1 |
| *Scientific career* | | 281 |
| 1st degree year | [1947, 1977, 1985, 1993, 2006] | 124 |
| 1st degree region | Africa (exc. SA) 279, Asia 96, Europe 400, North America 65, South Africa 1714, South America 13, | 129 |
| PhD year | [1949, 1989, 1998, 2004, 2013] | 154 |
| PhD region | Africa (exc. SA) 53, Asia 97, Europe 509, North America 189, South Africa 1688, South America 4, | 156 |
| 1st SA employment year | [1945, 1990, 1999, 2005, 2014] | 226 |
| No. of ratings | [1, 1, 1, 3, 7] | |
| Journal papers | [0, 17, 32, 58, 750] | – |
| Conference papers | [0, 0, 4, 15, 289] | – |
| Rated scientists | 2696 (total) | 283 |

Numerical values are described by the five statistics [min, 1st quartile, median, 3rd quartile, max]
No. ratings within the period (1990, 2005).
Journal papers and conference papers are both peer-reviewed. Other scientific output in terms of working papers, books, patents, etc., are not considered.
Skin color categories follow the racial categorization established during the Apartheid.

built on a national publication bonus system, and (iii) an extraction of South African publications from the Web of Science (making use of the author affiliation field). Overall, we identify 25% of the individuals as South African, 35% as being foreign, and for 40% of the individuals the origin remains ambiguous. Co-author networks, used for the analysis to calculate network variables of NRF-rated scientists, include only individuals with non-ambiguous origin (note that for NRF-rated scientists, the origin is never ambiguous, only for their co-authors appearing in their digital Curriculua Vitae).

Instead of creating one large co-author network of all SA science, we create one co-author network for each scientific field (e.g. physics) by including all rated SA scientists from that field and all their co-authors (local and foreign). This reduces the risk to confuse two different (local and/or foreign) scientists having the same name as being the same person, by avoiding the issue altogether for scientists working in different fields. The drawback is that referrals of foreigners across scientific fields are not traced.

The network is constructed on peer-reviewed journal publications. We exclude conference proceedings to avoid double counting of scientific collaborations from one year to another. Furthermore, each scientist enters her scientific output up to the time of filing for a rating (or a project) application, implying that the same article appears in the dataset twice if two different scientists file the same, co-authored paper.[8] The potential duplication of co-author links prompts us to work with unweighted, binary networks.

With these data at hand, we construct an unbalanced panel spanning the period from 1996 to 2014. The panel traces individual scientists from the first year of South African employment after obtaining a PhD until one year before their last observed rating. Individuals for which this information is missing cannot be included in the panel. In constructing the estimation panel, the panel dimension is further reduced due to missing information and constraints imposed by the estimation method. A detailed discussion is deferred to the Appendix. Descriptive statistics are given in Section 3.4, Estimation Sample.

[8] There is no unique paper identifier such as DOI in the dataset, and titles, journal names, etc., in the data are strings entered by the applicants in various, non-consistent ways.

## 3.2 Variables

### 3.2.1 Mathematical notation

In order to specify how exactly we measure variables, and also in view of the analysis that follows, some mathematical notation is helpful.

Each year $t$, we observe co-author ties between SA scientists $i$ and $j$, $(i \sim j)$, and ties between SA scientists $i$ and foreign scientists $k$, $(i \sim k)$. To characterize social capital, we define three networks: local co-authorship, foreign co-authorship, and faculty. The faculty network, $g^f$, is a bipartite network of scientists and affiliations, projected onto the scientists. The local, South African, co-authorship network, $g^s$, is simply co-authorship among South African scientists. The foreign co-authorship network, $g^a$, is a bi-partite network of South African scientists and foreign scientists.[9] All networks are time specific, subscript $t$ denoting the year in which a paper was published or an affiliation was observed. To calculate co-author social capital, we separately cumulate the two co-authorship networks over the past $\tau$ years into $G^s_t$ and $G^a_t$. Note that we proxy the concept "international social capital" of a scientist by her relations in the "foreign co-authorship network" and therefore refer to this measure in the following as "foreign social capital." In the analysis that follows, we use $\tau = 5$. In the robustness analysis, we vary that time window. A scientist, $i$, has three sources of social capital, which we capture simply as his or her neighborhoods in each of the three cumulated networks $G$: $N^s_{i,t}$; $N^a_{i,t}$; $N^f_{i,t}$.

### 3.2.2 Outcome variables

#### *3.2.2.1 Foreign ties*

Our interest is in the acquisition of foreign collaborators by South African scientists. The number of foreign co-author ties formed by a focal South African scientist $i$ at time $t$ is our dependent variable, denoted by $y_{i,t}$. In our panel data, we observe in total across all years and individuals the formation of 36,866 "foreign ties".

#### *3.2.2.2 Network motifs of foreign tie acquisitions*

We can differentiate several temporal network motifs.[10] Figure 1 illustrates the different network motifs of foreign ties described in the text below. The figure is drawn in such a way to illustrate that different types of ties could emerge in the same project or paper.

The motifs we discuss are dyads and triads because we are particularly interested in tie formation between a focal local scientist $i$ and a foreign scientist $k$ (a dyad) and the potential interaction between a focal scientist $i$ and her local peer $j$ in forming a tie with a foreign scientist $k$ (a triad). For each focal scientist $i$, we count how often she is involved in each motif. The tree diagram in Figure 1 clarifies that outcome variables consist of disjunct motifs, such that the sum over motifs corresponds to the total number of foreign ties formed.

First, consider ties to a foreign scientist $k$ with no prior South African co-author (see right branch "new foreigner" in Figure 1). A new tie between a focal scientist $i$ and the "new foreigner" $k$ could be unique to $i$, that is not involving any other local scientist $j$ (one South African co-authoring with the foreigner—*unique acquisition*), or joint with another local scientist $j$ (several South Africans collaborating as a team—*joint acquisition*). This categorization makes sense if we would like to understand whether foreign contacts new to South Africa tend to be established by local scientists in isolation or jointly.

Now consider ties to a foreigner $k$ who already has South African co-author(s) (see left branch "known foreigner"). A tie between a focal scientist $i$ and a foreigner $k$ could simply be a repeat of a prior tie between $i$ and $k$ (*repeated tie*). We consider this a repeated tie, whatever the involvement of a focal scientist's peer $j$. So what we focus on here is $i$'s deepening of existing foreign relations (the $k$'s). If not repeated, it could be a "referral" from a peer, i.e. there is no prior tie between the focal scientist $i$ and the foreign scientist $k$, but a prior tie between focal's peer $j$ and $k$, where the

[9] Note that in our dataset, we do not observe ties among foreign scientists.

[10] A network motif in general is an isomorphism class defined on a subgraph, the simplest is an undirected edge, the most prominent a closed triangle (see Holland and Leinhardt, 1976). Temporal network motifs are isomorphism classes on temporal networks taking into account time information on edges and potentially nodes. The simplest and most prominent temporal network motif is a repeated tie, that is the introduction of a tie in time $t$ that already has been created at some time before.

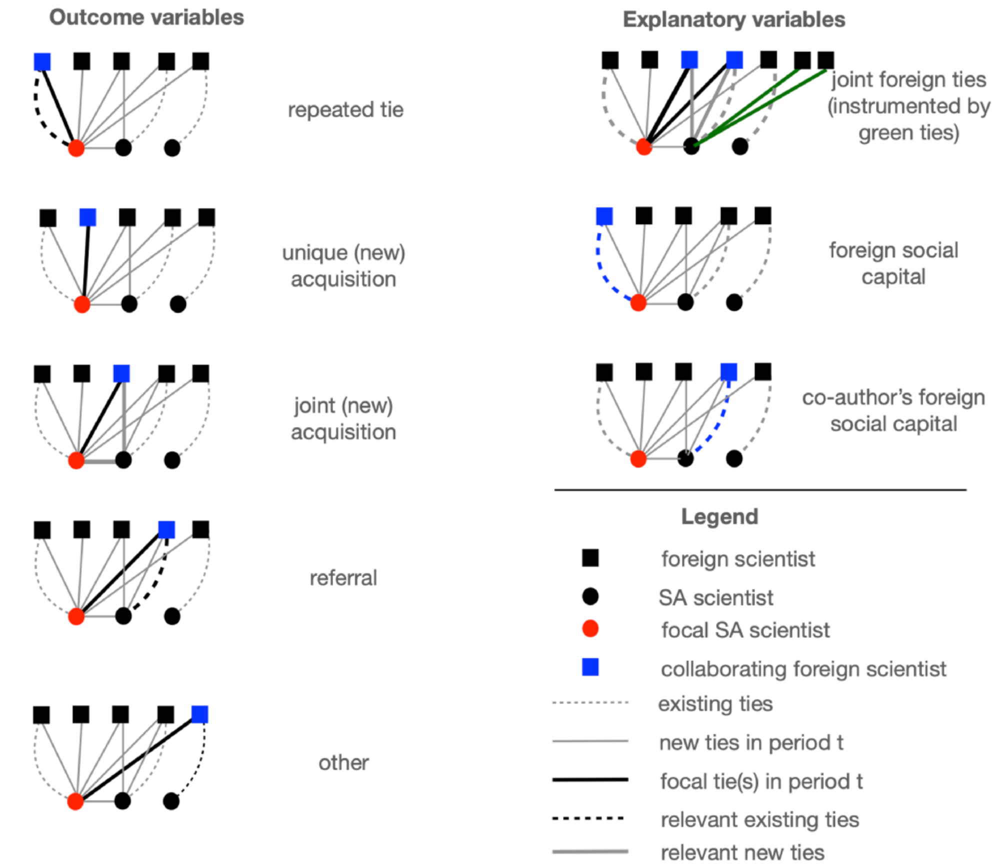


**Figure 1.** Network motifs of foreign tie formation.

peer *j* could be a (past or current) co-author of *i* (*co-author referral*) or a colleague of *i* (*faculty referral*). The "referral" motifs aim to capture the fact that part of a peer's *j* international social capital becomes a new foreign contact of the focal scientist *i*.

Finally, as a remainder category, there may be other configurations such that a focal scientist *i* creates a tie to a foreigner *k* where *k* has a prior tie to another local scientist *j*, but we do not observe *j* to be a peer of *i*, i.e., *j* is not in any of *i*'s neighborhoods as defined above (*Other*).

In the panel data, we count in total 36,866 "foreign tie" acquisitions. Note that the same foreign scientist may form multiple ties with multiple SA co-authors. Around two-thirds of foreign ties are formed with new foreigners (23,207) and one-third are formed with foreigners already having an SA tie (13,659). Looking at the new foreigners entering the system, we find that around one-third are unique to one SA co-author (7402), while two-thirds are jointly acquired by several co-authors (15,805). Turning to tie formation with foreigners that already had a prior tie with a South African scientist, 80% are repeated ties (11,014). Ten percent are created through referrals (1501), and the rest is untraced (1144). This breakdown of tie types can be seen graphically in Figure 2.

At this point, a short reflection on the temporal network motifs just introduced seems helpful. First, note that this categorization is not the only possible one because several categories could be divided into further categories and re-assembled in alternative ways. The aim is not to be exhaustive but to create categories (or network motifs) that make sense from the perspective of a focal local scientist and hence are intuitive for the analyst. In the regressions that follow, the number of foreign ties created in a given category will be the left-hand-side variable, and hence

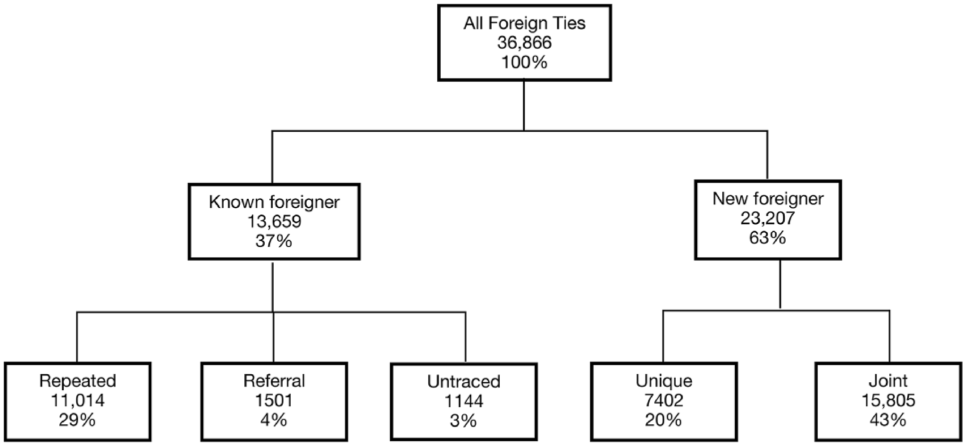


**Figure 2.** Tree diagram showing the distribution of different foreign tie types. The diagram includes all panel observations. The estimation sample includes fewer observations, and hence fewer foreign ties formed, due to missing observations and constraints imposed by panel estimation (detailed in the Appendix), but percentages remain the same.

the right-hand-side variables are factors that potentially contribute to the creation of such types of foreign ties.

Second, the motifs described above correspond always, sometimes, or never to a triangle (or a closed triad) in the observed aggregated network, i.e., the prior network plus current tie formation. Furthermore, the observation of a triangle in the aggregate network does not necessarily imply triadic closure. Triadic closure can be seen as a basic mechanism in the network formation that consists in the higher probability that two nodes separated by a shortest path of length two will eventually connect, relative to those at higher distances (and disproportionately so).[11] The motif *joint acquisition* corresponds to the introduction of a triangle into the prior network but is never triadic closure of a prior network's triple involving two local scientists and one foreign scientist. The reason is that, by our definition, the triad in the prior network has at most one link (one or no link between the local scientists, and no links between a local and the foreign scientist—the foreign scientist has never collaborated with a South African). Similarly referrals (including *co-author referral* and *faculty referral*) will always result in a triangle but do not necessarily imply triadic closure, whereas we require the existence of a prior tie between a peer $j$ and the foreign scientist $k$, we do not require a prior tie between the locals $i$ and $j$. For us, it suffices if $i$ and $j$ form a current tie to recognize them as local peers. *Repeated tie* may or may not involve the introduction of triangles or triadic closure. Finally, *unique acquisition* and the remainder category *other* never involve the introduction of a triangle nor triadic closure.

Third, we do not have reliable observations in our data on co-author ties between foreign scientists and as a consequence ignore these. Therefore, all motifs include only one foreign scientist. Clearly, what we term from the perspective of the South African scientist, ignoring links among foreign scientists, a unique acquisition of a foreign scientist $k$ could well be a referral of a foreign peer $l$ holding a prior tie to our focal local scientist $i$. This should be kept in mind when interpreting the regression results. For example, a positive effect of *foreign social capital* on *unique acquisition* may well involve a referral among foreign scientists taking place abroad. But neither can it be observed with our data nor is it the focus of the study at hand.

### 3.2.3 Explanatory variables

Various factors may influence foreign tie formation of SA scientists. We distinguish three sets of variables: attributes of the focal SA scientist (location of higher education, experience, scientific

[11] Thanks to an anonymous reviewer to help clarifying that.

quality, foreign social capital, and foreign affiliation), attributes of the focal's SA peers (peers' foreign ties, peers' foreign social capital, and peers' quality), and further characteristics of the focal individual (age, skin color, gender, and scientific discipline).

*3.2.3.1 Joint foreign ties*

SA scientists form and maintain co-author links to foreign scientists. The variable *joint foreign ties* measures how many (current) foreign co-authors of $i$ are also (current) foreign co-authors of $i$'s (current) SA co-author(s) $j$ in year $t$ (see Figure 1). Denote the number of foreign co-authors $i$ and $j$ create jointly in year $t$ as $y_{ij,t}$. *Joint tie formation* is then calculated for individual $i$ as $N_{ijk,t} = \sum_j y_{ij,t}$. The measure increases if two SA scientists write a paper together with a foreigner but also if the two SA scientists write one paper together and other papers in parallel with the same foreign scientist.

*3.2.3.2 Foreign social capital*

captures the international social capital of a scientist and is simply the size of the individual's neighborhood in the foreign co-authorship network (see Figure 1): $||N^a_{i,t}||$, which we will refer to simply as $N^a_{i,t}$.

*3.2.3.3 Peers' foreign social capital*

counts the total number of foreign co-authors accessible to an individual $i$ in exactly two steps through her South African social networks (see Figure 1). It is a count of the number of foreign collaborators in my neighbors' (cumulated) networks who are not in my own. There are two sources: my co-authorship neighbors ($N^{s,a}_{i,t}$) and my faculty neighbors ($N^{f,a}_{i,t}$).[12]

*3.2.3.4 Scientific quality*

is measured by the NRF rating. Researchers in our data undergo an NRF rating roughly every 5 years (see Section 3.1), and we use the ratings current at the time of the publication as a measure of quality ($q_{i,t}$). A scientist in a panel year before her first rating is assumed to be of the quality of her first (future) rating.

*3.2.3.5 Peers' scientific quality*

reflects the "quality" of the focal scientist's local (SA) colleagues. A focal scientist's South African social capital may be of higher or lower scientific excellence. The effect of excellence in SA social capital is ambiguous as it may affect the quest for, as well as the opportunity and success of, foreign tie formation. We measure the extent to which a focal scientist co-authored with high-quality scientists in the past simply by counting the number of rated researchers in her SA social capital, *rated co-authors* ($q^s_{i,t}$), and also how many rated scientists are found in the focal scientist's faculty (excluding $i$), *rated faculty* ($q^f_{i,t}$).

*3.2.3.6 Foreign affiliation*

states whether the scientist has had foreign employment. Our data do not allow to track all travels outside the country. However, it does contain foreign employments of our focal scientists. The variable *foreign affiliation* indicates whether or not there has been employment abroad in a certain year. The variable *prior foreign affiliations* ($s_{i,t}$) counts the number of foreign affiliation years during the preceding 5 years.

*3.2.3.7 Experience*

is measured in years in SA since PhD ($\tau_{i,t}$). This assumes that the experience clock of a scientist starts with her PhD and that scientific experience obtained in SA is particularly relevant for how scientists form foreign ties.[13]

[12] Formally: $||(\bigcup_{j\in(N^s_{i,t}\bigcup N^f_{i,t})} N^a_{j,t})\backslash N^a_{i,t}||$

[13] Alternatively, we considered years in SA and years since PhD separately in the regression, which did not affect the results. On the other hand, multiple versions of experience in the same regression creates multicollinearity issues.

*3.2.3.8 Higher education origin*

Higher education sets the stage for the subsequent scientific career. Foreign bachelor and *foreign PhD* indicate that initial condition. We indicate whether the first degree in higher education (mostly a bachelor degree, sometimes a master degree) was obtained abroad or in SA. Accordingly, we note where the PhD has been obtained. This yields four categories of HE (bachelor, PhD): SA-SA, Foreign-Foreign, SA-Foreign, and Foreign-SA.

*3.2.3.9 Scientific field*

Scientific collaboration varies over scientific fields. We use the definition of scientific fields used by the NRF for their rating process. Scientific field categories are the "usual" ones (see the Appendix in subsection Basic descriptives).

*3.2.3.10 Controls*

Control variables are *White* (based on color of skin) and *Male*. They may capture otherwise unobserved individual level heterogeneity in (social) opportunities.

Table 2 provides an overview of all variables.

## 3.3 Econometric methods

### 3.3.1 Model of foreign tie formation

The econometric model addresses the social process of foreign tie formation within the South African science system. The number of foreign co-author ties formed by a focal South African scientist $i$ at time $t$ is our dependent variable, $y_{i,t}$. A tie formation arises through the different network structures discussed in the previous section. These network motifs constitute further dependent variables, in order to identify the paths through which the explanatory variables work.

**Table 2.** Variables overview

| Variable name | Abbreviation | Mathematical symbol | Associated coefficient |
|---|---|---|---|
| *Outcome variables ($y_{i,t}$)* | | | |
| foreign ties (any) | for. ties | | |
| unique acquisition | unique acq. | | |
| joint acquisition | joint acq. | | |
| repeated tie | repeated | | |
| co-author referral | co-auth. ref. | | |
| faculty referral | fac. ref. | | |
| other | other | | |
| *Explanatory variables* | | | |
| joint foreign ties | joint for. | $\mathrm{N}_{ijk,t}$ | $\alpha$ |
| foreign social capital | for. soc. cap. | $\mathrm{N}^{a}_{i,t}$ | $\beta_0$ |
| peers' foreign social capital | | | |
| - *of co-authors* | co-auth. f.s.c. | $\mathrm{N}^{s,a}_{i,t}$ | $\beta_1$ |
| - *in faculty* | fac. f.s.c. | $\mathrm{N}^{f,a}_{i,t}$ | $\beta_2$ |
| scientific quality | quality | $q_{i,t}$ | $\gamma_0$ |
| peers' scientific quality | | | |
| - *of co-authors* | co-auth. quality | $q^{s}_{i,t}$ | $\gamma_1$ |
| - *in faculty* | fac. quality | $q^{f}_{i,t}$ | $\gamma_2$ |
| prior foreign affiliations | prior for. aff. | $s_{i,t}$ | $\delta$ |
| experience in SA | experience | $\tau_{i,t}$ | $\tau_t$ |
| higher education origin | HE | $x_{i,1}$ | $\theta_1$ |
| scientific field | sci. field | $x_{i,2}$ | $\theta_2$ |
| white | white | $x_{i,3}$ | $\theta_3$ |
| male | male | $x_{i,4}$ | $\theta_4$ |
| calendar year | year | $t_t$ | $t_t$ |

We consider a linear relationship[14] (in logs):

$$
\begin{aligned}
y_{i,t} = {} & \alpha N_{ijk,t} + \beta_0 N^{a}_{i,t} + \beta_1 N^{s,a}_{i,t} + \beta_2 N^{f,a}_{i,t} \\
& + \gamma_0 q_{i,t} + \gamma_1 q^{s}_{i,t} + \gamma_2 q^{f}_{i,t} + \delta s_{i,t} + \boldsymbol{x_i}'\boldsymbol{\theta} + t_t + \tau_t + \eta_i + u_{i,t}
\end{aligned} \tag{1}
$$

This equation is a dynamic peer effects model. The outcome of interest is an individual's foreign ties formed at a given time, $y_{i,t}$. Agents form foreign ties simultaneously, often by co-authoring the same papers ($N_{ijk,t}$). This simultaneity effect is captured in the first term of the equation by $\alpha$. Foreign tie formation can be conceptualized as a process of resource accumulation and use. Thus, current foreign tie formation may depend on the stock of foreign ties directly available to the focal agent $N^{a}_{i,t}$ but also on that available indirectly through her peers, i.e., local co-authors $N^{s,a}_{i,t}$ and local faculty $N^{f,a}_{i,t}$. In other words, the model allows for a resource effect arising from the agent's own resources ($\beta_0$) and a resource spill-over effect from local co-authors and faculty ($\beta_1$ and $\beta_2$, respectively). In addition, the scientific quality of the agent ($q_{i,t}$) may affect her ability to attract collaborators. Being connected to high-quality SA peers may also play a role, be it through the co-author network ($q^{a}_{i,t}$) or the faculty network ($q^{f}_{i,t}$). The peers' quality effect is commonly termed a contextual effect in the peer effects literature, here captured through $\gamma_1$ and $\gamma_2$. Prior foreign affiliations ($s_{i,t}$) provide rich opportunities for forming foreign ties. Further controls are included, in our case $\boldsymbol{x}$ is fixed over time as it incorporates *higher education origin*, *scientific field*, *White*, and *Male*.

The model allows further for nonlinear effects of calendar time as well as individual clocks starting at time of entry in the SA system after PhD, through time dummies $t_t$ and $\tau_t$. And there is an individual fixed effect $\eta_i$. Random shocks are captured by $u_{i,t}$.

Note that this typology of peer effects distinguishes sources of influence but not necessarily the nature of the foreign tie nor the mechanism through which it is formed. Above we identified several "types" of foreign ties (repeat, referral, etc.) and one might expect that different types of resources might have differential effects on different types of ties. To address this possibility, we treat each type of tie as a separate dependent variable, running the regression separately on each type of tie.

### 3.3.2 Estimation

Parameters of interest are estimated in two steps.

In the first step, we remove the effect of calendar years $\tau_t$ and all individual-level constant (observed) factors $\boldsymbol{X}$ through projection, also termed partial regression. The procedure avoids the incidental parameter issue and reduces heteroscedasticity of the error term in the generalized method of moments (GMM) regression. The partial regression transforms all factors that remain in the model by essentially demeaning them. Thus, it introduces necessarily some correlation into the (transformed) error term across time and further factors. We assume the resulting correlation to be negligible because averaging is over many observations.

In the second step, we estimate the (transformed) model by a systems GMM (Arellano and Bond, 1991). The systems GMM considers a level equation,

$$
\begin{aligned}
y_{i,t} = {} & \alpha N_{ijk,t} + \beta_0 N^{a}_{i,t} + \beta_1 N^{s,a}_{i,t} + \beta_2 N^{f,a}_{i,t} \\
& + \gamma_0 q_{i,t} + \gamma_1 q^{s}_{i,t} + \gamma_2 q^{f}_{i,t} + \delta s_{i,t} + \tau_t + \eta_i + u_{i,t}
\end{aligned} \tag{2}
$$

and an equation in differences,

[14] The number of foreign ties is a count with excess zeros (three quarters of scientist-year observations are without foreign ties), and over dispersion (mean 1.693, s.d. 5.250). We nevertheless adopt here a linear-in-logs model in order to handle better network dependence in observables and unobservables.

$$\Delta y_{i,t} = \alpha \Delta N_{ijk,t} + \beta_0 \Delta N_{i,t}^{a} + \beta_1 \Delta N_{i,t}^{s,a} + \beta_2 \Delta N_{i,t}^{f,a} + \gamma_0 \Delta q_{i,t} + \gamma_1 \Delta q_{i,t}^{s} + \gamma_2 \Delta q_{i,t}^{f} + \delta \Delta s_{i,t} + \Delta \tau_t + \Delta u_{i,t} \tag{3}$$

The difference equation removes the individual fixed effect $\eta_i$ and furthermore breaks the first-order correlation in the error term through differencing, i.e., $\Delta u_{i,t}$.

Most right-hand side variables are neither strictly nor weakly exogenous. For these endogenous factors, we have to find valid (and sufficiently strong) instruments.

Estimation of peer effects is a tricky issue (see Bramoullé *et al.*, 2020) that has been tackled in many ways. But there are only a few studies on network peer effects in a panel setting. The literature on network econometrics provides identification results for network cross-sections using instruments based on lags in network space (Bramoullé *et al.*, 2009; Liu *et al.*, 2014). This however is only a promising approach as long as the network is sufficiently exogenous as a whole (as for example for random assignment to dormitories as in Sacerdote 2001) or at least in parts (for example sudden deaths of colleagues in Azoulay *et al.* 2010). For the instrument to be valid, one needs also the fact that there are no other pathways except through the endogenous factor of how the instrument may affect the outcome, a common critique to the sudden death literature.

Our estimation strategy is therefore to exploit the time dimension of our sample. The large literature on dynamic models in panel GMM (up to Arellano and Bond's system GMM) provides some guidance here. Note that for the level equation, instruments need to be "net of" individual fixed effects (typically created through differencing of instruments). Furthermore, autoregressive correlation of error terms over time remains potentially an issue in the level equation.

Clearly, each instrument affects all parameter estimates. However, it is helpful to specify for each r.h.s. factor at least one valid instrument. We do that in the following.

#### *3.3.2.1 Simultaneity effect*

Multiple local scientists often collaborate jointly (or in parallel) with foreign scientists, creating an endogenous peer effect. The simultaneity effect in the model is captured by $N_{ijk,t}$. This is an endogenous factor in the sense that $E\left[(N_{ijk,t})u_{it}\right] \neq 0$ for several reasons. First, common foreign ties are by construction part of the outcome $y_{i,t}$ and therefore correlated with the error term. Second, tie formation among South African scientists is part of the same (network formation) process. Imagine for example a paper between two locals $i$, $j$, and a foreigner $k$ to capture some (unobserved by the econometrician) research opportunity. In that case, the research opportunity would enter all three components $u_{i,t}$, $u_{j,t}$, and $(i \sim j)$, creating a positive correlation.

We propose as an instrument $(N_{\neg ijk,t})$, i.e., the number of foreign ties of $i$'s local co-authors that are **not** collaborating with $i$ (see also Figure 1). This measure is not correlated with $u_{i,t}$ as long as $E[u_{i,t}, u_{j,t}^{*}] = 0$, where $u_{j,t}^{*}$ is the part of $j$'s shock that gives rise to foreign tie formation not involving $i$.[15] In the level equation, the unobserved component includes individual $i$'s fixed effect but those are purged out in the difference equation. This prompts us to impose the population orthogonality condition $E\left[N_{\neg ijk,t}\Delta u_{i,t}\right] = 0$.[16]

For the level equation, a similar reasoning holds, and we use the differenced instrument $\Delta N_{\neg ijk,t}$.

#### *3.3.2.2 Social capital effect*

The estimation model is akin to an auto-regressive model, with foreign social capital $N_{i,t}^{a}$ taking the role of a lagged outcome.

To fix ideas, consider $N_{i,t}^{a} \approx \sum_{s=1}^{5} y_{i,t-s}$ (this is approximate because the same foreign tie is in fact not counted multiple times in the stock of foreign social capital). Because $N_{i,t-1}^{s} \approx \sum_{s=2}^{6} y_{i,t-s}$, lagged social capital may serve as instrument in the difference equation $E\left[N_{i,t-1}^{a}\Delta u_{i,t}\right] \approx$

[15] Obviously, it must be also the case that individual fixed effects of $j$ are independent of $i$'s error term.

[16] In the context of a more comprehensive network formation model with individual tie formation as dependent variable, the equivalent condition would be that correlation of shocks dies out within two steps in the network. For SA scientist $i$ and $j$ and foreigners $k$ and $m$, such a path would be $kj - \{ji, jm\} - im$.

$E\left[\sum_{s=2}^{6} y_{i,t-s}(u_{i,t} - u_{i,t-1})\right] = 0$, in the case that errors are not correlated over time. On the other hand, $\Delta N^{a}_{i,t} \approx y_{i,t-1} - y_{i,t-6}$, and hence for the level equation, we impose the condition $E\left[\Delta N^{a}_{i,t} u_{i,t}\right] = 0$.

*3.3.2.3 Peers' social capital effect*

The estimation model incorporates an effect of a local scientist *i*'s peers' foreign social capital on *i*'s foreign tie formation through the term $N^{s,a}_{i,t}$ (social capital in the SA prior co-author network) and $N^{f,a}_{i,t}$ (social capital of faculty colleagues). One can imagine various mechanisms that create a positive (or negative) relationship, including referrals (of multiple steps) and demonstration effects. The endogeneity issue is similar to the individual's own social capital effect but supposedly less severe because correlation of *i*'s shocks and unobserved components with others' social capital is probably less strong. Individual fixed effects may lead still to correlation through SA network formation through a birds-of-a-feather effect. We therefore follow the same strategy as for the social capital effect, that is, purging out individual fixed effects through differencing, i.e., differencing the equation or differencing the instruments. Hence, we obtain for the difference equation $E\left[N^{s,a}_{i,t-1} \Delta u_{i,t}\right] = 0$ and for the level equation $E\left[\Delta N^{s,a}_{i,t} u_{i,t}\right] = 0$.

*3.3.2.4 Scientific excellence effect*

Scientific excellence of the scientist is considered a slowly changing characteristic, largely independent of current shocks. We consider $q_{i,t}$ therefore to be exogenous, i.e., $E\left[q_{i,t} u_{i,t}\right] = 0$.

Note that our measure of scientific excellence may actually be considered as a good proxy for individual fixed effects ($\eta_i$). Thus, the individual fixed effects we are dealing with in the model are actually fixed effects net of ratings. Hence, one may also term them systematic measurement errors in our proxy of foreign-oriented scientific excellence through NRF ratings.

*3.3.2.5 Peers' scientific excellence effect*

Scientific excellence in one's local social capital (SA co-author and/or faculty) may enhance foreign tie formation, but it may also be a substitute for it. Local scientific excellence in the personal SA co-author network and SA faculty may be influenced by individual fixed effects. For the co-author network, we again create lags for the difference equation and differences for the level equation to get rid of fixed effects in one term of the moment condition.

In the case of excellence in faculty, we assume that sorting into faculties is mostly on observed excellence (with outward focus), which is well captured by the rating. Therefore, we attempt to instrument that factor by itself.

*3.3.2.6 Controls*

Controls are gender, skin color, and time, which we all assume to be strictly exogenous. Time is measured since entry into the SA system as faculty (after obtaining a PhD), which we include as dummies.

## 3.4 Estimation sample

Scientists with missing information—in sociodemographic variables as well as publication details—are excluded from the panel. This leaves us with 2160 individuals in 21,119 observations, corresponding to a reduction in the original population by about 20%. Our estimation procedure imposes further requirements on the data, which reduces the number of individuals in the panel by another 10%. In detail, the systems GMM approach asks for a panel where each individual is observed in at least three time periods. In our setting, we are dealing essentially with a lagged dependent that is accumulated over 5 years. In order to increase variation of respective variables within individuals, we include only scientists observed over at least 5 years. On the other hand, we exclude scientists from the panel after being more than 20 years in faculty, in order to maintain a sufficient number of observations for the year dummy estimates.[17]

[17] Excluding individuals from the panel however does at no point reduce the precision with which the social environment of scientists remaining in the sample is measured, as the measurement of co-author networks and faculty composition is based on all available information. We further found that the exclusion of individuals does not change

The final estimation sample contains 1552 individual scientists observed in 16,055 scientist-years between 1996 and 2014. Table 3 shows the mean, standard deviation, and Pearson's correlation coefficient of the variables in the estimation sample.

## 4. Results

The econometric model has been estimated on the sample of rated South African researchers in the sciences (Table 4). In a later section, we do the estimations separately on sub-samples of researchers with South African and foreign higher education (Figure 3) and separately for scientific fields (Figure 4). Insights obtained on the whole sample are generally robust, largely carry through to researchers with higher education from SA as well as from abroad, and tend to hold across individual scientific fields.

### 4.1 Foreign tie formation in South African science

First consider Table 4, showing coefficient estimates when regressing on different types of tie formation (all types, unique, joint, repeated, co-author referral, faculty referral, and others). For each tie type, the odd-numbered column excludes jointly created foreign ties in that period (*joint for.*); the even-numbered column includes it. In total (columns 1 and 2), we count 25,013 foreign ties being formed. This total aggregates six different tie formation patterns that are considered in subsequent columns.

Before turning to the coefficient estimates, we observe model specification test statistics at the bottom of the Table. In agreement with the estimation assumption that there is no serial correlation above order 1 in the first differences of the error term (Arellano and Bond, 1991), across all models, $m1$ strongly rejects the null of *no* first-order serial correlation at a significance level below 0.001, whereas $m2$ accepts *no* second-order serial correlation. Sargan-Hansen's test of over-identifying restrictions, see *sargan*, delivers $p$-values at acceptable (high) levels.[18]

Standard tests for systems GMM are complemented by network correlation test statistics. In detail, the table provides a permutation test of network correlation of residuals obtained from the difference equations (Moran's I diff.) and from the level equations (Moran's I level). We observe in differences and in levels a highly significant but moderate network auto-correlation below 0.1. Note that (high) network auto-correlation of residuals does not imply a bias in coefficient estimates, because GMM orthogonality conditions do not involve network instruments in our set-up. However, standard asymptotic GMM standard errors are likely to be biased upwards if the iid assumption on the error terms does not hold. Therefore, the table presents bootstrapped standard errors obtained from random resampling of individuals, which we have indeed found to be higher than the asymptotic ones.[19]

We turn now to the coefficient estimates.

Model (1) in Table 4 regresses foreign ties on lagged factors only. The elasticity of foreign tie formation on a focal scientist's foreign social capital (for. soc. cap.) is estimated to be highly significant at 0.36, meaning that a 1% increase in one's foreign social capital (accumulated in the past) can be expected to result in a 0.36% increase in current foreign tie formation. Elasticity with respect to co-authors' foreign social capital is estimated to be much lower, at 0.06, but significant at a level below 1%. Prior foreign affiliations, as expected, also have a significantly positive effect on foreign tie formation. On the other hand, foreign social capital at the faculty (beyond the agents' and co-authors' foreign social capital) is small and insignificant. Scientific

the relative composition of the sample (in the Appendix), which is an argument against the idea that the panel reduction creates an estimation bias.

[18] Yet, note that in estimations of repeated ties, co-author referrals and faculty referrals the sargan $P$-value goes below 10 percent. This raises the concern that our GMM instruments may be invalid, but could be also caused by the sensitivity of the test to its iid assumption regarding unobserved components. In further estimations based on more homogenous sub-samples, and hence more homogenous unobserved components, the sargan $P$-value tends to be much higher (see Appendix Tables). This suggests that low Sargan $P$-values in the whole sample estimation may result from relatively high heteroscedasticity of error terms.

[19] One can remark also that the model specification tests applied, m- and Sargan-tests, also rely on the assumption of iid error terms, and may therefore be biased (up- or down-ward). However, given the modest adjustment of standard errors achieved with the bootstrap, we don't expect that bias to be strong.

**Table 3.** Descriptive statistics (16,055 observations on 1552 scientists)

| | Mean | Std.dev. | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) for. ties | 1.558 | 4.841 | | | | | | | | | | | | | |
| (2) unique acq. | 0.302 | 1.292 | 0.511 | | | | | | | | | | | | |
| (3) joint acq | 0.685 | 2.827 | 0.816 | 0.162 | | | | | | | | | | | |
| (4) repeated | 0.455 | 2.028 | 0.729 | 0.295 | 0.31 | | | | | | | | | | |
| (5) co-auth. ref. | 0.055 | 0.44 | 0.396 | 0.071 | 0.239 | 0.298 | | | | | | | | | |
| (6) fac ref. | 0.066 | 0.466 | 0.414 | 0.082 | 0.258 | 0.299 | 0.955 | | | | | | | | |
| (7) other | 0.049 | 0.415 | 0.494 | 0.204 | 0.39 | 0.247 | 0.235 | 0.238 | | | | | | | |
| (8) joint for. | 4.274 | 23.934 | 0.733 | 0.106 | 0.733 | 0.487 | 0.388 | 0.396 | 0.406 | | | | | | |
| (9) for. soc. cap. | 4.816 | 11.859 | 0.591 | 0.276 | 0.369 | 0.595 | 0.282 | 0.284 | 0.291 | 0.415 | | | | | |
| (10) co-auth. f.s.c. | 19.228 | 60.088 | 0.314 | 0.101 | 0.255 | 0.214 | 0.334 | 0.33 | 0.198 | 0.357 | 0.457 | | | | |
| (11) fac. f.s.c. | 160.054 | 335.557 | 0.139 | 0.086 | 0.123 | 0.077 | 0.059 | 0.086 | 0.047 | 0.114 | 0.182 | 0.339 | | | |
| (12) quality L | 0.033 | 0.179 | −0.039 | −0.025 | −0.029 | −0.028 | −0.017 | −0.017 | −0.02 | −0.015 | −0.057 | −0.016 | −0.057 | | |
| (13) quality Y | 0.155 | 0.362 | 0.032 | −0.007 | 0.042 | 0.013 | 0.013 | 0.014 | 0.027 | 0.051 | 0.006 | 0.032 | 0.11 | −0.079 | |
| (14) quality P | 0.01 | 0.102 | 0.017 | 0.011 | 0.008 | 0.02 | 0.011 | 0.009 | −0.002 | 0.012 | 0.024 | 0.023 | 0.023 | −0.019 | −0.044 |
| (15) quality C | 0.516 | 0.5 | −0.04 | −0.023 | −0.034 | −0.032 | 0.001 | 0.001 | −0.011 | −0.029 | −0.037 | −0.009 | −0.037 | −0.192 | −0.442 |
| (16) quality AB | 0.142 | 0.349 | 0.124 | 0.113 | 0.077 | 0.104 | 0.023 | 0.027 | 0.034 | 0.043 | 0.173 | 0.048 | 0.003 | −0.076 | −0.174 |
| (17) co-auth. quality | 2.584 | 3.236 | 0.268 | 0.083 | 0.251 | 0.164 | 0.211 | 0.212 | 0.12 | 0.257 | 0.37 | 0.565 | 0.195 | −0.041 | 0.004 |
| (18) fac. quality | 23.18 | 22.602 | 0.09 | 0.051 | 0.093 | 0.034 | 0.041 | 0.064 | 0.016 | 0.075 | 0.1 | 0.237 | 0.763 | −0.077 | 0.12 |
| (19) prior for. aff. | 0.23 | 0.826 | 0.128 | 0.146 | 0.061 | 0.104 | 0.023 | 0.031 | 0.082 | 0.043 | 0.141 | 0.045 | 0.028 | −0.015 | −0.047 |
| (20) experience | 8.646 | 5.316 | 0.046 | 0.022 | 0.044 | 0.027 | 0.011 | 0.012 | 0.021 | 0.026 | 0.085 | 0.093 | 0.03 | −.028 | −.316 |
| (21) HE SA-SA | 0.634 | 0.482 | −0.064 | −0.068 | −0.02 | −0.071 | −0.012 | −0.013 | −0.038 | −0.022 | −0.106 | 0.028 | −0.007 | 0.026 | 0.051 |
| (22) HE SA-for. | 0.09 | 0.286 | −0.01 | −0.005 | −0.005 | −0.007 | −0.019 | −0.015 | −0.009 | −0.002 | −0.021 | −0.034 | 0.029 | 0.007 | 0.032 |
| (23) HE for.-SA | 0.068 | 0.252 | −0.007 | −0.014 | 0.005 | −0.015 | 0.001 | −0.001 | −0.005 | 0.005 | −0.016 | 0.027 | 0.023 | −0.005 | 0.005 |
| (24) HE for.-for. | 0.208 | 0.406 | 0.087 | 0.093 | 0.024 | 0.098 | 0.027 | 0.026 | 0.054 | 0.025 | 0.151 | −0.026 | −0.026 | −0.033 | −0.086 |
| (25) white | 0.769 | 0.421 | 0.034 | 0.026 | 0.029 | 0.018 | 0.021 | 0.02 | 0.008 | 0.019 | 0.045 | 0.004 | −0.045 | −0.122 | −0.014 |
| (26) male | 0.717 | 0.451 | 0.002 | 0.037 | −0.024 | 0.021 | −0.024 | −0.022 | −0.009 | −0.032 | 0.012 | −0.039 | −0.095 | −0.045 | −0.071 |
| (27) year | 2004.456 | 4.622 | 0.149 | 0.059 | 0.14 | 0.088 | 0.07 | 0.082 | 0.076 | 0.131 | 0.189 | 0.271 | 0.383 | −0.039 | 0.111 |

| | (14) | (15) | (16) | (17) | (18) | (19) | (20) | (21) | (22) | (23) | (24) | (25) | (26) | (27) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| quality C (15) | −0.106 | | | | | | | | | | | | | |
| quality AB (16) | −0.042 | −0.42 | | | | | | | | | | | | |
| co-auth. quality (17) | −0.004 | 0.01 | 0.133 | | | | | | | | | | | |
| fac. quality (18) | 0.022 | −0.033 | 0.031 | 0.213 | | | | | | | | | | |
| prior for. aff. (19) | 0.039 | −0.016 | 0.127 | −0.018 | 0.036 | | | | | | | | | |
| experience (20) | −0.069 | 0.14 | 0.265 | 0.227 | 0.043 | 0.068 | | | | | | | | |
| HE SA-SA (21) | −0.049 | 0.059 | −0.091 | 0.143 | 0.023 | −0.11 | 0.111 | | | | | | | |
| HE SA-for. (22) | 0.101 | −0.046 | 0.018 | −0.037 | 0.057 | 0.014 | −0.006 | −0.413 | | | | | | |
| HE for.-SA (23) | 0.006 | 0.008 | −0.022 | 0.018 | 0.032 | −0.019 | −0.081 | −0.355 | −0.085 | | | | | |
| HE for.-for. (24) | −0.017 | −0.043 | 0.108 | −0.154 | −0.087 | 0.132 | −0.077 | −0.675 | −0.161 | −0.138 | | | | |
| white (25) | 0 | 0.066 | 0.135 | 0.119 | 0.015 | −0.017 | 0.181 | 0.331 | −0.051 | −0.126 | −0.279 | | | |
| male (26) | 0.017 | 0.005 | 0.119 | −0.042 | −0.114 | 0.063 | 0.08 | −0.171 | 0.032 | 0.048 | 0.151 | −0.118 | | |
| year (27) | −0.014 | −0.024 | −0.077 | 0.277 | 0.307 | 0.053 | 0.157 | −0.083 | −0.002 | 0.071 | 0.056 | −0.158 | −0.071 | |

Mean, standard deviations, and Pearson's correlation coefficient of outcomes and right-hand-side variables.
Observations with missing information excluded from the sample.
The variable scientific fields is not included.

**Table 4.** Foreign tie formation of rated scientists in the sciences, panel GMM

| | Foreign ties (25,013; 100%) | | Unique acquisition (4,854; 19.4%) | | Joint acquisition (11,003; 44.0%) | | Repeated tie (7,305; 29.2%) | | Co-author referral (891; 3.6%) | | Faculty referral (1,063; 4.2%) | | other (788; 3.2%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
| joint for. | – | 0.367*** | – | 0.004 | – | 0.305*** | – | 0.125*** | – | 0.035*** | – | 0.058*** | – | 0.049*** |
| | – | (0.024) | – | (0.018) | – | (0.021) | – | (0.017) | – | (0.007) | – | (0.01) | – | (0.011) |
| for. soc. cap. | 0.361*** | 0.211*** | 0.068*** | 0.065*** | 0.152*** | 0.038** | 0.263*** | 0.214*** | 0.009 | −0.001 | 0.014 | −0.004 | 0.009 | −0.005 |
| | (0.028) | (0.024) | (0.015) | (0.018) | (0.022) | (0.014) | (0.019) | (0.018) | (0.007) | (0.007) | (0.008) | (0.008) | (0.006) | (0.007) |
| co-auth. f.s.c. | 0.059** | 0.019 | 0.003 | 0.001 | 0.05** | 0.021 | 0.005 | −0.005 | 0.016*** | 0.013** | 0.016** | 0.011* | 0.009* | 0.006 |
| | (0.021) | (0.012) | (0.009) | (0.01) | (0.016) | (0.012) | (0.011) | (0.01) | (0.004) | (0.004) | (0.005) | (0.004) | (0.005) | (0.005) |
| fac. f.s.c. | −0.019 | −0.016 | −0.016 | −0.015 | 0 | 0.002 | −0.023 | −0.022 | −0.001 | −0.001 | 0.002 | 0.001 | 0.002 | 0.001 |
| | (0.032) | (0.022) | (0.021) | (0.02) | (0.026) | (0.016) | (0.02) | (0.018) | (0.008) | (0.008) | (0.009) | (0.009) | (0.008) | (0.007) |
| quality L | 0.021 | −0.018 | −0.02 | −0.018 | 0.021 | 0.005 | −0.014 | −0.036 | 0.006 | 0.001 | 0.018 | 0.009 | 0 | −0.003 |
| | (0.092) | (0.083) | (0.063) | (0.069) | (0.075) | (0.055) | (0.052) | (0.05) | (0.012) | (0.015) | (0.018) | (0.018) | (0.013) | (0.014) |
| quality Y | 0.055 | 0.074 | 0.075* | 0.076 | −0.03 | −0.009 | 0.028 | 0.036 | 0 | 0.001 | −0.009 | −0.004 | −0.005 | −0.004 |
| | (0.071) | (0.056) | (0.036) | (0.039) | (0.05) | (0.037) | (0.041) | (0.044) | (0.015) | (0.016) | (0.02) | (0.018) | (0.014) | (0.013) |
| quality P | 0.176 | 0.154 | 0.178 | 0.172 | 0.077 | 0.017 | −0.027 | −0.041 | −0.003 | −0.011 | −0.023 | −0.038 | −0.011 | −0.009 |
| | (0.267) | (0.238) | (0.201) | (0.192) | (0.25) | (0.213) | (0.156) | (0.165) | (0.105) | (0.108) | (0.104) | (0.089) | (0.048) | (0.052) |
| quality C | −0.003 | 0.035 | 0.05 | 0.054 | −0.028 | 0.002 | −0.042 | −0.027 | −0.001 | 0.001 | 0.001 | 0.003 | −0.003 | −0.003 |
| | (0.051) | (0.048) | (0.029) | (0.03) | (0.044) | (0.031) | (0.03) | (0.03) | (0.011) | (0.012) | (0.014) | (0.015) | (0.012) | (0.013) |
| quality AB | 0.018 | 0.069 | 0.08 | 0.086 | 0.017 | 0.051 | −0.071 | −0.057 | −0.014 | −0.013 | −0.017 | −0.016 | −0.01 | −0.013 |
| | (0.085) | (0.064) | (0.053) | (0.054) | (0.071) | (0.04) | (0.051) | (0.054) | (0.018) | (0.017) | (0.02) | (0.022) | (0.019) | (0.019) |
| co-auth. quality | 0.038 | −0.025 | −0.017 | −0.017 | 0.051* | −0.006 | 0.023 | 0.004 | 0.005 | −0.002 | 0.003 | −0.009 | −0.009 | −0.018* |
| | (0.042) | (0.026) | (0.023) | (0.023) | (0.025) | (0.021) | (0.021) | (0.019) | (0.008) | (0.008) | (0.008) | (0.008) | (0.009) | (0.008) |
| fac. quality | 0.004 | 0 | 0.009 | 0.01 | 0.004 | 0.001 | 0.001 | −0.004 | 0 | 0 | 0.002 | 0.002 | −0.002 | −0.001 |
| | (0.042) | (0.032) | (0.026) | (0.025) | (0.03) | (0.021) | (0.023) | (0.022) | (0.01) | (0.009) | (0.011) | (0.01) | (0.009) | (0.008) |
| prior for. aff. | 0.053** | 0.043*** | 0.036*** | 0.036** | −0.001 | −0.009 | 0.031* | 0.033** | 0.002 | 0.002 | 0.003 | 0.004 | 0.006 | 0.005 |
| | (0.018) | (0.012) | (0.01) | (0.012) | (0.014) | (0.009) | (0.012) | (0.012) | (0.005) | (0.005) | (0.006) | (0.005) | (0.005) | (0.004) |
| m1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| m2 | 0.21 | 0.115 | 0.079 | 0.079 | 0.196 | 0.413 | 0.558 | 0.553 | 0.359 | 0.365 | 0.276 | 0.335 | 0.263 | 0.243 |
| sargan | 0.493 | 0.416 | 0.762 | 0.695 | 0.562 | 0.887 | 0.136 | 0.052 | 0.078 | 0.062 | 0.093 | 0.068 | 0.093 | 0.245 |
| moran.s.diff | 0.047(0) | 0.006(0.063) | −0.012(1) | −0.012(0.999) | 0.082(0) | 0.057(0) | 0.033(0) | 0.025(0) | 0.051(0) | 0.044(0) | 0.049(0) | 0.039(0) | 0.027(0) | 0.023(0) |
| moran.s.level | 0.069(0) | 0.012(0) | −0.008(0.98) | −0.007(0.981) | 0.121(0) | 0.082(0) | 0.056(0) | 0.043(0) | 0.1(0) | 0.08(0) | 0.1(0) | 0.073(0) | 0.043(0) | 0.032(0) |
| indiv. | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 | 1552 |
| obs. | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 | 16055 |

SET sample, all models include time dummies for individual experience, $\tau_t$.
Estimation after residual regression to remove scientific field $\psi_f$ and calendar year shocks $t_t$.
Coefficient standard errors obtained through 100 bootstrap estimations.
m1 and m2 provide *P*-value for Arellano and Bond's (1991) m-test on first-order and second-order serial correlation in the first-differenced residuals.
Sargan provides the (statistic, d.o.f., and p-val) of Sargan–Hansen test.
Moran's I provides Moran's I statistic (*P*-value) on residuals in differences and levels, with *P*-value from permutation test (999 permutations on each year-network).
*, **, and *** denote significance levels of 5%, 1%, and 0.1%, respectively.

quality, as measured through ratings, are in general consistent with the idea that higher quality is associated with foreign tie formation.[20] But these estimates are not significant, once controlling for individual fixed effects and (accumulating) foreign social capital. The same is true for scientific quality among co-authors and in faculty.[21]

Model (2) includes the variable joint foreign tie formation (*joint for.*) to capture the contemporaneous contribution of local co-authors to the focal scientist's current foreign tie formation. The coefficient is estimated to be high, 0.36, and highly significant. Thus, collaboration with other South African scientists is an important contributor to foreign tie formation. But observe that compared to the first model, the coefficients of (own) foreign social capital (for. soc. cap.) and co-authors' foreign social capital (co-auth. f.s.c.) drop considerably. The effect of prior foreign affiliations also gets weaker though to a lesser extent.

This leads to an important result: the first two models combined suggest that (current) *joint* foreign tie formation mediates[22] the effect of foreign social capital (accumulated in the past) on (current) acquisition of foreign ties. Foreign social capital is a relevant resource in foreign tie formation. But it is activated through collaboration among local scientists. This implies that collaboration among local scientists plays an important role in the formation of their foreign social capital.

A more detailed interpretation of the coefficient estimate of "*joint* foreign ties" of SA co-authors requires us to recall that we instrument this variable by SA co-authors' "*separate* foreign ties" (i.e., those not involving the focal scientist). Roughly, the main condition for that instrument to be valid is that random shocks affecting the focal's foreign ties are not correlated with shocks affecting the foreign ties of his SA co-authors that do not involve the focal scientist. If the instrument is valid in that sense, the positive coefficient estimate implies that SA co-authors positively contribute to tie formation between the focal scientist and his foreign collaborators.[23] It is possible however that the instrument is not perfect. For example, a research program could result in shocks that are correlated across partly overlapping (international) co-author teams.[24] In that case, the coefficient estimate of "*joint* foreign ties" would capture the presence of some unobserved factor (in the example the research program) that contributes to foreign tie formation for everyone in the collective involving the focal scientist and his SA co-authors. If the instrument is valid, there is a positive effect from SA co-authors to the focal SA scientist. If the instrument is invalid, then the focal scientist and his SA co-authors build on something common in their foreign tie formation. One way or the other, some form of collaboration among locals seems beneficial to foreign tie formation.

Unique acquisitions (models 3 and 4)—a focal South African establishes a unique co-authorship tie with a foreigner who is new to South Africa—account for around 20% of foreign tie formation. In this case, the most relevant resource is the focal's foreign social capital and having had a foreign affiliation. Peers' foreign social capital does not play a role nor do we observe a (mediating) effect of local scientific interaction.

Joint acquisitions (models 5 and 6), i.e., a focal South African establishes together with peers a co-authorship tie with a foreigner who is new to South Africa, account for 44%—the biggest share of foreign tie formation. The focal scientists' and peers' foreign social capital are both valuable resources when it comes to joint acquisitions (model 5), but which are again effectively leveraged through joint foreign ties (model 6). In order to avoid confusion here, recall that the model

[20] There is a positive effect of having a rating compared to the base category "unsuccessful rating," and the effect is increasing for senior ratings (AB higher than C), and (P higher than Y).

[21] Note that ratings change slowly and so are likely to be closely related to individual fixed effects. Corresponding coefficients are thus mostly identified through the level equation resulting in weaker identification.

[22] We follow here the basic reasoning of Judd and Kenny (1981), namely that the left-hand-side variable 'social capital' affects significantly the outcome, e.g. 'foreign ties', and this estimated effect reduces significantly once the mediating variable 'joint tie formation' is included.

[23] Note that the results of Sargan-Hansen's test of over-identifying restrictions in Table 4 suggest that instruments are indeed valid.

[24] In the main regressions, Table 4, such an effect could arise through national programs at the level of the discipline, since we control for discipline and calendar year effects but not their interaction. However, extended regressions on individual disciplines (Figure 3 below) control for discipline-time-specific shocks and therefore factors affecting the instrument would be confined to research programs localized to the level of the lab for example. Because estimates are similar in both series of regressions, local programs seem to be the more realistic possibility here.

outcome, joint acquisitions, relates to the number of joint new acquisitions claimed by a focal scientist, whereas the estimate of the explanatory variable, joint foreign ties, relates to the focal scientists' peers contribution in achieving this outcome. Consistently, prior foreign affiliation does not contribute to joint acquisitions.

Repeating a foreign tie (models 7 and 8) represents about 30% of all ties. It strongly depends on the focal author's foreign social capital, as the elasticity estimate of 0.26 in model 7 suggests. Prior foreign affiliations also contribute somewhat, with a coefficient of 0.03. The (mediating) effect of joint foreign ties is weaker when it comes to repeating ties but is still considerable. Thus, local scientific collaboration also contributes to persistence of established foreign ties.

Referrals, be it among co-authors (models 9 and 10) or within faculty (models 11 and 12), are rarely observed, and this may limit identification of causal effects. Nonetheless, taken at face value, the effect of a focal's foreign social capital is estimated to be negligible. This makes sense because the foreign co-author in question is not an own acquaintance but that of a local peer. However, the elasticity of referrals on peers' social capital is estimated to be relatively small, about 0.016. Similarly, the effect of joint foreign tie formation seems limited, perhaps because joint efforts are more often directed to acquiring new foreign ties.

There are three general conclusions we can draw here. First is that the foreign social capital that a South African scientist accumulates from past collaborations increases the probability that he or she will create future international collaborations. Second is that there are "spillovers": if South Africans A and B have collaborated in the past, the larger is the stock of B's foreign social capital, the more likely it is that A will have foreign collaboration in the future. Finally, when the number of current jointly acquired foreign co-authors is included as an explanans, the coefficients on social capital tend to fall significantly. This is consistent with the idea that *A*'s taking advantage of *B*'s foreign social capital is facilitated if A and B collaborate jointly with the foreign scientist.

## 4.2 Extensions: Education and discipline

In this section, we consider two possible confounding effects. First, we consider that there may be differences between scientists trained within South Africa and those trained abroad. Second, there may be disciplinary differences in the way that co-authorships and foreign co-authorships in particular are formed. We examine both those issues here.

### 4.2.1 Foreign tie formation by higher education origin

Scientists with foreign higher education tend to enter the South African system with higher foreign social capital than do their new South African colleagues. Being "raised" abroad may also not only create a "taste for foreign ties" but bring real additional advantages in international collaboration.

From this perspective, it may be surprising that estimations of foreign tie formation on social capital are remarkably similar for South African- and foreign-trained scientists. Figure 3 plots estimations of different modes of tie formation (along the x-axis) against associated coefficients (along the y-axis) for the population of SA-trained and locally trained scientists including and excluding joint foreign tie formation (varying colors).

There are some small differences, which while making sense intuitively are mostly insignificant. South African–trained scientists seem in general to benefit more from their co-authors' foreign social capital (see column "for. ties" and row "co-auth. f.s.c."). On the other hand, foreign-trained scientists seem to benefit more from foreign affiliations. It allows them not only to form unique acquisitions, as it does for South African–trained scientists (see column "unique acq." and row "prior for. aff."), but moreover is important for them in repeating existing foreign ties (see column "repeated" and row "prior for. aff.").

### 4.2.2 Foreign tie formation by scientific fields

How research is done depends on the specific scientific field. Figure 4 suggests that this is also true when it comes to foreign tie formation—despite the fact that standard errors get relatively large due to small sample sizes. In physics for example, elasticities on foreign tie formation of own and co-authors' foreign social capital are relatively high (gold dots). Biologists and physicists benefit

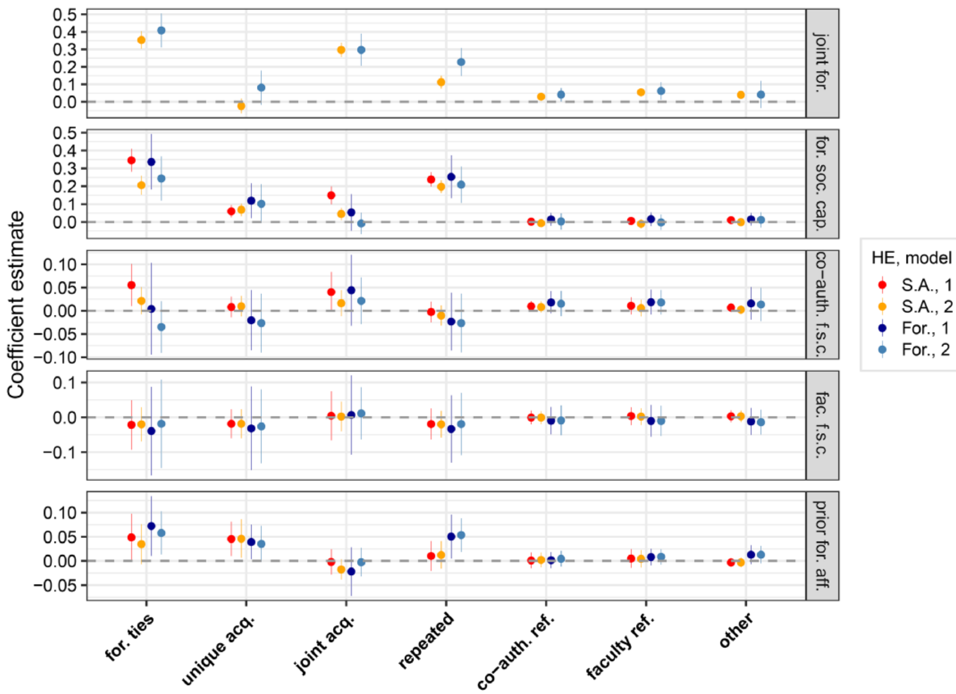


**Figure 3.** Coefficient estimates of foreign tie formation (horizontal axis) on social capital variables (right vertical axis). Model 1 (2) excludes (includes) joint foreign tie formation (joint for.). Estimations obtained separately on scientists with South African higher education (S.A. HE) and on scientists with foreign higher education (For. HE). Bars stretch out over twice the bootstrapped standard errors up and down the coefficients (dots). Sample and model statistics are provided in the appendix.

from foreign stays to form unique acquisitions, Earth and marine scientists seem to mostly repeat existing collaborations abroad, and for social scientists, the effect of having a foreign affiliation is estimated to be zero in either case.

However, despite all these differences across the sciences, we observe that whenever foreign social capital is relevant, it tends to be mediated by scientific collaboration among South African scientists.

## 4.3 Robustness

The model has been built up gradually, and the last step was to remove dummies for the controls “white” and “male” from GMM estimation and rather include them into the partial regression before GMM. Estimation results for the other parameters have not changed, but GMM model statistics have slightly improved.

### 4.3.1 Time window for social capital

Regression results in the main text include social capital variables on the right-hand-side and longitudinal network motifs on the left-hand-side that are obtained on the co-author network aggregated over the past 5 years. We did the same regressions for a time window of 3 years and 7 years, respectively. First, the total number of foreign ties formed remains the same, whereas counts of the motifs change slightly. Logically, a shorter time window moves some cases of “known foreigner” (i.e. repeated tie, referral, and others) to “new acquisition” (unique and joint). Observed shifts of motifs are very modest; for 3 years up to one percentage point and for 7 years up to half a percentage point. Thus, most foreign ties that are repeated or referred are not older than 3 years, and very few are older than 5 years.

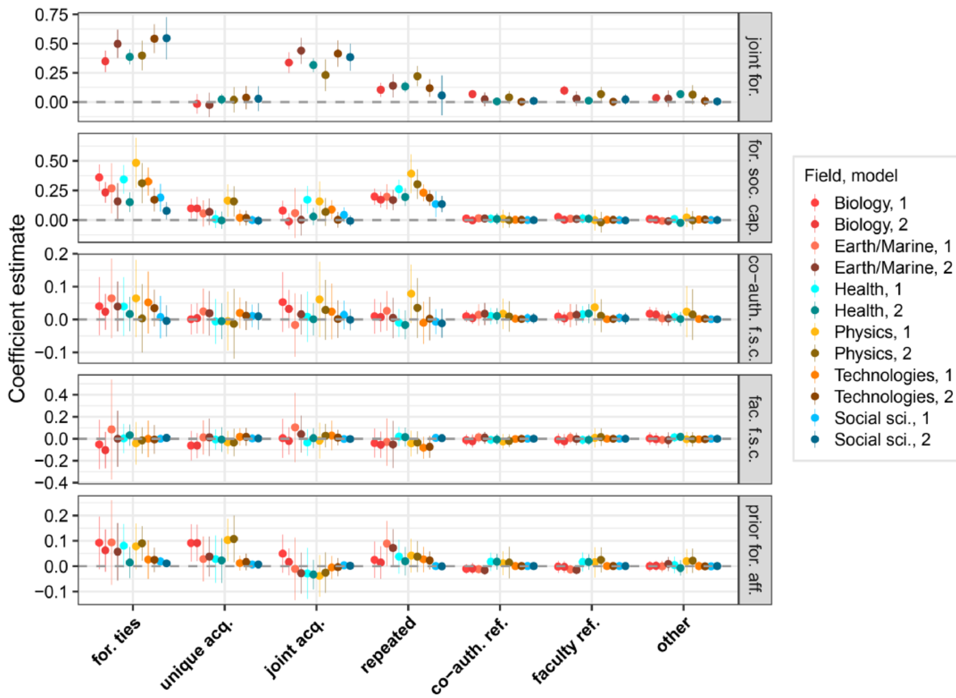


**Figure 4.** Coefficient estimates of foreign tie formation (horizontal axis) on social capital variables (right vertical axis). Model 1 (2) excludes (includes) joint foreign tie formation (joint for.). Estimations obtained on individual scientific fields. Bars stretch out over twice the bootstrapped standard errors up and down the coefficients (dots). Sample and model statistics are provided in the appendix.

For both alternative time windows, the model specification tests continue to support validity of estimations, and estimation results are robust. Coefficient estimates change never beyond two standard deviations. In the case of a 3-year window of social capital, peers' foreign social capital effect and the mediation effect through joint tie formation become somewhat stronger (in particular in joint acquisition), whereas the effect of own social capital becomes slightly weaker (in particular in unique acquisitions and repeated ties). The opposite happens if the time window is 7 years, i.e., slightly weaker effect of peers' foreign social capital and stronger effect of own social capital. One possible explanation is that own foreign social capital measured on a longer time window is a better proxy for international embeddedness, whereas peers' international contacts become more relevant the more recent they are.

### 4.3.2 Excess zeros in the outcome

In many scientist-years, no foreign ties are formed. This results in excess zeroes in the (non-transformed) outcome variables for which the log-linear model applied is not an optimal choice. As discussed earlier, we refrain from applying a zero-inflated count data model because this would create several complications in our model (in particular when it comes to the partial regression before GMM and network effects in GMM). In order to verify that this issue does not drive our results, we repeated all estimations on samples restricted to actively publishing individuals, i.e., publishing on average two papers per year. Results are found to be robust.

## 5. Discussion and conclusion

Many of our results are both consistent with prior literature and intuitively plausible. The likelihood of an individual scientist forming foreign ties is increased by a strong stock of existing foreign ties as well as by a prior foreign affiliation. Where a South Africa–based scholar was educated abroad, repeated collaboration with the same foreign ties is more likely than for a locally educated scholar.

However, these insights reflect a preoccupation with the individual scientist. This atomistic view has characterized much of prior literature on the benefits of collaboration in general and in particular of collaborations with scientists from abroad for countries behind the frontiers of knowledge creation. Instead, a key contribution of this study is our insistence that the generation of knowledge is a collective endeavour. We suggest that an exclusive focus on the individual and her human capital tells only a small part of the overall story. In this paper, we show the potentially helpful effect of not simply human but also social capital and the role of the scientific community in which the scientist finds herself in enabling global collaborations.

Using a robust dataset from South Africa, we advance the literature on international scientific collaborations in countries with smaller and less-developed science systems. Differentiating between the different ways in which foreign ties are acquired allows us to advance a systemic understanding of how science advances. Individual tie acquisition of new foreign collaborators does happen, but it represents only about 20% of all collaborations. A focus on an individual and her attributes, including her international exposure, seems less helpful, given that the great majority of foreign tie acquisitions happen jointly, through repeated ties or even via referrals.

Our work contributes to a better understanding of the systemic functioning of scholarship in developing countries. Our first contribution is the insight that a given scientist is more likely to form ties with foreign co-authors the more her local collaborators already have foreign ties. In other words, the propensity to form foreign ties is not simply a characteristic of the individual but instead is affected by foreign social capital that is collectively possessed. This insight already suggests that the value of foreign scientific collaboration should be understood at the level of the peer community, rather than at the individual level.

Our second contribution is to highlight that this process occurs in the context of local collaboration. Collaborations with local scientists play a strong mediating effect. Thus, the foreign social capital of the peer community translates into foreign tie acquisition when there are strong ties inside the local science system.

This mediating effect is even seen for repeated ties. Practically speaking, a tie is likely to be repeated if a foreign co-author found a collaboration worthwhile enough to work again with the same co-author. Although our evidence shows that similar resources are employed as in the acquisition of unique ties (the focal scholar's own foreign social capital and prior foreign affiliations), the mediating effect of collaborations with local scientists is again positive. This suggests that a culture of collaboration locally is useful not only for the attraction but also for the retention of foreign co-authors.

Our insights have extensive implications. Scholars implicitly acknowledge the importance of the overall system when they point out how systemic challenges can prevent foreign-trained PhDs from doing research in low-income countries, even when those academics wish to do so (Fellesson and Mählck, 2017). Our work suggests the potential gains for development from systemic engagement with the local system.

The term "science system" is hardly contested, but we suggest that more attention should be given to understanding the systemic dimension of knowledge creation behind the technology frontier. Universities throughout the world are engaged in collaboration at the institutional level, often using "cooperation agreements" as part of their advertising copy. Indeed, such international cooperation is part of what makes the "system". Many have observed that ties made by globally mobile individuals enable the globalization of science through their linkages (e.g., Cao *et al.*, 2020). We observe here, though, that much as the ties at the individual level are important, the second-order connections (friends of friends) also create value at individual and wider levels. These second-order connections may be particularly important in still-developing systems, which may not have the resources to support the international (particularly regarding sojourns abroad) ambitions of every scientist, through their international agreements.

Thus, our work provides evidence of the importance of a granular understanding of how individual scientists inside the local system contribute not only to the functioning but also to the internationalization of that system.

To be concrete, our analysis confirms the value of traditional policy measures for internationalization, in particular the hiring of (renowned) foreign scientists and support of foreign stays. What our results point out, though, is that maintaining an inclusive collaboration environment within the local science system is instrumental for foreign tie formation. Foreign ties are rarely formed or kept by a local scientist in isolation. Instead, foreign ties seem to flourish in local environments that are in themselves collaborative. This suggests that more attention should be given to the formation and strengthening of scientists' ties inside the local system, paradoxically, also to strengthen the formation of foreign ties.

The implications of our work for policymakers are clear, albeit somewhat counterintuitive. Provided that there are already some foreign ties, helping scholars to build ties inside the local system and finding ways of fostering collaborations between local scientists do not simply constitute a way of strengthening the local system but also of strengthening the global connectedness of the local science base. By strengthening (collaborations inside) the local science system, policymakers can also increase international connectedness.

This is not to say that international social capital does not matter, on the contrary. International social capital at the level of both the individual and the community is key. But it is activated by local collaborative relationships. This insight suggests an important avenue for future research: to better understand the relative allocation of resources (time, money, and effort) to develop a stock of foreign capital versus to the fostering of local collaborations.

The benefits of internationalization are not in doubt. Science is a global endeavor, and all countries—but especially those with smaller and less-developed science systems—benefit to the extent that they are integrated with that system. But our work also suggests the value of a more thorough engagement with the systemic and system-wide functioning of science, also at the local level, to serve not just the global project, but particularly development.

## References

Arellano, M. and S., Bond (1991), 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, **58**, 277–297.

Ayres, I. and F. E., Vars (2000), 'Determinants of Citations to Articles in Elite Law Reviews', *Journal of Legal Studies*, **29**(S1), 427–450.

Azoulay, P., J. S., Graff Zivin and J., Wang (2010), 'Superstar Extinction', *The Quarterly Journal of Economics*, **125**(2), 549–589.

Barnard, H., R., Cowan and M., Müller (2012), 'Global Excellence at the Expense of Local Diffusion, or a Bridge between Two Worlds? Research in Science and Technology in the Developing World', *Research Policy*, **41**(4), 756–769.

Baruffaldi, A. and F., Poege (2020) 'A Firm Scientific Community: Industry Participation and Knowledge Diffusion,' *IZA Discussion Paper* 13419.

Baum, J. A., R., Cowan and N., Jonard (2010), 'Network-Independent Partner Selection and the Evolution of Innovation Networks', *Management Science*, **56**(11), 2094–2110.

Bordons, M., I., Gomez, M. T., Fernandez, M., Angeles Zulueta and A., Mendez (1996), 'Local, Domestic and International Scientific Collaboration in Biomedical Research', *Scientometrics*, **37**(2), 279–295.

Bozeman, B., J. S., Dietz and M., Gaughan (2001), 'Scientific and Technical Human Capital: An Alternative Model for Research Evaluation', *International Journal of Technology Management*, **22**(7-8), 716–740.

Bramoullé, Y., H., Djebbari and B., Fortin (2009), 'Identification of Peer Effects through Social Networks', *Journal of Econometrics*, **150**(1), 41–55.

Bramoullé, Y., H., Djebbari and B., Fortin (2020), 'Peer Effects in Networks: A Survey', *Annual Review of Economics*, **12**, 603–629.

Cao, C., J., Baas, C. S., Wagner and K., Jonkers (2020), 'Returning Scientists and the Emergence of China's Science System', *Science and Public Policy*, **47**(2), 172–183.

Carayol, N., L., Bergé, L., Cassi and P., Roux (2019), 'Unintended Triadic Closure in Social Networks: The Strategic Formation of Research Collaborations between French Inventors', *Journal of Economic Behavior & Organization*, **163**, 218–238.

Coey, C. (2018), 'International Researcher Mobility and Knowledge Transfer in the Social Sciences and Humanities', *Globalisation, Societies and Education*, **16**(2), 208–223.

Coleman, J. S. (1988), 'Social Capital in the Creation of Human Capital', *American Journal of Sociology*, **94**, 95–120.

Cowan, R. and N., Jonard (2003), 'The Dynamics of Collective Invention', *Journal of Economic Behavior & Organization*, **52**(4), 513–532.

Cowan, R., N., Jonard and J. -B., Zimmermann (2007), 'Bilateral Collaboration and the Emergence of Innovation Networks', *Management Science*, **53**(7), 1051–1067.

Crane, D. (1972) *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, University of Chicago Press, ISBN:0226118576.

Dahlander, L. and D. A., McFarland (2013), 'Ties That Last: Tie Formation and Persistence in Research Collaborations over Time', *Administrative Science Quarterly*, **58**(1), 69–110.

Didegah, F. and M., Thelwall (2013), 'Which Factors Help Authors Produce the Highest Impact Research? Collaboration, Journal and Document Properties', *Journal of Informetrics*, **7**, 861–873.

Essers, D., F., Grigoli and E., Pugacheva (2022), 'Network Effects and Research Collaborations: Evidence from IMF Working Paper Co-authorship', *Scientometrics*, **127**, 7169–7192.

Fafchamps, M., M. J., van der Leij and S., Goyal (2010), 'Matching and Network Effects', *Journal of the European Economic Association*, **8**(1), 203–231.

Fellesson, M. and P., Mählck (2017), 'Untapped Research Capacities? Mobility and Collaboration at the Intersection of International Development Aid and Global Science Regimes', *International Journal of African Higher Education*, **4**(1), 1–24.

Fry, C. V. (2023), 'Bridging the Gap: Evidence from the Return Migration of African Scientists', *Organization Science*, **34**(1), 404–432.

Gibson, J. and D., McKenzie (2014), 'Scientific Mobility and Knowledge Networks in High Emigration Countries: Evidence from the Pacific', *Research Policy*, **43**(9), 1486–1495.

Granovetter, M. (1985), 'Economic Action and Social Structure: The Problem of Embeddedness', *American Journal of Sociology*, **91**(3), 481–510.

Gulati, R. (1995), 'Social Structure and Alliance Formation: A Longitudinal Analysis', *Administrative Science Quarterly*, **40**(4), 619–652.

Holland, P. W. and S., Leinhardt (1976), 'Local Structure in Social Networks', *Sociological Methodology*, **7**, 1–45.

Ismail, A., G., Ansell and H., Barnard (2020), 'Fitting In, Standing Out, and Doing Both: Supporting the Development of a Scholarly Voice', *Journal of Management Education*, **44**(4), 473–489.

Jonkers, K. and L., Cruz-Castro (2013), 'Research upon Return: The Effect of International Mobility on Scientific Ties, Production and Impact', *Research Policy*, **42**(8), 1366–1377.

Judd, C. M. and D. A., Kenny (1981), 'Process Analysis: Estimating Mediation in Treatment Evaluations', *Evaluation Review*, **5**(5), 602–619.

Kwiek, M. (2020a), 'Internationalists and Locals: International Research Collaboration in a Resource-Poor System', *Scientometrics*, **124**, 57–105.

Kwiek, M. (2020b), 'What Large-Scale Publication and Citation Data Tell Us about International Research Collaboration in Europe: Changing National Patterns in Global Contexts', *Studies in Higher Education*, **46**(12), 2629–2649.

Kyvik, S. and I., Reymert (2017), 'Research Collaboration in Groups and Networks: Differences across Academic Fields', *Scientometrics*, **113**, 951–967.

Lane, J. N., I., Ganguli, P., Gaule, E., Guinan and K. R., Lakhani (2021), 'Engineering Serendipity: When Does Knowledge Sharing Lead to Knowledge Production?', *Strategic Management Journal*, **42**(6), 1215–1244.

Li, F. and L., Tang (2019), 'When International Mobility Meets Local Connections: Evidence from China', *Science and Public Policy*, **46**(4), 518–529.

Liu, X., E., Patacchini and Y., Zenou (2014), 'Endogenous Peer Effects: Local Aggregate or Local Average?', *Journal of Economic Behavior & Organization*, **103**, 39–59.

Martin-Sempere, M. J., J., Rey-Rocha and B., Garzon-Garcia (2002), 'The Effect of Team Consolidation on Research Collaboration and Performance of Scientists. Case study of Spanish University Researchers in Geology', *Scientometrics*, **55**(3), 377–394.

Mazzoleni, R. and R. R., Nelson (2007), 'Public Research Institutions and Economic Catch-Up', *Research Policy*, **36**(10), 1512–1528.

Mohnen, M. (2022), 'Stars and Brokers: Knowledge Spillovers Among Medical Scientists', *Management Science*, **68**(4), 2513–2532.

Müller, M., R., Cowan and H., Barnard (2018), 'On the Value of Foreign PhDs in the Developing World: Training versus Selection Effects in the Case of South Africa', *Research Policy*, **47**(5), 886–900.

Nelson, R. R. (2004), 'The Challenge of Building an Effective Innovation System for Catch-Up', *Oxford Development Studies*, **32**(3), 365–374.

Netz, N., S., Hampel and V., Aman (2020), 'What Effects Does International Mobility Have on Scientists' Careers? A Systematic Review', *Research Evaluation*, **29**(3), 327–351.

Nkomo, S. M. (2015), 'Challenges for Management and Business Education in a "Developmental" State: The Case of South Africa', *Academy of Management Learning & Education*, **14**(2), 242–258.

NRF (2005) 'Facts and Figures 2005: The NRF Evaluation and Rating System', https://web.archive.org/web/20060926141616/http://www.nrf.ac.za/evaluation/Content/Documents/Rating/FactsFigures.pdf, accessed 3 March 2022.

Pouris, A. (2007), 'The National Research Foundation's Rating System: Why Scientists Let their Ratings Lapse', *South African Journal of Science*, **103**, 439–441.

Pouris, A. and Y. S., Ho (2014), 'Research Emphasis and Collaboration in Africa', *Scientometrics*, **98**, 2169–2184.

Puuska, H. -M., R., Muhonen and Y., Leino (2014), 'International and Domestic Co-Publishing and their Citation Impact in Different Disciplines', *Scientometrics*, **98**, 823–839.

Rivera-Leon, L. (2021) '*Characterizing Scientific Productivity in Middle Income Economies: an Economics of Science Perspective*, PhD Thesis,' Maastricht University.

Sacerdote, B. (2001), 'Peer Effects with Random Assignment: Results for Dartmouth Roommates', *The Quarterly Journal of Economics*, **116**(2), 681–704.

Scarazzati, S. and L., Wang (2019), 'The Effect of Collaborations on Scientific Research Output: the Case of Nanoscience in Chinese Regions', *Scientometrics*, **121**(2), 839–868.

Scellato, G. C. F. and P., Stephan (2015), 'Migrant Scientists and International Networks', *Research Policy*, **44**, 108–120.

Sooryamoorthy, R. (2019), 'Scientific Knowledge in South Africa: Information Trends, Patterns and Collaboration', *Scientometrics*, **119**, 1365–1386.

Tahamtan, I., A. S., Afshar and K., Ahamdzadeh (2016), 'Factors Affecting Number of Citations: a Comprehensive Review of the Literature', *Scientometrics*, **107**, 1195–1225.

Trippl, M. (2013), 'Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level', *Regional Studies*, **47**(10), 1653–1667.

Turpin, T., R., Woolley, J., Marceau and S., Hill (2008), 'Conduits of Knowledge in the Asia Pacific', *Asian Population Studies*, **4**(3), 247–265.

Yang, X., X., Cai and T., Li (2022), 'Peer Effects of the Young Returnee Scientists: Evidence from the State Key Laboratories in China', *Science and Public Policy*, **49**(5), 739–750.

Ynalvez, M. A. and W. M., Shrum (2011), 'Professional Networks, Scientific Collaboration, and Publication Productivity in Resource-Constrained Research Institutions in a Developing Country', *Research Policy*, **40**(2), 204–216.

Zhao, Q. and J., Guan (2011), 'International Collaboration of Three 'Giants' with the G7 Countries in Emerging Manobiopharmaceuticals', *Scientometrics*, **87**, 159–170.

# Appendix 1

## 1.1 Sample

Social characteristics (for SET in the first part of Table 1) exist for almost the entire dataset (including SSH), with only seven individuals missing information on birth year, gender, skin color, or citizenship. Birth year ranges between 1926 and 1985 with an interquartile range of 1953–1969 and a median of 1961. About one-third of the scientists are female, two-thirds are male. The majority, two-thirds (3037), of scientists is white, according to the apartheid racial categorization.[25] A majority, 80%, has South African citizenship (including permanent residents), only 5% are from the rest of Africa. Most scientists with non-South African citizenship, 10% of the sample, have European citizenship, mostly from the UK followed by Germany.

Information on scientific careers is less complete, with about 11 missing (some) information (see the second part of Table 1). The scientific domain shows twice as many scientists in the sciences (SET) as in social sciences and humanities (SSH). However, for around 10% of the population, we are missing data on academic degrees (bachelor or master as first academic degree and PhD) and employments.[26]

Until 2002, the NRF rated only researchers within the domain of science, engineering, and technologies (SET) where ratings diffused relatively fast and seem to have saturated by 1995

[25] Racism, in particular during the Apartheid regime, produced strong socio-economic differences by color of skin for which we control with this variable.

[26] For individuals where first South African employment is known, we have very complete information on the overall employment trajectory.

(suggested e.g. by NRF, 2005, p.18, Fig.: total number of rated researchers in SET: 1985–2003). In 2002, the rating system was expanded to include social sciences and humanities (SSH). In our dataset, in that first year around 207 SSH researchers obtained their first NRF rating compared to 32 SET researchers. In the subsequent three years, until 2005, further catching-up happens. The ratio of SSH to SET in terms of first ratings then normalizes to around 1:2. On the other hand, the panel that we construct from the rating data displays each year (from 1996 to 2014) that same ratio of 1:2 in terms of active researchers. Thus, the relatively late inclusion of SSH does not lead to a shifting composition of the scientific domain in our panel.

There is no strong pattern of missing information over disciplines, except for arts with 18% missing. Also, foreign and local graduates have similarly few missing observations. Missing information is particularly high for older scientists, born in the 1940s and 1950s, and those scientists with a last rating in the year 2002, 2003, or 2004. It seems possible that in particular older scientists with some reputation managed to sidestep a complete digital filing in the earlier years of the NRF data platform, perhaps with a reference to earlier dossiers in paper format.

All individuals in the sample have (by sample construction) at least one valid rating during the period 1990–2015, with seven being the maximum number of ratings observed for one individual.

For scientific output (peer-reviewed journal and conference papers), the extent of missing information is not immediately clear because no (or low) publication counts are in principle possible. The scientific discipline with the largest number of researchers showing no journal articles is arts (10%) where other outputs may be more relevant. In this discipline, collaboration is probably in general not well proxied by paper co-authors. In all other disciplines, only 1% or less did not file any journal articles. In the natural sciences, nearly all scientists filed peer-reviewed journal articles. In these disciplines, zero publication entries may indeed signal missing data.

### 1.1.1 Network sample

Co-author networks are used not only to count foreign ties of our focal scientists but also to measure the network context in which foreign ties are acquired. Therefore, we aim at an exhaustive description of the network, covering South African scientific collaboration as much as possible, by including in particular papers of rated scientists that are not part of the main sample.

The data creation process itself has implications for network construction. Scientists entered their (past) publications during a period from 2002 and 2014 for various programs run by the NRF. They enter their scientific output up to the time of filing for a rating (or a project) application. Thus, a unique article enters the data twice in case two different scientists file the same, co-authored paper.[27] The potential duplication of co-author links prompts us to work with unweighted, binary networks. On the other hand, papers are missed if the paper is published after the last filing of a scientist. This leads us to drop focal scientists from the panel at the year before their last filing because foreign ties are observed only partly or not at all thereafter. A second effect is a potential measurement error on the collaboration network among South African scientists.

Publication data fields include among other things publication year, publication type, (co-)authors, title, and journal. After some initial cleaning,[28] we obtain 308,412 publication entries with publication dates between 1990 and 2014; 235,712 peer-reviewed journal articles and 72,700 peer-reviewed conference proceedings. The main analysis is based on peer-reviewed journal articles. This avoids duplication of collaboration ties when a conference paper is subsequently published in a journal.

Disambiguation of co-authors is based on information within the NRF dataset. We proceed sequentially. In a first step, the scientist filing a paper is identified by name (surname and initials). In a second step, co-author names are matched to a list of 5,255 scientists rated by NRF between 1984 and 2017. In a third step, remaining names are matched to non-rated NRF users. These

[27] Because there is no unique paper identifier available, we do not know the exact number of unique papers. Titles, journal names, etc. in the data are strings entered by the applicants in various ways and hence cannot serve as identifiers without further cleaning.

[28] The initial cleaning entails essentially two parts. First, joining two NRF data snapshots with different data structures but partly overlapping time periods. Second, identification of individual author names (surname and initials) within each paper's author list as these have been entered by the applicant in a free format string.

include for example South African PhDs but also local and foreign scientists. The sequential matching is based on the following assumption: if there exist two individuals with the same name as one of the authors, the author is more likely (i) the filing scientist rather than the other individual, else (ii) the NRF rated scientists rather than the other individual, else (iii) some scientist dealing with the NRF rather than an individual who does not. In case identification within each of these three steps is ambiguous, we keep the author as unidentified throughout. Finally, for all author names not found in the NRF dataset, a one-to-one correspondence between author name and individual is assumed. True, this final decision may conflate two individuals with the same name. We checked on the whole sample that this issue is not too severe. In addition, we construct networks only within disciplines such that different individuals from different disciplines are not conflated.[29]

Our disambiguation procedure leads to the following numbers: the 235,712 peer-reviewed journal articles give rise to 866,024 author name–paper relations ('author-papers'). Around 25% of author papers are associated with the postulating rated scientist, 15% can be attributed to non-postulating rated scientists, 17% are found among non-rated scientists registered in the NRF system, and for the remaining 41% we take the name as the individual. Around 4% of author papers are ambiguous in that there are multiple individuals in the NRF dataset with the same name and initials. For the main analysis, we remove these ambiguous cases. Sixty-one focal scientists are not associated with any peer-reviewed journal article.[30] This disambiguation of names yields 163,591 authors (individuals).

### 1.2 Basic descriptives

Basic descriptives include statistics on Social Science and Humanities (SSH), but the focus is clearly on Science, Engineering, and Technologies (SET), and ultimately our analysis will be restricted to those domains.

#### 1.2.1 Scientific fields

Table A1 provides the number of panel observations (year-scientists) and individuals (scientists) by scientific discipline along the formation of foreign and SA co-author ties. Within the 'Science and Technology' domain, physicists have on average most collaborations per year (9.25 collaborations per scientist-year) and the highest share of foreign ties (55% of all ties are foreign). For computer scientists, we measure relatively few collaboration ties and a relatively low share of foreign ties. This may be due to the prevalence of conference proceedings over journal articles in the field. Most natural sciences have on average around 30% foreign ties (per scientist-year).

In Social Science and Humanities, Arts is the most foreign-oriented field conditional on its relatively small number of overall co-author ties. Economics and social sciences form most co-author ties per scientist-year with the highest foreign orientation among SSH but still low compared to the foreign tie formation observed in SET fields.

#### 1.2.2 Higher education and the development of foreign social capital

Higher education may be obtained in South Africa, abroad, or both. We classify origin of higher education by the first higher education degree[31] and the PhD. Higher education is denoted "South African" if both the first higher degree and PhD have been obtained in South Africa. It is denoted "foreign" when both degrees have been obtained abroad. The "second-diagonal" cases are "South African-foreign," and "foreign-South African."

The origin of higher education is related to the age of the PhD graduate. In general, foreign-trained PhDs earn their doctorate earlier than locally trained PhDs. Focal scientists with a fully foreign higher education earn their doctorates at a median age of 31, which corresponds to the age of the median US PhD (https://ncses.nsf.gov/pubs/nsf20301/data-tables). Focal scientists fully

[29] This implies that, by construction, we do not observe diffusion of co-author ties across disciplines.

[30] 36 did not postulate an article (mostly in arts). 25 postulated an article that has been removed during the initial cleaning process (e.g., due to missing publication date).

[31] In most cases, the first higher education degree is a bachelor degree or comparable variant; if that is not available, we use the master degree as a proxy.

**Table A1.** Tie formation by scientific fields

| | Observations | Individuals | SA ties | Foreign ties | Total ties | Ties obs. | Foreign ties |
|---|---|---|---|---|---|---|---|
| *Science, Engineering and Technologies (SET)* | | | | | | | |
| Agricultural | 2562 | 251 | 10012 | 3858 | 13870 | 5.41 | 0.28 |
| Biological | 4552 | 417 | 22092 | 10831 | 32923 | 7.23 | 0.33 |
| Chemical | 1692 | 162 | 8310 | 2360 | 10670 | 6.31 | 0.22 |
| Earth and marine | 2402 | 231 | 7799 | 4159 | 11958 | 4.98 | 0.35 |
| Health | 4909 | 488 | 31119 | 14373 | 45492 | 9.27 | 0.32 |
| ICT | 1407 | 150 | 1666 | 432 | 2098 | 1.49 | 0.21 |
| Mathematical | 1899 | 169 | 2616 | 1404 | 4020 | 2.12 | 0.35 |
| Physical | 2237 | 222 | 9246 | 11451 | 20697 | 9.25 | 0.55 |
| Technologies | 3353 | 349 | 9354 | 3068 | 12422 | 3.71 | 0.25 |
| *Social Sciences and Humanities (SSH)* | | | | | | | |
| Arts | 933 | 89 | 195 | 132 | 327 | 0.35 | 0.40 |
| Economic | 1631 | 159 | 2213 | 670 | 2883 | 1.77 | 0.23 |
| Humanities | 4911 | 418 | 2755 | 441 | 3196 | 0.65 | 0.14 |
| Law | 1662 | 150 | 836 | 128 | 964 | 0.58 | 0.13 |
| Social | 3529 | 326 | 4348 | 1197 | 5545 | 1.57 | 0.22 |

Fisher's exact test rejects the Null hypothesis that foreign and SA tie formation is independent of scientific fields at a significance level of below 0.00049 (over all scientific fields, within SET, and within SSH).

formed in South Africa take around 3 years longer (median 34). Scientists with "South African–foreign" higher education graduate with a PhD slightly earlier (median 33), while 'foreign–South African' take three more years (median 37). The interquartile range for SA formed scientists is 30–38 years and shifted down by two to four years for fully trained foreign. We observe that same tendency across scientific fields with some smaller variation. In general, age at PhD is higher for SSH than for SET.

Time to first SA employment after PhD, and hence experience gathered elsewhere, also varies by origin of higher education. 90% of focal scientists with a PhD earned in South Africa start an SA employment within 5 years from the doctorate. 80% with 'South African–foreign' education 'return home' for SA employment within 5 years. On the other hand, only 50% of foreign-educated focal scientists entered the SA higher education system within 5 years after their PhD. Most fully foreign-trained scientists enter between 2 and 10 years after PhD (interquartile range), with 7 years of experience on average. On the other hand, 75% of SA scientists start first employment after PhD within 2 years from graduation (4 years for 'South African–foreign'). This pattern, again, holds across scientific disciplines.

Table A2 looks at how foreign ties are acquired, from time of entry into the SA system (i.e., $t \geq 0$), by origin of education, for the SET and SSH subsamples, respectively. First consider the absolute numbers, which are average number of ties (of a certain kind) over all scientist-year observations from time of entry into the SA system (i.e., $t \geq 0$). For foreign-trained scientists, we see more foreign ties than for the other groups. That holds for the total number of foreign ties (last column), the different ways of acquisitions (the other columns), and for both subsamples (SET and SSH).

Furthermore, patterns of foreign tie formation differ by origin of higher education. Consider the proportions of tie formation motifs given in the brackets (Table A2). Foreign-trained scientists seem more "independent" in foreign tie formation as they establish relatively often unique ties to new foreign co-authors (and less joint acquisitions) and are more likely to repeat existing ties (and use to a lesser extent referrals within SA).

## 1.3. Extensions

This section provides the statistics underlying the figures in Section 4.2

**Table A2.** Motifs of tie formation by higher education for subsamples SET and SSH, average (proportion)

| | new unique | new joint | repeated | referral | others | foreign ties |
|---|---|---|---|---|---|---|
| *Science, Engineering and Technologies (SET)* | | | | | | |
| SA | 0.26 (0.183) | 0.684 (0.480) | 0.378 (0.265) | 0.063 (0.044) | 0.039 (0.028) | 1.425 (1) |
| For | 0.622 (0.231) | 0.908 (0.337) | 0.966 (0.359) | 0.092 (0.034) | 0.102 (0.038) | 2.69 (1) |
| SA-For | 0.312 (0.205) | 0.639 (0.419) | 0.47 (0.308) | 0.066 (0.043) | 0.037 (0.024) | 1.525 (1) |
| For-SA | 0.241 (0.161) | 0.765 (0.51) | 0.371 (0.247) | 0.077 (0.052) | 0.045 (0.03) | 1.5 (1) |
| *Social Sciences and Humanities (SSH)* | | | | | | |
| SA | 0.038 (0.209) | 0.094 (0.515) | 0.047 (0.257) | 0.002 (0.012) | 0.001 (0.007) | 0.183 (1) |
| For | 0.081 (0.310) | 0.114 (0.438) | 0.065 (0.249) | 0.001 (0.003) | 0 (0) | 0.261 (1) |
| SA-For | 0.045 (0.252) | 0.087 (0.486) | 0.042 (0.236) | 0.002 (0.013) | 0.002 (0.013) | 0.179 (1) |
| For-SA | 0.096 (0.303) | 0.163 (0.514) | 0.047 (0.149) | 0.009 (0.029) | 0.002 (0.005) | 0.317 (1) |

**Table A3.** Higher education origin (ind., obs.), outcome (freq.), and model statistics (m1, m2, sargan, moran's I diff., moran's I level) of estimations shown in Figure 3.

| Model 1 | | | | | Model 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| m1 | m2 | sargan | moran diff | moran level | m1 | m2 | sargan | moran diff | moran level |
| *South African HE (961, 10180), for. ties (13474)* | | | | | | | | | |
| <0.001 | 0.198 | 0.286 | <0.001 | <0.001 | <0.001 | 0.095 | 0.796 | 0.003 | 0.005 |
| *South African HE (961, 10180), unique acq. (2401)* | | | | | | | | | |
| <0.001 | 0.043 | 0.846 | 0.793 | 0.511 | <0.001 | 0.044 | 0.84 | 0.77 | 0.563 |
| *South African HE (961, 10180), joint acq. (6542)* | | | | | | | | | |
| <0.001 | 0.117 | 0.164 | <0.001 | <0.001 | <0.001 | 0.369 | 0.76 | <0.001 | <0.001 |
| *South African HE (961, 10180), repeated (3524)* | | | | | | | | | |
| <0.001 | 0.688 | 0.085 | <0.001 | <0.001 | <0.001 | 0.744 | 0.077 | <0.001 | <0.001 |
| *South African HE (961, 10180), co-auth. ref. (524)* | | | | | | | | | |
| <0.001 | 0.052 | 0.167 | <0.001 | <0.001 | <0.001 | 0.052 | 0.132 | <0.001 | <0.001 |
| *South African HE (961, 10180), faculty ref. (628)* | | | | | | | | | |
| <0.001 | 0.04 | 0.31 | <0.001 | <0.001 | <0.001 | 0.04 | 0.184 | <0.001 | <0.001 |
| *South African HE (961, 10180), other (379)* | | | | | | | | | |
| <0.001 | 0.262 | 0.316 | <0.001 | <0.001 | <0.001 | 0.263 | 0.172 | <0.001 | <0.001 |
| *foreign HE (333, 3345), for. ties (7953)* | | | | | | | | | |
| <0.001 | 0.545 | 0.485 | 0.217 | 0.011 | <0.001 | 0.925 | 0.032 | 0.802 | 0.057 |
| *foreign HE (333, 3345), unique acq. (1795)* | | | | | | | | | |
| <0.001 | 0.879 | 0.78 | 0.985 | 0.981 | <0.001 | 0.975 | 0.776 | 0.96 | 0.899 |
| *foreign HE (333, 3345), joint acq. (2733)* | | | | | | | | | |
| <0.001 | 0.3 | 0.907 | <0.001 | <0.001 | <0.001 | 0.225 | 0.952 | <0.001 | <0.001 |
| *foreign HE (333, 3345), repeated (2814)* | | | | | | | | | |
| <0.001 | 0.677 | 0.193 | 0.953 | 0.029 | <0.001 | 0.485 | 0.114 | 0.86 | 0.226 |
| *foreign HE (333, 3345), co-auth. ref. (263)* | | | | | | | | | |
| <0.001 | 0.129 | 0.728 | <0.001 | <0.001 | <0.001 | 0.148 | 0.896 | <0.001 | <0.001 |
| *foreign HE (333, 3345), faculty ref. (301)* | | | | | | | | | |
| <0.001 | 0.191 | 0.831 | <0.001 | <0.001 | <0.001 | 0.173 | 0.856 | <0.001 | <0.001 |
| *foreign HE (333, 3345), other (310)* | | | | | | | | | |
| <0.001 | 0.262 | 0.816 | 0.704 | 0.975 | <0.001 | 0.275 | 0.793 | 0.702 | 0.98 |

**Table A4.** Scientific field (ind., obs.) and outcome (freq.) and model statistics (m1, m2, sargan, moran's I diff., moran's I level) of estimations shown in Figure 4

| Model 1 | | | | | Model 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| m1 | m2 | sargan | moran diff | moran level | m1 | m2 | sargan | moran diff | moran level |
| *Biological sciences (288, 2890), for. ties (5767)* | | | | | | | | | |
| < 0.001 | 0.402 | 0.664 | <0.001 | <0.001 | <0.001 | 0.875 | 0.525 | 0.3 | 0.106 |

(continued)

**Table A4.** (Continued)

| Model 1 | | | | | Model 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| m1 | m2 | sargan | moran diff | moran level | m1 | m2 | sargan | moran diff | moran level |
| *Biological sciences (288, 2890), unique acq. (1334)* | | | | | | | | | |
| <0.001 | 0.906 | 0.733 | 0.999 | 0.834 | <0.001 | 0.922 | 0.613 | 1 | 0.832 |
| *Biological sciences (288, 2890), joint acq. (2656)* | | | | | | | | | |
| <0.001 | 0.943 | 0.789 | <0.001 | <0.001 | <0.001 | 0.654 | 0.86 | 0.002 | 0.002 |
| *Biological sciences (288, 2890), repeated (1360)* | | | | | | | | | |
| <0.001 | 0.599 | 0.348 | 0.002 | <0.001 | <0.001 | 0.195 | 0.39 | 0.05 | 0.004 |
| *Biological sciences (288, 2890), co-auth. ref. (184)* | | | | | | | | | |
| <0.001 | 0.074 | 0.543 | <0.001 | <0.001 | <0.001 | 0.032 | 0.483 | <0.001 | <0.001 |
| *Biological sciences (288, 2890), faculty ref. (231)* | | | | | | | | | |
| <0.001 | 0.321 | 0.263 | <0.001 | <0.001 | <0.001 | 0.081 | 0.471 | <0.001 | <0.001 |
| *Biological sciences (288, 2890), other (186)* | | | | | | | | | |
| <0.001 | 0.671 | 0.291 | 0.069 | 0.032 | <0.001 | 0.629 | 0.303 | 0.082 | 0.071 |
| *Earth and marine sciences (143, 1448), for. ties (2184)* | | | | | | | | | |
| <0.001 | 0.75 | 0.609 | 0.025 | 0.001 | <0.001 | 0.8 | 0.853 | 0.978 | 0.864 |
| *Earth and marine sciences (143, 1448), unique acq. (510)* | | | | | | | | | |
| <0.001 | 0.72 | 0.625 | 1 | 0.753 | <0.001 | 0.759 | 0.598 | 0.995 | 0.66 |
| *Earth and marine sciences (143, 1448), joint acq. (947)* | | | | | | | | | |
| <0.001 | 0.413 | 0.427 | <0.001 | <0.001 | <0.001 | 0.771 | 0.273 | 0.016 | 0.015 |
| *Earth and marine sciences (143, 1448), repeated (590)* | | | | | | | | | |
| <0.001 | 0.575 | 0.108 | 0.157 | 0.018 | <0.001 | 0.657 | 0.161 | 0.226 | 0.057 |
| *Earth and marine sciences (143, 1448), co-auth. ref. (67)* | | | | | | | | | |
| <0.001 | 0.513 | 0.791 | 1 | 0.315 | <0.001 | 0.48 | 0.674 | 1 | 0.463 |
| *Earth and marine sciences (143, 1448), faculty ref. (80)* | | | | | | | | | |
| <0.001 | 0.437 | 0.762 | 0.999 | 0.28 | <0.001 | 0.382 | 0.707 | 1 | 0.46 |
| *Earth and marine sciences (143, 1448), other (57)* | | | | | | | | | |
| 0.004 | 0.778 | 0.658 | 0.011 | 0.001 | 0.003 | 0.758 | 0.907 | 0.003 | 0.001 |
| *Health Sciences (296, 3012), for. ties (6266)* | | | | | | | | | |
| <0.001 | 0.123 | 0.029 | <0.001 | <0.001 | <0.001 | 0.265 | 0.263 | 0.089 | 0.273 |
| *Health Sciences (296, 3012), unique acq. (928)* | | | | | | | | | |
| <0.001 | 0.121 | 0.686 | 0.299 | 0.048 | <0.001 | 0.124 | 0.634 | 0.347 | 0.036 |
| *Health Sciences (296, 3012), joint acq. (3250)* | | | | | | | | | |
| <0.001 | 0.189 | 0.24 | <0.001 | <0.001 | <0.001 | 0.81 | 0.648 | <0.001 | <0.001 |
| *Health Sciences (296, 3012), repeated (1589)* | | | | | | | | | |
| <0.001 | 0.517 | 0.007 | <0.001 | <0.001 | <0.001 | 0.547 | 0.01 | 0.002 | <0.001 |
| *Health Sciences (296, 3012), co-auth. ref. (232)* | | | | | | | | | |
| <0.001 | 0.1 | 0.289 | 0.002 | <0.001 | <0.001 | 0.099 | 0.226 | 0.003 | <0.001 |
| *Health Sciences (296, 3012), faculty ref. (286)* | | | | | | | | | |
| <0.001 | 0.084 | 0.276 | 0.003 | <0.001 | <0.001 | 0.083 | 0.181 | 0.004 | <0.001 |
| *Health Sciences (296, 3012), other (213)* | | | | | | | | | |
| <0.001 | 0.679 | 0.59 | 0.015 | <0.001 | <0.001 | 0.574 | 0.482 | 0.02 | 0.002 |
| *Physical sciences (107, 1152), for. ties (4444)* | | | | | | | | | |
| <0.001 | 0.616 | 0.776 | 0.057 | 0.004 | <0.001 | 0.948 | 0.619 | 0.589 | 0.015 |
| *Physical sciences (107, 1152), unique acq. (650)* | | | | | | | | | |
| <0.001 | 0.326 | 0.619 | 0.951 | 0.152 | <0.001 | 0.329 | 0.586 | 0.943 | 0.109 |
| *Physical sciences (107, 1152), joint acq. (1482)* | | | | | | | | | |
| <0.001 | 0.878 | 0.966 | <0.001 | 0.001 | <0.001 | 0.948 | 0.663 | 0.004 | 0.001 |
| *Physical sciences (107, 1152), repeated (1776)* | | | | | | | | | |
| <0.001 | 0.726 | 0.461 | 0.159 | 0.002 | <0.001 | 0.755 | 0.28 | 0.281 | 0.072 |
| *Physical sciences (107, 1152), co-auth. ref. (276)* | | | | | | | | | |
| 0.003 | 0.635 | 0.331 | <0.001 | <0.001 | 0.004 | 0.654 | 0.422 | <0.001 | <0.001 |
| *Physical sciences (107, 1152), faculty ref. (299)* | | | | | | | | | |
| 0.001 | 0.603 | 0.499 | <0.001 | <0.001 | 0.001 | 0.59 | 0.657 | 0.001 | 0.001 |

(continued)

**Table A4.** (Continued)

| Model 1 | | | | | Model 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| m1 | m2 | sargan | moran diff | moran level | m1 | m2 | sargan | moran diff | moran level |
| *Physical sciences (107, 1152), other (237)* | | | | | | | | | |
| <0.001 | 0.134 | 0.381 | 0.559 | 0.877 | <0.001 | 0.09 | 0.359 | 0.645 | 0.823 |
| *Technologies and applied sciences (228, 2350), for. ties (1777)* | | | | | | | | | |
| <0.001 | 0.098 | 0.747 | <0.001 | <0.001 | <0.001 | 0.226 | 0.812 | 0.064 | 0.004 |
| *Technologies and applied sciences (228, 2350), unique acq. (350)* | | | | | | | | | |
| <0.001 | 0.727 | 0.371 | 0.258 | 0.109 | <0.001 | 0.661 | 0.461 | 0.191 | 0.062 |
| *Technologies and applied sciences (228, 2350), joint acq. (859)* | | | | | | | | | |
| <0.001 | 0.142 | 0.147 | <0.001 | <0.001 | <0.001 | 0.949 | 0.699 | 0.006 | <0.001 |
| *Technologies and applied sciences (228, 2350), repeated (522)* | | | | | | | | | |
| <0.001 | 0.588 | 0.428 | <0.001 | <0.001 | <0.001 | 0.417 | 0.251 | 0.008 | <0.001 |
| *Technologies and applied sciences (228, 2350), co-auth. ref. (12)* | | | | | | | | | |
| 0.024 | 0.84 | 1 | 0.407 | 0.904 | 0.024 | 0.822 | 1 | 0.861 | 0.96 |
| *Technologies and applied sciences (228, 2350), faculty ref. (20)* | | | | | | | | | |
| 0.001 | 0.844 | 0.978 | 0.993 | 0.94 | 0.001 | 0.842 | 0.999 | 0.999 | 0.98 |
| *Technologies and applied sciences (228, 2350), other (26)* | | | | | | | | | |
| <0.001 | 0.391 | 0.874 | 0.056 | 0.057 | <0.001 | 0.443 | 0.994 | 0.063 | 0.042 |
| *SSH (869, 9380), for. ties (1798)* | | | | | | | | | |
| <0.001 | 0.012 | 0.863 | <0.001 | <0.001 | <0.001 | 0.246 | 0.663 | 0.032 | 0.001 |
| *SSH (869, 9380), unique acq. (453)* | | | | | | | | | |
| <0.001 | 0.366 | 0.424 | 0.019 | 0.005 | <0.001 | 0.365 | 0.483 | 0.022 | 0.004 |
| *SSH (869, 9380), joint acq. (848)* | | | | | | | | | |
| <0.001 | 0.034 | 0.369 | <0.001 | <0.001 | <0.001 | 0.577 | 0.205 | <0.001 | 0.001 |
| *SSH (869, 9380), repeated (463)* | | | | | | | | | |
| 0.001 | 0.506 | 0.972 | 0.586 | 0.415 | < 0.001 | 0.43 | 0.919 | 0.814 | 0.699 |
| *SSH (869, 9380), co-auth. ref. (12)* | | | | | | | | | |
| 0.011 | 0.32 | 0.998 | 0.925 | 1 | 0.01 | 0.298 | 0.967 | 0.995 | 1 |
| *SSH (869, 9380), faculty ref. (23)* | | | | | | | | | |
| 0.001 | 0.322 | 0.991 | 0.973 | 1 | 0.001 | 0.295 | 0.976 | 0.985 | 1 |
| *SSH (869, 9380), other (11)* | | | | | | | | | |
| 0.027 | 0.317 | 0.986 | 0.742 | 0.764 | 0.027 | 0.594 | 1 | 0.977 | 0.919 |