Scientometrics (2019) 118:1–19
https://doi.org/10.1007/s11192-018-2954-9

Identification of research communities in cited and uncited publications using a co-authorship network

Zewen Hu[1,2] · **Angela Lin**[2] · **Peter Willett**[2]

Received: 30 September 2017 / Published online: 10 November 2018

Abstract
Patterns of co-authorship provide an effective means of probing the structures of research communities. In this paper, we use the CiteSpace social network tool and co-authorship data from the Web of Science to analyse two such types of community. The first type is based on the cited publications of a group of highly productive authors in a particular discipline, and the second on the uncited publications of those highly productive authors. These pairs of communities were generated for three different countries—the People's Republic of China (PRC), the United Kingdom (UK) and the United States of America (USA)—and for four different disciplines (as denoted by Web of Science subject categories)—Chemistry Organic, Engineering Environmental, Economics, and Management. In the case of the UK and USA, the structures of the cited and uncited communities in each of the four disciplines were markedly different from each other; in the case of the PRC, conversely, the cited and uncited PRC communities had broadly similar structures that were characterised by large groups of connected authors. We suggest that this may arise from a greater degree of guest or honorary authorship in the PRC than in the UK or the USA.

Keywords Uncited publications · Co-authorship · Honorary authorship · Social network analysis · Collaborative pattern · Research community

Introduction

There has been much research over the years into the factors that affect the number of citations that a published article attracts. Typical factors include the number of authors (Smart and Bayer 1986), the journal impact factor (Elkins et al. 2010), whether international collaboration is involved (Inzelt et al. 2009; Prathap 2013), whether a journal is open access (Craig et al. 2007), author reputations (Makino 1998) etc. Recent examples of such studies include those by Thelwall and Wilson (2014), Onodera and Yoshikane (2015) and by Tahamtan et al. (2016), whose detailed review identified no fewer than 28 factors in three

✉ Zewen Hu
huzewen915@163.com

[1] School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, People's Republic of China

[2] Information School, University of Sheffield, Sheffield S10 2TN, UK

broad categories, these relating to the article itself, to the article's author(s), and to the journal in which the article was published.

Such studies have focused on cited articles but there has also been some interest in uncitedness, i.e., the study of articles that never attract any citations (Hu and Wu 2014; Hu et al. 2018; Schwartz 1997; Zhao 2015). There are many reasons why this might be the case for a particular publication: it may be of low quality, be poorly written or difficult to understand, be published in an inappropriate journal, be valuable but undiscovered or forgotten. That said, MacRoberts and MacRoberts (2010) have suggested that even uncited publications may influence subsequent research in a range of ways. There have been several studies of the statistical characteristics of uncited publications, e.g. (Burrell 2013; Egghe 2010; Liang et al. 2015). Lou and He (2016) reported a correlation between uncitedness and author affiliations. Stern (1990) analysed the bibliographic characteristics of uncited biomedical papers and found that the number of authors had a smaller influence on whether articles were cited. Hu and Wu (2018) have recently reported the reasons for citing or not citing a publication, including 'prestigious authors', and 'academic tastes and interests similar to citers' that can drive authors to form large research communities. In this paper we investigate differences of research communities between cited and uncited articles from a novel perspective, that of using social network analysis (SNA).

SNA is being increasingly used to study bibliometric phenomena, involving the co-occurrences in publications of authors, topics, institutions and countries, and similarly for the co-citations of same; here, we focus on co-authorship. Co-authorship of a publication can be thought of as documenting a collaboration between two or more authors, and these collaborations form a network in which the nodes represent authors and the edges between pairs of authors denote a joint publication. The use of such networks to reveal the structures of academic communities was pioneered by Newman (2001a, b, 2004). For example, a co-authorship study of publications from biology, physics, and mathematics, identified a single large connected group of authors in each case, representing 82–92% of the total number of author nodes. Newman (2004) noted that this suggested that a large fraction of each of these research communities could be regarded as working in what he described as a "linked research enterprise", with the remaining author nodes forming a large number of much smaller connected components.

Furthermore, research collaboration in groups has been identified as important factor to drive productivity and enhance citation of research (Smart and Bayer 1986; Cohen 1991; Etzkowitz 1992). These factors have been examined by a large number of follow-up studies: size and productivity (Seglen and Aksnes 2000; Guimerà et al. 2005; Maaike et al. 2015), the difference in cooperation within research groups (Adams 2012; Kyvik et al. 2017), the role of the leader within groups (Pudovkin et al. 2012). Glanzel (2002) showed that the relationship between collaboration and the quality of the research varies across different subject fields. The evidence from the patterns and structure of social science co-authorship networks showed that authors with many collaborators and high scientific prestige gain more connections from newcomers than do their colleagues (Moody 2004). From studying the structure of co-authorship networks in four scientific disciplines (Physics, Mathematics, Biotechnology and Sociology) in four 5-year periods (1986–2005), Kronegger et al. (2011) revealed that, regardless of the research discipline, the co-authorship structure very quickly consolidates into a multi-core, semi-periphery to periphery structure.

There is subsequently a large, and growing, literature associated with the analysis of co-authorship networks, as exemplified in work of the patterns and structure of the community of researchers from the field of mathematical research (Grossman 2002), computer science (Izquierdo et al. 2016; Franceschet 2011), and the measure of authors' centrality

in co-authorship networks (Lu and Feng 2009), the detection and identification of research groups in co-authorship networks (Calero et al. 2006; Perianes-Rodriguez et al. 2010; Wang et al. 2012, 2013), the relations between research groups and quality of research through co-authorship analysis (Yan and Ding 2009; Abbasi et al. 2011; Reyes-Gonzalez et al. 2016), the stability of co-authorship structures (Cugmas et al. 2016), as well as scientific collaborations between China and the European Union reflected by co-authorship structures (Wang and Wang 2017; Wang et al. 2017), inter alia.

However, their studies mainly focus on the structure of research communities from the angle of unclassified publications that might not see into the mutual relations between the structure of research community and quality of research. This paper focuses on both cited and uncited publications, and uses analysis of co-authorship networks to look into the difference of research communities between two types of publications across different disciplines and countries. So, a SNA tool, CiteSpace (Chen 2004, 2006; Chen et al. 2010) has been used to address two principal research questions. First, are there significant differences between networks based on cited and uncited articles; second, are there significant differences between networks based on work conducted in different countries [specifically the People's Republic of China (PRC), the United Kingdom (UK), and the United States of America (USA)].

Experimental details

The basic data for the study were obtained from the Web of Science (hereafter WoS) Core Collection in December 2016. Four different WoS subject categories were selected that contained a sufficient number of both cited and uncited publications in all of the three chosen countries for subsequent analysis (where the selected publications comprised articles, reviews and proceedings papers). These categories were Chemistry Organic, Engineering Environmental, Economics, and Management that respectively belong to three broadly but fundamentally different disciplines, i.e., one physical science, one engineering and two social science disciplines. The differences in the three chosen disciplines might help to provide persuasive evidence for the difference in the structure of research communities across different disciplines. The three countries chosen were the UK and the USA, which have for long been two of the most productive academic nations, and the PRC, which is a nation that has risen to academic prominence over the last decade or so.

For each category, all publications for the period 2000–2014 from the three chosen countries were identified, together with the citations to those publications for the period 2000–2016. The resulting WoS dataset contained 109,843 publications from Chemistry Organic, 78,738 from Engineering Environmental, 151,827 from Economics and 112,371 from Management, as detailed in Table 1. For example, 47,504 of the Management publications had USA authors, of which 37,723 were cited and the remaining 9781 were uncited.

From these publications, we identified the most productive authors for each combination of subject and country. For each year of publication (2000, 2001 etc.), the 50 most productive authors were identified (and also the 51st, 52nd etc. if they had the same number of publications in a given year as the 50th most productive author for that year), subject to them having at least two publications in the chosen year. So, for example, in 2001 the most productive USA author from cited publications in Economics was Acemoglu D with 8 publications, then Kleit AN with 7 publications and so on. Furthermore, name ambiguity might affect the quality of research results to a certain extent. So, we checked name of each

Table 1 Numbers of publications for the period 2000–2014, and citations to those for the period 2000–2016

Subject category	Total	USA		PRC		UK	
		Cited	Uncited	Cited	Uncited	Cited	Uncited
Chemistry organic	109,843	54,158	1434	35,354	2097	16,288	512
Engineering environmental	78,738	29,142	4978	20,056	16,451	6885	1226
Economics	151,827	72,974	17,293	21,520	7294	25,438	7308
Management	112,371	37,723	9781	7093	38,917	14,219	4638

author in all co-authorship networks and merged the repeated authors through checking whether two authors own same affiliation. As an illustration, two authors–Allenby B and Allenby BR from uncited co-authorship network in Engineering Environmental of USA, own the same affiliation–department of Civil, Environmental and Sustainable Engineering, Arizona State University, Tempe, USA. Then we merged two authors in co-authorship network through the merging function named the Alias List in CiteSpace III.

The CiteSpace III system (available from http://cluster.cis.drexel.edu/~cchen/citespace/) was then used to generate two co-authorship networks for each combination of subject and country: one based on the cited publications for these productive authors and the other based on their uncited publications. The connected components in the resulting networks then corresponded to research communities (Girvan and Newman 2002), comprising groups of authors who are linked either directly (e.g., author A has published jointly with author B) or indirectly (e.g., authors A and C are linked if A has published jointly with B, who has also published jointly with author C).

The incidence of uncited publications

Citation studies often pay little attention to uncited publications. However, since one of the principal foci of this study is the difference between cited and uncited materials, we start with a brief review of the characteristics of the uncited publications in our WoS sample. Table 2 lists the percentages of uncited publications for each combination of subject and country, e.g., for the 2002 Chemistry Organic publications, 1.3% of the American, 4.9% of the Chinese and 1.9% of the British publications attracted no citations at all in the period 2002–2016. The bottom two rows of the table list the mean and standard deviation over the 15 years 2000–2014.

We can draw two principal conclusions from this table. First, far fewer Chemistry Organic publications remain uncited than is the case for the other three subject categories, where the uncited percentage can be as much as an order of magnitude greater. The Chemistry Organic uncited rates range from 1.0% (2003 USA publications) to 10.2% (2014 PRC publications), but even the latter figure is notably less than the figure of 18.8% quoted by Hamilton (Hamilton 1991) for 1984 Chemistry Organic publications that were uncited in the period 1984–1988. In addition to the mean values, the year-to-year fluctuations (as quantified by the standard deviations) are also the smallest for this subject.

Second, the uncited percentages for PRC publications are consistently larger than for USA and UK publications (where the mean uncited percentages are very similar for all four disciplines). Indeed, for the 60 combinations (15 years of publication and four subject categories), the Chinese uncited percentage is the largest for all but three of them

Table 2 Percentages of uncited articles in each year of publication for each combination of subject and country

Year	Chemistry organic			Engineering environmental			Economics			Management		
	USA	PRC	UK	USA	PRC	UK	USA	PRC	UK	USA	PRC	UK
2000	1.7	8.0	1.2	28.2	14.9	20.6	11.2	12.2	11.2	21.3	32.8	9.2
2001	1.7	6.9	3.8	22.9	48.6	20.3	9.6	29.1	8.4	10.3	44.7	4.4
2002	1.3	4.9	1.9	19.5	6.4	15.4	8.1	73.6	8.7	19.0	58.1	15.5
2003	1.0	5.6	1.3	19.7	45.8	26.6	11.8	74.5	15.3	11.5	46.5	5.3
2004	1.2	4.5	1.9	12.9	53.0	14.5	9.3	71.0	9.9	11.3	81.4	9.6
2005	2.0	3.0	1.9	12.0	44.3	11.9	12.3	61.0	10.6	11.3	87.7	11.3
2006	1.6	4.0	2.5	11.8	37.6	7.2	15.0	73.5	19.7	14.2	89.1	18.9
2007	2.7	3.8	3.1	11.8	47.3	19.3	18.7	87.5	20.0	20.2	90.8	24.1
2008	1.5	4.7	2.0	8.9	49.2	12.8	18.5	80.0	22.4	18.3	85.9	26.2
2009	2.3	4.1	2.3	16.1	44.5	12.0	20.9	83.4	25.8	18.8	88.6	27.8
2010	2.4	4.2	3.6	14.3	36.4	21.2	20.5	78.1	20.7	22.4	88.4	27.6
2011	2.1	4.7	4.9	8.8	32.9	5.8	21.3	63.0	28.2	23.7	82.8	34.4
2012	3.7	4.6	3.6	9.4	57.2	5.5	24.1	67.1	28.5	21.5	75.8	26.7
2013	5.5	7.0	6.0	9.6	51.6	12.0	27.0	67.9	30.4	27.1	80.4	33.0
2014	9.5	10.2	8.0	14.8	33.4	18.7	35.2	63.1	36.2	36.5	77.5	35.7
Mean	2.7	5.4	3.2	14.7	40.2	14.9	17.6	65.6	19.7	19.2	74.0	20.7
SD	0.02	0.02	0.02	0.06	0.14	0.06	0.08	0.20	0.09	0.07	0.19	0.11

(Chemistry Organic in 2011 and Engineering Environmental in 2000 and 2002). Moreover, the differences are very substantial: between two and three times in the mean values for all but Chemistry Organic. This is especially so in the two social science subjects where the uncited percentages are in excess of 60% for all but the very earliest years (specifically, 2000–2001 for Economics and 2000–2003 for Management). Thus, while citations to Chinese research in general are increasing in many subject fields (Tang et al. 2015; Xie et al. 2014), there are still very large numbers that fail to achieve any form of impact in the shape of WoS citations. At least in the four disciplines considered here, this occurs far less for the USA and UK publications, thus mirroring the more general pattern of low Chinese impact noted by Radosevic and Yoruk (2014) and by Zhu et al. (2014).

Co-authorship patterns in cited and uncited publications

The co-authorship networks for the uncited and cited USA publications are shown in Figs. 1 and 2 respectively, the four diagrams in each case corresponding to the Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right) datasets for the most productive authors. Each connected component in the network can be considered as a research community that comprises individual productive authors who are linked together by the fact that they have co-authored papers with other authors in the network. The numbers and sizes of the connected components are shown in Tables 3 and 4, these corresponding to the networks in Figs. 1 and 2, respectively.

Fig. 1 The co-authorship mapping from uncited USA publications for Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right)

Inspection of the four networks in Fig. 1 show that they all exhibit a comparable structure in which there are large numbers of small families (the largest being a single Engineering Environmental grouping containing 24 members). For example, the network based on the uncited USA Chemistry Organic publications contained 158 productive author nodes and 118 co-authorship links: these yielded a total of 36 families containing between 2 and 7 individuals, with the remaining 47 productive authors unlinked. The ratio of the number of links to the number of nodes is 0.75, with the corresponding figures for Economics and Management being 0.63 and 0.53, respectively; that for Engineering Environmental is 1.05, representing a more highly linked structure than for the other three disciplines. However, it will be clear that in all networks, the great majority of the families consist of just two or three authors. Figure 2 and Table 4 describe the corresponding networks based on the cited publications. It will be seen that the structures here are very different: while there are again many small, two-or three-membered communities, all four networks are dominated by a single, enormous community, a sort of extended research family: that for Economics contains 97 authors while those for the other three disciplines all contain over 200 authors.

The corresponding sets of networks for PRC and UK publications are shown in Figs. 3 and 4 (uncited publications) and Figs. 5 and 6 (cited publications), respectively. The eight

Fig. 2 The co-authorship mapping from cited USA publications for Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right)

UK networks are similar to those for the USA in that the uncited ones (Fig. 4) again consist of many small communities while the four cited ones (Fig. 6) are each dominated by a single, extended research family. This is also the case with the four networks based on the Chinese cited publications in Fig. 5. Where the Chinese networks differ are when the uncited publications are considered (in Fig. 3) since these all have the single, very large community that otherwise characterises the various cited-publication networks.

To save space, we have not included full tables analogous to Tables 3 and 4 here. Instead, the data for all of the networks is summarised in Table 5. This table organizes the communities into four sizes: those with ≥200 authors, with 20–100 authors, with 2–19 authors, and singletons; a blank cell denotes no communities within a given range of sizes. For example, if we read across the row for ≥200 Engineering Environmental authors, the Chinese uncited network has one community containing 450 authors (this was indeed the largest community identified in any of the analyses), and the USA, Chinese and UK cited networks each has one family containing 225, 365 and 318 authors, respectively. This table highlights the fact that the Chinese uncited networks are totally different in structure from

Table 3 The sizes and numbers of research communities in four subject categories from Fig. 1

Family size	Research communities from uncited USA publications			
	Chemistry organic	Engineering environmental	Economics	Management
24		1		
16			1	
15		2		
14		1		
13		1		
12			1	
10		3		2
8		3	1	
7	1	4		
6		9	1	2
5	5	12	3	5
4	6	31	12	10
3	7	45	23	35
2	17	98	91	115
1	50	250	349	358

Table 4 The sizes and numbers of research communities in four subject categories from Fig. 2

Family size	Research communities from cited USA publications			
	Chemistry organic	Engineering environmental	Economics	Management
280				1
225		1		
217	1			
97			1	
14		1		
13			1	
12	1			
11	1			
10	1			
9			1	
8		1	2	
7		1	1	2
6	1	1		
5		3	1	3
4	2	4	7	11
3	6	9	17	20
2	14	22	71	76
1	103	155	430	380

Fig. 3 The co-authorship mapping from uncited Chinese publications for Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right)

the USA and UK uncited networks, whereas the three sets of cited networks are comparable in structure. It also makes clear that, in most cases, the majority of the authors are in small or singleton clusters, as would be expected given the power law behaviour of co-authorship networks first noted by Newman (Newman 2001b, 2004). A χ^2 analysis of the data in Table 5 is detailed in Table 6, each cell of which contains the χ^2 value and the associated probability of a significant difference ($\nu = 3$) between the distributions of family sizes for the uncited and cited networks. There is no significant difference, at the 0.05 level of statistical significance, for two of the Chinese datasets, those for Engineering Environmental and Management, and the difference is only marginal ($p = 0.049$) for Economics. That is completely different from all USA and UK datasets.

The structural characteristics of the 24 networks (cited and uncited for four disciplines in three nations) are further summarised in Table 7 in terms of their density, modularity and centrality. The density is the ratio of the number of actual edges in the network to the total possible number of edges. The modularity is the fraction of the edges within the network minus the fraction of the edges that would be expected if they were assigned at random, so that a large positive value reflects a highly structured network. The centrality is the betweenness centrality for each node in the network, i.e., the number of shortest

Fig. 4 The co-authorship mapping from uncited UK publications for Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right)

paths passing through that node. While some of the parameter values differ little between countries or between cited and uncited networks, others support the view that the Chinese networks differ markedly from their American and British counterparts. For example, the numbers of authors with non-zero centrality in the largest community, where the Chinese uncited values are much larger than the other two values across all four disciplines. Again, both the percentage of authors in the largest community and the mean centralities are larger, and the percentage of singleton communities smaller, for China in the sets of both cited and uncited publications.

The influence of the larger threshold for productive scholars

A reviewer suggested that the use of a larger number of productive scholars might affect the results obtained. To check whether this is in fact the case, the UK and USA Economics data was re-analyzed using the top-100 scholars, rather than the top-50 as previously. The structural characteristics of the resulting networks are detailed in Table 8, which reveals a very similar pattern of behaviour to that observed in the Economics portion of Table 7. Analogous results are obtained if we compare the numbers and sizes of the various communities obtained using the two sets of scholars, as shown in Table 9. The top-50 co-authorship networks are

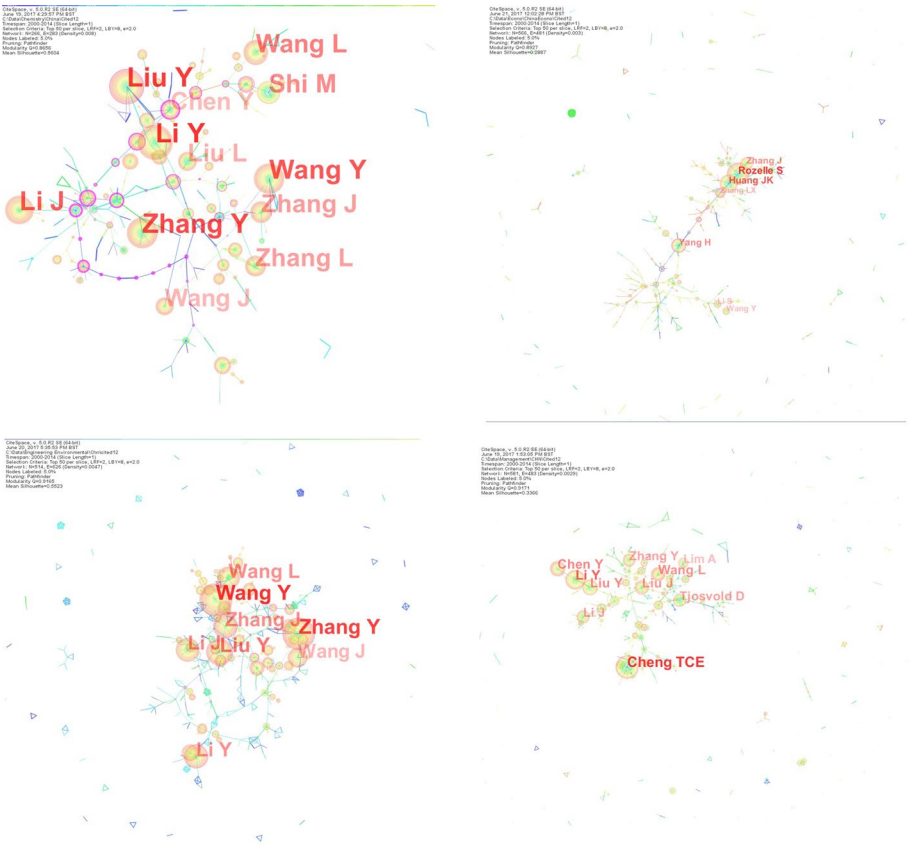

Fig. 5 The co-authorship mapping from cited Chinese publications for Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right)

dominated by large clusters (containing 97 and 177 authors for the USA and the UK respectively) for the cited publications and exactly the same behaviour is observed for the top-100 networks (240 and 364 authors for the USA and the UK respectively). That apart, the great majority of the authors are in small or singleton clusters containing between 1 and 19 members: 100% for the uncited USA and UK publications, and 82% and 76% for the cited USA and UK publications. There is again a statistically significant difference between the cited and uncited networks: the χ^2 values for USA and UK are 2.739 and 3.768 respectively, and the associated probability is 0.000 in both cases. It would hence seem that the choice of threshold has little or no effect on the overall structure of the research communities or on the differences between the cited and uncited publications.

Discussion and conclusions

As noted in the Introduction, we sought to answer two separate questions, viz whether there are differences between networks based on cited and uncited articles, or based on different countries. It is clear from the above that these two questions are inter-related in that

Fig. 6 The co-authorship mapping from cited UK publications for Chemistry Organic (upper-left), Engineering Environmental (lower-left), Economics (upper-right) and Management (lower-right)

the results for the two types of Chinese network differ considerably from those for the other two countries. Collaboration—whether at the institutional, national or international level—has long been recognised as an important factor in the ability of a research publication to attract citations. It might be hence expected that there would be greater evidence of collaboration in cited, as against uncited, publications. The co-authorship networks considered in this paper have demonstrated that this is certainly the case for USA and UK publications in four disparate research disciplines, since there are significant differences in structure between the cited and uncited networks. The former are characterised by a single, extended community involving a large fraction of the total sets of productive authors, together with a large number of much smaller, or even singleton, components; whereas the extended community is absent from the latter. There is some degree of variation in the precise figures but this is hardly unexpected given the very different ways in which disciplines are organized in the physical sciences, engineering and the social sciences, and the differences are small when compared with the corresponding results for the Chinese researchers.

We are unaware of any previous studies that have suggested that the Chinese results might be very different from those of other countries. Our results show that there is a much greater degree of commonality between the cited and uncited networks for the PRC, with both types of network exhibiting a single, highly extended community that encompasses a

Table 5 Communities based on uncited and cited publications

Size of community	Number of communities and of their members for uncited publications			Number of communities and of their members for cited publications		
	USA	PRC	UK	USA	PRC	UK
Chemistry organic						
≥ 200 authors		1 (233)		1 (217)	1 (240)	1 (250)
20–100 authors						
2–18 authors	36 (111)	30 (179)	19 (190)	26 (93)	4 (110)	25 (178)
Singleton	47	65	17	103	16	78
Engineering environmental						
≥ 200 authors		1 (450)		1 (225)	1 (365)	1 (318)
20–100 authors	1 (24)					2 (66)
2–19 authors	209 (708)	42 (103)	56 (179)	42 (137)	32 (90)	165 (612)
Singleton	250	63	75	155	59	170
Economics						
≥ 200 authors		1 (228)			1 (255)	
20–100 authors				1 (97)		4 (177)
2–19 authors	133 (356)	37 (89)	108 (312)	101 (271)	54 (146)	97 (266)
Singleton	349	143	304	430	165	353
Management						
≥ 200 authors		1 (245)		1 (280)	1(316)	1 (217)
20–100 authors						1 (35)
2–19 authors	169 (432)	30 (74)	123 (346)	113 (295)	49 (136)	145 (409)
Singleton	358	85	206	380	129	369

Each cell contains the number of communities within a given size range, with the total number of authors within those communities bracketed

Table 6 Chi squared analysis for differences between the cited and uncited networks

Nation	Chemistry organic		Engineering environmental		Economics		Management	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p
USA	156.9	0.000	567.5	0.000	111.6	0.000	293.5	0.000
PRC	66.05	0.000	0.670	0.715	6.019	0.049	4.682	0.096
UK	175.3	0.000	122.8	0.000	164.0	0.000	175.0	0.000

large fraction of the complete set of productive authors. So why have the extensive linkages, as reflected in the co-authorship network, not resulted in the expected impact in terms of citation in the published literature? One possible reason may be a cultural phenomenon arising from the widespread use of gift or honorary authorship, i.e., the inclusion in a list of authors of individuals who have made little, or no, substantive contribution to the research described in the publication. Here, a combination of the 'publish or perish' syndrome, payment for publication in WoS journals, and an in-built respect for authority figures in China mean that it is common for new, or less qualified, staff to include better

Table 7 Characteristics of the networks for uncited and cited publications

Network characteristic	Uncited publications			Cited publications		
	USA	PRC	UK	USA	PRC	UK
Chemistry organic						
Density of the complete network	0.020	0.006	0.025	0.004	0.008	0.005
Modularity of the complete network	0.955	0.869	0.899	0.900	0.866	0.916
Number of authors with non-zero centrality in the largest community	0	104	4	107	136	136
Mean centrality for all authors in the complete network	0.000	0.045	0.015	0.042	0.138	0.068
Mean productivity for all authors in the largest community	3.000	4.692	3.375	29.120	69.901	24.880
Percentage of authors in the largest community	4.4	61.8	15.0	53.5	90.2	61.6
Percentage of singleton communities	29.7	17.2	18.9	24.9	6.0	19.2
Engineering environmental						
Density of the complete network	0.002	0.003	0.005	0.003	0.005	0.003
Modularity of the complete network	0.982	0.912	0.962	0.905	0.917	0.946
Number of authors with non-zero centrality in the largest community	0	169	0	89	171	31
Mean centrality for all authors in the complete network	0.000	0.074	0.000	0.048	0.061	0.023
Mean productivity for all authors in the largest community	2.333	14.560	2.688	16.358	26.162	6.287
Percentage of authors in the largest community	2.4	73.1	6.3	43.5	71.0	27.3
Percentage of singleton communities	25.5	10.2	29.5	30.0	11.5	14.6
Economics						
Density of the complete network	0.002	0.003	0.002	0.001	0.003	0.002
Modularity of the complete network	0.925	0.913	0.977	0.935	0.893	0.904
Number of authors with non-zero centrality in the largest community	0	86	0	10	90	4
Mean centrality for all authors in the complete network	0.000	0.058	0.000	0.011	0.039	0.010
Mean productivity for all authors in the largest community	8.190	17.310	3.850	20.390	8.440	11.870
Percentage of authors in the largest community	2.3	49.6	1.5	16.3	45.1	7.7
Percentage of singleton communities	49.5	31.1	49.3	72.2	29.2	44.3

Table 7 (continued)

Network characteristic	Uncited publications			Cited publications		
	USA	PRC	UK	USA	PRC	UK
Management						
Density of the complete network	0.001	0.004	0.002	0.001	0.004	0.001
Modularity of the complete network	0.980	0.899	0.986	0.922	0.778	0.950
Number of authors with non-zero centrality in the largest community	0	115	0	60	82	30
Mean centrality for all authors in the complete network	0.000	0.064	0.000	0.018	0.025	0.015
Mean productivity for all authors in the largest community	6.190	32.080	3.300	11.200	8.465	8.357
Percentage of authors in the largest community	1.3	60.6	1.8	29.3	45.5	21.1
Percentage of singleton communities	45.0	21.0	37.3	39.8	18.6	35.8

Table 8 Characteristics of the networks for uncited and cited publications

Network characteristic	Uncited publications		Cited publications	
	USA	UK	USA	UK
Density of the complete network	0.001	0.001	0.001	0.001
Modularity of the complete network	0.972	0.983	0.936	0.943
Number of authors with non-zero centrality in the largest community	0	0	27	58
Mean centrality for all authors in the complete network	0.000	0.000	0.000	0.001
Mean productivity for all authors in the largest community	8.875	3.741	15.18	9.249
Percentage of authors in the largest community	1.3	2.9	18	24
Percentage of singleton communities	47.8	75.8	45	40.3

Table 9 Communities based on uncited and cited publications

Size of community	Number of communities and of their members for uncited publications				Number of communities and of their members for cited publications			
	TOP 50		TOP 100		TOP 50		TOP 100	
	USA	UK	USA	UK	USA	UK	USA	UK
≥ 200 authors							1 (240)	1 (364)
20–100 authors					1 (97)	4 (177)		
2–19 authors	133 (356)	108 (312)	224 (665)	169 (222)	101 (271)	97 (266)	163 (471)	184 (549)
Singleton	349	304	608	694	430	353	577	617

Each cell contains the number of communities within a given size range, with the total number of authors within those communities bracketed

established, senior colleagues as authors in an attempt to increase the chance that a submission will be accepted for publication in a prestigious journal. It must be emphasised that we are not suggesting that this is a purely Chinese phenomenon since it clearly occurs in many countries and disciplines (Sokol 2008; Strange 2008; Zaki 2011); however, it is known to be particularly well established in the Chinese context (Hvistendahl 2013; Liao et al. 2017; Macfarlane 2017). For example, an empirical analysis by Hao et al. (2009) found that guest authorship was involved in 28.6% of papers published in 2008 in the *Chinese Medical Journal*, the great majority of the guest authors being heads of departments or institutions. In contrast, Wislar et al. (2011) quoted a markedly lower figure of 11.4% for 2008 articles in three USA journals (*Annals of Internal Medicine*, *Journal of the American Medical Association*, and *New England Journal of Medicine*).[1]

[1] Wislar et al. also discussed ghost authorship in these 2008 journal issues, i.e., authors who were involved in the research or in the writing of a paper but are unnamed there (Bavdekar 2012). This also is a major problem in PRC, with Hu and Wu (2013) providing a detailed analysis of the websites of 127 Chinese companies offering large-scale ghost-writing services.

Chen and MacFarlane (2016) have suggested that guanxi (a generic term for the networks of relationships that are used to oil the wheels of business and more generally) and the intensive norms of reciprocity that dominate academic life in China are very different from the culture of authorship in Western academe. Tang et al. suggested guanxi as one of several factors that could influence the citation counts for highly cited publications in nanotechnology, a topic that has been the focus of considerable Chinese research efforts (Tang et al. 2015). The work reported here extends that of Tang et al. in that our results suggest that guanxi may also occur in other disciplines (and not just in nanotechnology) and that this effect applies to uncited as well as to cited articles: consequently, this is a factor that might usefully form the basis for future bibliometric studies.

In this paper, we have focused on the role of gift authorship in China, but—as Rethinaraj and Chakravarty (2017) noted—such behaviour is common amongst researchers in Asia. In particular, the phenomenon has been extensively discussed in the Indian context (see, e.g., Bavdekar 2012; Daniel 2016; Mandal et al. 2015; Zaki 2011), and it would hence be of interest to determine whether this can be identified in co-authorship networks of the sort investigated here. Furthermore, it would also be worthwhile to explore how a co-authorship network might change if individual sets of publications were grouped not only by whether they were cited or not cited, but also by how many citations they received. Our study still belongs to a probing analysis, in the following studies, more disciplines and countries might be selected to verify the robust of the findings in this manuscript.

Acknowledgements This study was supported by the National Natural Science Foundation of China (Grant No. 71603128), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20160974), the China Scholarship Council for funding (Grant No. 201608320057). Furthermore, Zewen Hu thanks PhD candidate lucyantie Mazalan from University of Sheffield for help in Chi squared analysis.

References

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics, 5*(4), 594–607.

Adams, J. (2012). Collaborations: The rise of research networks. *Nature, 490*(7420), 335–336.

Bavdekar, S. B. (2012). Authorship issues. *Lung India, 29*(1), 76–80.

Burrell, Q. L. (2013). A stochastic approach to the relation between the impact factor and the uncitedness factor. *Journal of Informetrics, 7*(3), 676–682.

Calero, C., Buter, R., Cabello-Valdes, C., & Noyons, E. (2006). How to identify research groups using publication analysis: An example in the field of nanotechnology. *Scientometrics, 66*(2), 365–376.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America, 101*(supplement), 5303–5310.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology, 61*(7), 1386–1409.

Chen, S., & MacFarlane, B. (2016). Academic integrity in China. In T. Bretag (Ed.), *Handbook of academic integrity* (pp. 99–105). Singapore: Springer.

Cohen, J. E. (1991). Size, age, and productivity of scientific and technical research groups. *Scientometrics, 1*(3), 395–416.

Craig, I. D., Plume, A. M., Mcveigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?: A critical review of the literature. *Journal of Informetrics, 1*(3), 239–248.

Cugmas, M., Ferligoj, A., & Kronegger, L. (2016). The stability of co-authorship structures. *Scientometrics, 106*(1), 163–186.

Daniel, R. S. (2016). Gift authorship belittles faith on integrity of publications and India's future. *Current Science, 110*(9), 1610–1611.
Egghe, L. (2010). The distribution of the uncitedness factor and its functional relation with the impact factor. *Scientometrics, 83*(3), 689–695.
Elkins, M. R., Maher, C. G., Herbert, R. D., Moseley, A. M., & Sherrington, C. (2010). Correlation between the journal impact factor and three other journal citation indices. *Scientometrics, 85*(1), 81–93.
Etzkowitz, H. (1992). Individual investigators and their research groups. *Minerva, 30*(01), 28–50.
Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology, 62*(10), 1992–2012.
Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821–7826.
Glanzel, W. (2002). Coathorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends, 50*(3), 461.
Grossman, J. W. (2002). Patterns of collaboration in mathematical research Six degrees of separation? *SIAM News, 35*(9), 1–3.
Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*(5722), 697–702.
Hamilton, D. P. (1991). Research papers: Who's uncited now? *Science, 251*(4989), 25.
Hao, X., Qian, S., You, S., & Wang, M.-Y. (2009). Ghost writers and honorary authorship: A survey from the *Chinese Medical Journal*. In: International Congress on Peer Review and Biomedical Publication. Vancouver: American Medical Association at http://www.peerreviewcongress.org/abstracts_2009.html.
Hu, Z. W., & Wu, Y. S. (2013). An empirical analysis on number and monetary value of ghostwritten papers in China. *Current Science, 105*(9), 1230–1234.
Hu, Z. W., & Wu, Y. S. (2014). Regularity in the time-dependent distribution of the percentage of never-cited papers: An empirical pilot study based on the six journals. *Journal of Informetrics, 8*(1), 136–146.
Hu, Z., & Wu, Y. (2018). A probe into causes of non-citation based on survey data. *Social Science Information, 57*(1), 139–151.
Hu, Z., Wu, Y., & Sun, J. (2018). A quantitative analysis of determinants of non-citation using a panel data model. *Scientometrics, 116*(2), 843–861.
Hvistendahl, M. (2013). China's publication bazaar. *Science, 342*(6162), 1035–1039.
Inzelt, A., Schubert, A., & Schubert, M. (2009). Incremental citation impact due to international co-authorship in hungarian higher education institutions. *Scientometrics, 78*(1), 37–43.
Izquierdo, J. L. C., Cosentino, V., & Cabot, J. (2016). Analysis of co-authorship graphs of core-ranked software conferences. *Scientometrics, 109*(03), 1665–1693.
Kronegger, L., Ferligoj, A., & Doreian, P. (2011). On the dynamics of national scientific systems. *Quality & Quantity, 45*(5), 989–1015.
Kyvik, S., Reymert, I., & Glänzel, W. (2017). Research collaboration in groups and networks: differences across academic fields. *Scientometrics, 113*(2), 1–17.
Liang, L. M., Zhong, Z., & Rousseau, R. (2015). Uncited papers, uncited authors and uncited topics: A case study in library and information science. *Journal of Informetrics, 9*(01), 50–58.
Liao, Q.-J., Zhang, Y.-Y., Fan, Y.-C., Zheng, M.-H., Bai, Y., Eslick, G. D., et al. (2017). Perceptions of Chinese biomedical researchers towards academic misconduct: A comparison between 2015 and 2010. *Science and Engineering Ethics,* in press.
Lou, W., & He, J. (2016). Does author affiliation reputation affect uncitedness? *Proceedings of the Association for Information Science & Technology, 52*(1), 1–4.
Lu, H., & Feng, Y. (2009). A measure of authors' centrality in co-authorship networks based on the distribution of collaborative relationships. *Scientometrics, 81*(2), 499–511.
Maaike, B., Horlings, E., Groenewegen, P., Van der Weijden, I., & Van den Besselaar, P. (2015). Organizational factors influencing scholarly performance: A multivariate study of biomedical research groups. *Scientometrics, 102*(01), 25–49.
Macfarlane, B. (2017). The ethics of multiple authorship: Power, performativity and the gift economy. *Studies in Higher Education, 42*(7), 1194–1210.
MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: a study of uncited and seldom-cited literature. *Journal of the American Society for Information Science and Technology, 61*(1), 1–13.
Makino, J. (1998). Productivity of research groups-relation between citation analysis and reputation within research communities. *Scientometrics, 43*(1), 87–93.

Mandal, M., Bagchi, D., & Basu, S. R. (2015). Scientific misconducts and authorship conflicts: Indian perspective. *Indian Journal of Anaesthesia, 59*(7), 400–405.
Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review, 69*(2), 213–238.
Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64,* 016131.
Newman, M. E. J. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404–409.
Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America, 101*(supplement), 5200–5205.
Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology, 66*(4), 739–764.
Perianes-Rodriguez, A., Olmeda-Gomez, C., & Moya-Anegon, F. (2010). Detecting, identifying and visualizing research groups in co-authorship networks. *Scientometrics, 82*(02), 307–319.
Prathap, G. (2013). Second order indicators for evaluating international scientific collaboration. *Scientometrics, 95*(2), 563–570.
Pudovkin, A., Kretschmer, H., Stegmann, J., & Garfield, E. (2012). Research evaluation. Part 1: Productivity and citedness of a German medical research institution. *Scientometrics, 93*(01), 3–16.
Radosevic, S., & Yoruk, E. (2014). Are there global shifts in the world science base? Analysing the catching up and falling behind of world regions. *Scientometrics, 101*(03), 1897–1924.
Rethinaraj, T. S., & Chakravarty, S. (2017). "Unethical authorship is research misconduct" at https://sc-lab.org/pdf/2017-research-misconduct.pdf. Accessed 28th January 2018.
Reyes-Gonzalez, L., Gonzalez-Brambila, C. N., & Veloso, F. (2016). Using co-authorship and citation analysis to identify research groups: A new way to assess performance. *Scientometrics, 108*(3), 1–21.
Schwartz, C. A. (1997). The rise and fall of uncitedness. *College and Research Libraries, 58*(01), 19–29.
Seglen, P. O., & Aksnes, D. W. (2000). Scientific productivity and group size. A bibliometric analysis of Norwegian microbiological research. *Scientometrics, 49,* 125–143.
Smart, J. C., & Bayer, A. E. (1986). Author collaboration and impact: a note on citation rates of single and multiple authored articles. *Scientometrics, 10*(5–6), 297–305.
Sokol, D. K. (2008). The dilemma of authorship. *British Medical Journal, 336*(7642), 478.
Stern, R. E. (1990). Uncitedness in the biomedical literature. *Journal of the American Society for Information Science and Technology, 41,* 193–196.
Strange, K. (2008). Authorship: Why not just toss a coin? *American Journal of Physiology Cell Physiology, 295*(3), C567–C575.
Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics, 107*(03), 1195–1225.
Tang, L., Shapira, P., & Youtie, J. (2015). Is there a clubbing effect underlying Chinese research citation increases? *Journal of the Association for Information Science and Technology, 66*(9), 1923–1932.
Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics, 8*(4), 963–971.
Wang, L., & Wang, X. (2017). Who sets up the bridge? Tracking scientific collaborations between China and the European Union. *Research Evaluation, 26*(2), 124–131.
Wang, L., Wang, X., & Philipsen, N. J. (2017). Network structure of scientific collaborations between China and the EU member states. *Scientometrics, 113*(2), 765–781.
Wang, X., Xu, S., Liu, D., & Liang, Y. (2012). The role of Chinese—American scientists in China—US scientific collaboration: A study in nanotechnology. *Scientometrics, 91*(3), 737–749.
Wang, X., Xu, S., Wang, Z., Peng, L., & Wang, C. (2013). International scientific collaboration of China: Collaborating countries, institutions and individuals. *Scientometrics, 95*(3), 885–894.
Wislar, J. S., Annette, F., Fontanarosa, P. B., & Deangelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: A cross sectional survey. *British Medical Journal, 343*(7835), d6128.
Xie, Y., Zhang, C., & Lai, Q. (2014). China's rise as a major contributor to science and technology. *Proceedings of the National Academy of Sciences of the United States of America, 111*(26), 9437–9442.
Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A co-authorship network analysis. *Journal of the Association for Information Science and Technology, 60*(10), 2107–2118.
Zaki, S. A. (2011). Gift authorship—a cause for concern. *Lung India, 28*(3), 232–233.
Zhao, S. X. (2015). Uncitedness of reviews. *Current Science, 109*(8), 1377–1378.
Zhu, J., Hassan, S.-U., Mirza, H. T., & Xie, Q. (2014). Measuring recent research performance for Chinese universities using bibliometric methods. *Scientometrics, 101*(1), 429–443.