Aslib Journal of Information Management

Efficacy of a giant component in co-authorship networks: Evidence from a Southeast Asian dataset in economics

Sameer Kumar

Efficacy of a giant component in co-authorship networks

Evidence from a Southeast Asian dataset in economics

Sameer Kumar
Asia-Europe Institute, University of Malaya, Kuala Lumpur, Malaysia

Received 13 December 2014
Revised 18 September 2015
Accepted 5 October 2015

Abstract

Purpose – The purpose of this paper is to investigate whether a sparse and relatively small giant component (GC) will capture highly productive authors.

Design/methodology/approach – The author used a geographically dispersed data set involving authors in the field of economics in ten countries in Southeast Asia and applied social network analysis methods to investigate the structure and dynamics of GCs.

Findings – Results reveal that a GC, characterized by both low density and small size, can still capture a significant percentage (68 per cent of the top 25) of the most productive authors. There seems to be a topological backing for this occurrence. The number of direct connections (or "degree") in the GC was correlated with research productivity, such that high-degree authors were almost twice as productive as low-degree authors. It is probable that productive authors having higher than average degrees may be the cause of the formation of the GC. The author hypothesize that irrespective of its size or sparseness, GCs in co-authorship networks may still represent the seat of main intellectual activity in the network.

Originality/value – This is one of the first studies to quantitatively analyse the ability of a co-authorship-based less-prominent GC to capture prominent authors.

Keywords Economics, Research collaborations, Research productivity, Co-authorship network, Giant component, Southeast Asia

Paper type Research paper

Introduction

A social network is a connection between entities having some kind of relationship. Any aspect of our life may be viewed from the perspective of networks. For example, we can describe a friendship network as one in which two persons are connected as friends, or an airline network where two airports are connected by their carrier flights. Likewise, we can construct a co-authorship network by connecting two authors who have co-authored a research paper. Since research collaboration occurs even when researchers do not subsequently co-author a research paper, co-authorship is only a partial indicator of research collaboration (Katz and Martin, 1997). However, as a tangible and verifiable indicator of research collaborations (Glänzel and Schubert, 2004), co-authorship of a research article is often considered to be a reliable proxy of research collaboration. Co-authorship is commonly used to gauge the association between researchers at individual, institutional and national levels.

A giant component (GC) is the largest connected component in a network. In a co-authorship network (Olmeda-Gomez *et al.*, 2009), GCs often represent the seats of main intellectual activity in a community of researchers (Fatt *et al.*, 2010). The size of a

The study is supported by University of Malaya, project numbers: RP020D-14AFR and UM.C/625/1/HIR/MOHE/SC/13/3.

Aslib Journal of Information Management
Vol. 68 No. 1, 2016
pp. 19-32

2050-3806
DOI 10.1108/AJIM-12-2014-0172

GC is a well-studied research topic (Kumar and Jan, 2013b; Yin *et al.*, 2006; Kumar, 2015). In a network, the degree distribution (pattern of the number of direct connections node has) in a GC, as compared to the whole network, is somewhat skewed towards high-degree authors (Newman, 2010). A number of recent co-authorship network studies (Uddin *et al.*, 2012; Yan *et al.*, 2010; Kuzhabekova, 2011) have shown that the degree of vertices is positively correlated with the productivity of the authors. Could this mean that the GCs in co-authorship networks could trap a larger percentage of productive authors than the whole network? The purpose of this study is to investigate the relevance of this topological phenomenon to co-authorship networks.

GC in co-authorship networks

In a network, most nodes initially exist either in isolation or in small clusters. There is a "tipping point" (percolation level), a special value of probability:

$$p = \frac{1}{n} \tag{1}$$

(where n represents the number of nodes), above which a GC forms (Newman, 2008).

One of the most commonly studied characteristics in co-authorship network studies is the size of the GC. The size of a GC is significant as it may convey how cohesive (a community of entities that are closely linked to one another) or scattered (a community of entities that are not linked to one another) a network is. A GC in a network could also represent the "core" field of research and other smaller components may represent groups of authors working in specialized areas of the field.

While investigating the co-authorship networks of various disciplines, Newman (2001, 2004) found the size of the GC of science-based subjects to be anywhere between 57.2 per cent (computer science) and 92.6 per cent (biomedical research). Kretschmer (2004) noted that GCs generally occupy about 40 per cent of network nodes. However, small GCs have also been found. For example, while investigating the co-authorship network of library and information science authors in China, Yan and Ding (2009) detected the GC to comprise just 20.77 per cent of the network nodes. Similarly, other studies have also reported GCs of varying sizes (Yin *et al.*, 2006; Kumar and Jan, 2013b; de Souza and Barbastefano, 2011; Biscaro and Giupponi, 2014) in their respective fields. In a study by Erfanmanesh *et al.* (2012), the authors reported that the GC of scientometrics-based co-authorship networks consisted of the most prolific authors in the field.

The sizes of GCs vary between disciplines. It is generally higher in the natural or biomedical sciences (or physical sciences in general) than in the social sciences or humanities. The existence of fewer social science papers is a global phenomenon, the evidence for which is directly shown by the fact that in Asia, for example, there are 20 papers published in the physical sciences to one published social sciences paper (WoS, SCI and SSCI databases, September 2014). A greater number of papers indicate more authors and more collaborative activity in a given field (although the level of collaboration per paper may differ, depending on the field), which leads to the formation of more prominent GCs in the physical than in the social sciences. In addition, the possibility of a discovery (Bettencourt *et al.*, 2009), which is more common in the physical than in the social sciences, can spur increased activity in the field, thereby increasing the possibility for the faster formation of a GC.

In this study, we define a GC as "prominent" if it interconnects with more than 50 per cent of nodes (> 50 per cent) in the network (see Figure 1).

Once a component attracts more than 50 per cent of the total network nodes, it is impossible for any other component to exceed this component in size, irrespective of the activity that takes place in other network components. For example, let's assume a network has x nodes. If the largest GC has a node of more than $x/2$ nodes (or more than 50 per cent of the total number of nodes), it would have a total greater than or equal to $(x/2)+1$ nodes, making it impossible for any other component in the network to become larger than it (at best, the other component could garner $x-((x/2)+1)$ nodes or less than 50 per cent of nodes). More simply, if a network has 100 nodes, the GC must cross the threshold of containing 50 connected nodes (or 51 nodes) to become "prominent" (see Figure 1).

Research objective

Most of the previous studies of the GCs of co-authorship networks have focused on the size of or reasons for GC formation (percolation level, maturity of field, etc.). These studies have rarely looked at the efficacy or ability of GCs (irrespective of their size or density) in capturing productive authors. Rarely have they also investigated the link between collaborative ties, research productivity and the ability of the GC to trap productive authors. This study represents an attempt to fill this gap.

We expect that a GC in a dense network may trap the prominent actors, as these networks often have prominent GCs. However, our interest is in "less prominent" GCs, based on the premise that a data set in which the actors represent a large geographical region with lower levels of collaborative associations would also be unlikely to have a prominent GC. Specifically, do networks with relatively small GCs still capture productive authors? Here we use a sparse data set of economics researchers in Southeast Asian nations to determine whether less prominent GCs still capture prominent authors. In addition, we also examine other network aspects, including their topological properties, to gain a deeper insight into the subtleties of researcher associations.

In this study, we address the following main research question:

RQ1. Does a less prominent and low-density GC of a co-authorship network still capture highly productive authors?

The results of this study will add to the body of the GC literature and this is perhaps one of the first studies to test the efficacy of less prominent GCs in co-authorship networks.

Research methods

Our first task was to identify a sparse network for study. We consider a network to be sparse when it has low density, a long average path length and a relatively small GC (i.e. occupying less than 25 per cent of the total nodes in the network). Our guess was

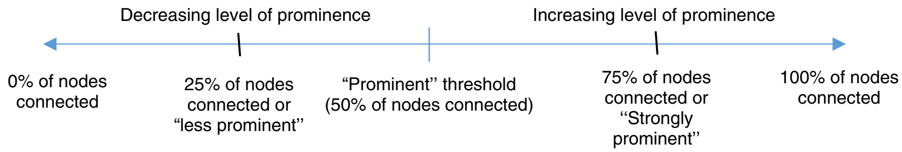


Figure 1. Different "prominence" levels of a giant component

that a geographically diverse data set gathered from a social sciences field would likely meet our requirements. After a few iterations, we chose as a case study the field of economics and the geographical region comprising the nations in Southeast Asia.

We harvested records from 1979 to 2010 from the Institute for Scientific Information Web of Science Social Sciences Citation Index (SSCI) databases with at least one author per article record from one of the following Southeast Asian countries – Singapore, Malaysia, Philippines, Laos, Cambodia, Vietnam, Thailand, Indonesia, Myanmar and Brunei. These Southeast Asian countries also form an economic association named Association for the Southeast Asian Nations (ASEAN) (the only Southeast Asian nation not part of ASEAN is East Timor). This query resulted in the extraction of 2,062 records.

Authors publish in a wide range of journals, and many of these may not be part of the SSCI database. As such, the boundary of this research is artificial as it includes only those articles indexed by the SSCI database. In addition, there is a possibility that two or more authors, who may have the same names, could be taken as one author and thus have their works and co-authorship merged in this study. On the other hand, an author might be taken as two or more authors in cases where his or her name is mentioned with different name variations. There have been cases in which a few authors have mentioned their names differently at different times in different papers. In such cases, an author would be identified as two or more authors (depending on the number of variations) and his work split between the author's name variations. In order to address this issue, we have used the Jaro (Jaro-Winkler text similarity) algorithm, which is built into Sci2 software (Sci2 Team, 2009). We cleaned the data using the Jaro similarity matrix at a 0.9 similarity level and then performed a manual check to avoid any inadvertent merging of names.

The basis of association is the author's co-authorship in a multi-authored paper (Kumar and Jan, 2013a, 2014; Uddin *et al.*, 2012). A co-authorship network is formed when two or more authors write a paper together. In graph terms, authors become nodes and the edges are the papers written together. Our co-authorship network includes all authors in the data set – those who have collaborated as well as those who have not collaborated. When visualized, isolates are depicted as dots in the network and provide a complete picture of the network. The size of a GC is calculated in proportion to all the authors (including isolates) in the network. Co-authors of a paper form an undirected edge (without arrows) between them, meaning that they have a mutual relationship; two authors writing a paper together would be well acquainted with each other.

The number of papers, unlike the number of citations, is the outcome over which the authors have direct control. Hence, to determine author productivity, the quantitative aspect, i.e., the number of papers authored or co-authored by the author, is taken into account. To quantitatively determine an author's productivity, there are two common methods – the whole method and the fractional method. We have undertaken the whole method for the calculation of productivity of authors. Using the whole method, every co-author, irrespective of the number of co-authors on the paper, receives credit for publishing a paper. In the fractional method, depending on the number of authors, the credit is either equally divided or is based on a weighted formula whereby a fractional credit is assigned (i.e. the first author may get more points than the third co-author). Although fractional counting may more accurately indicate the contribution by each author, this issue is often debated. For example, in certain papers the first author may have done the major portion of the work. In other instances, it may happen that all co-authors had contributed equally.

However, these aspects are rarely disclosed, hence bibliometric data are unable to trap finer nuances. For this reason, the whole paper counting method is often preferred (Kumar and Jan, 2013a).

We conducted a social network analysis in this study, which uses a set of established mathematical algorithms to investigate the network topologies of a network (Wasserman and Faust, 1994). The network is explored at two levels – the global level and the local level. At the global level, the network examines the topology of the whole network using clustering coefficients, average path lengths and degree of distribution and community formations. Centrality measures, such as degree, closeness and betweenness, view a network from the node (or local) level. The degree of distribution is the total number of direct connections in a node, and the betweenness and closeness centralities determine the relative importance of each node in the network.

The density of a network is a ratio of the number of links in the network and the maximum number of possible links (Otte and Rousseau, 2002). A low density indicates a sparse network in which the flow of resources is slow. A dense network (or a high-density network), on the other hand, indicates more connectedness and such networks typically have a faster flow of resources.

Geodesic distance is the shortest path between any two random nodes in a network (Newman, 2003). The diameter of a network is the maximum (or longest) geodesic distance between two random nodes in the network. Most networks, remarkably, demonstrate small geodesic distances, and thus, the concept of "six degrees of separation" or "small world" (Watts and Strogatz, 1998) has flourished. This means that, on average, any two random nodes in a network are at a distance of six "hops" from each other. Average geodesic distance is simply an average of the geodesic distance of the nodes in the network. It is quite possible that there may be more than one geodesic path between any two nodes in a network.

Also known as "transitivity", the clustering coefficient is the possibility that two nodes will connect if they have a common partner (Wasserman and Faust, 1994). Naturally formed networks (such as co-authorship networks) generally demonstrate high clustering coefficients. Short geodesic distances and high clustering coefficients are classic features of "small world" networks.

All our analyses and visualizations were performed with Sci2 and NodeXL software tools (Sci2 Team, 2009; Smith *et al.*, 2009).

Results and discussion

Research productivity

Of the 2,062 records extracted, those affiliated with Singapore have the most papers (1,026 papers), followed by the Philippines (295), Indonesia (244), Malaysia (242), Thailand (211), Vietnam (56), Cambodia (4) and Brunei (1). Laos and Myanmar have no records in our data set (see Figure 2). While data harvesting demonstrated the domination of Singapore in terms of number of papers produced, other nations contributed about 50 per cent of the papers, which is a significant figure. Singapore's developed nation status, coupled with English being one of its official languages, is likely the contributing factors for its high paper productivity. Singapore has two universities – National University and Singapore and Nanyang Technological University – that are ranked among the top 50 universities in the world (Symonds, 2013). These two universities contributed a significant portion of the papers produced by Singapore. Although national research productivity is not at issue here, this statistic does show that in this Southeast Asian region, countries such as Singapore significantly outperform other nations.

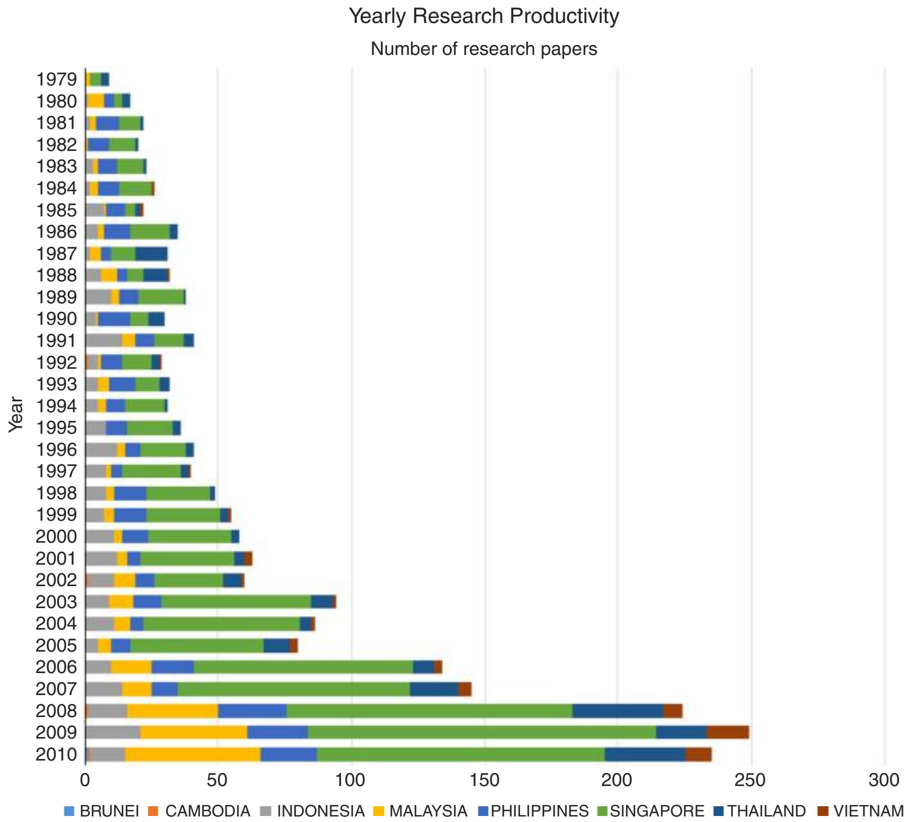


Figure 2. Southeast Asian nations in research productivity

Our examination of authorship begins with the identification of authors who either co-authored or solo-authored a paper. Some authors write both solo and co-authored papers. In our data set, for example, Rasiah, R., was a solo author in four of the 11 articles he authored. The data shows that 697 papers were written by single authors, making this one of the most prominent forms of authorship (accounting for 33.80 per cent of the total number of papers).

Increase in collaborative associations over the years

The growth in the rate of published papers represents the increasing incidence of collaboration (Persson *et al.*, 2004). As more authors collaborate, the number of papers per author increases. Over 38.40 per cent of the papers in our data set were written by two authors, 19.44 per cent were written by three authors, 5.28 per cent were written by four authors and 3.05 per cent were written by five or more authors. This data clearly reveals that the dyad is the most common collaborative form between authors, followed by triad and quad forms of association. Beyond quad association, the percentage of collaborative relationships falls sharply. In situations when co-authorship is the result of a large cross-country study, in which several authors (sometimes ten or more) co-write a paper with one or two main authors serving as co-ordinators between all authors, there is the likelihood that not every author is well acquainted with every other author. In co-authorship networks, subtle aspects such as this cannot be identified and thus represent one of its main research limitations.

Also known as hyper-authorships (in the physical sciences, it can sometimes run in the hundreds), the phenomenon of a large number of authors (11 or more) having written a paper is rare in our data set, contributing just 0.34 per cent (seven papers) of the total number of papers. The maximum number of authors in a paper in our data set is 25, and the average number of authors per paper is 2.12.

Figure 3 shows that the number of co-authored papers has increased over the past 30 years. From 2000 on we also see a rise of papers authored by five-plus authors, which signals an increase in collaboration. Research has also spread over larger geographical areas, which may have demanded the efforts of larger groups of researchers. However, we observe that in terms of numbers, solo authorship still predominates. The number of solo-authored papers has also gradually increased over the past 20 years; there were 20 solo-authored papers written in 1991, 27 in 2000 and 42 in 2010.

The overall network

The co-authorship network consists of 2,681 nodes (authors) with 3,779 undirected edges between them. These authors include authors from Southeast Asian countries and their foreign collaborators. The network is fragmented with a small GC, a large number of weakly connected components and a number of isolates. There are 764 weakly connected components. Graph density is 0.00105, which indicates that the network is very sparse. In addition, the GC is not prominent (or small), consisting of 544 nodes, which is about 20 per cent of the total number of nodes. There are 250 isolates in the network. Isolates are those authors within the data set who have not co-authored a paper. Large GCs form when there is a high level of collaborative activity between the authors of a community, which is sustained over a period of time. In contrast, the small-sized GC in this study indicates a low level of collaborative activity.

Network metrics of the whole graph and the GC are presented in Table I.

The second component (G2) is about one fifth the size of the GC (see Figure 4). If just one author of a component (i.e. G1) were to co-author a paper with an author in another component (i.e. G2), these two components would merge. This is how a GC increases in size.

The GC has a maximum geodesic distance of 23 and an average geodesic distance of 9.30. Its graph density of 0.0099190 signifies that it too is sparse in nature. In general, high-density graphs have lower geodesic distances and vice versa. Its average geodesic distance of approximately nine and its high clustering coefficient confirms its "small world" nature.

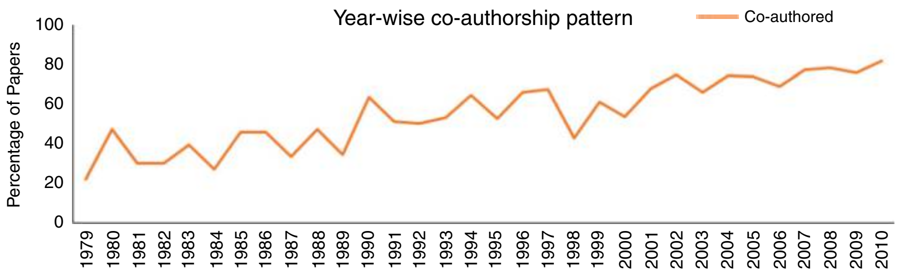


Figure 3.
Percentage wise co-authored papers have seen an increase

Most of the components are either dyads, triads or quads, which illustrates that a large number of authors work in small isolated groups, and but for a few exceptions, most write only a small number of papers.

The dynamics of GC formation

We can view the development of the GC in three time frames: 1979-1990, 1979-2000 and 1979-2010. The relative size (or emergence) of a GC is calculated by measuring the ratio of the nodes in the GC and the total number of nodes in the whole network (Barabasi *et al.*, 2002). The results of our study show that the initial formation of the GC is quite fluid while small components are consolidating, and it is not clear which component might actually grow into a visible GC. In its first time frame, the GC has only 15 nodes, yet it is able to capture six of the most productive authors. However, as the network grows and enters its second phase, the third largest component actually adds more nodes than the erstwhile largest component and ultimately supersedes to become the GC. The new GC, although very small, still captures five of the top 25 most productive authors. As the network enters its third phase, the largest component of the first phase again adds nodes and reclaims its rank as the largest component, comprising 544 nodes. This GC is now able to capture 17 of the top 25 most productive authors (see Figure 5).

Throughout the three time frames, authors within the GC have written more papers and are also better connected than authors in the rest of the network (see Table II).

Correlation between author degree and research productivity

In the GC (of the overall network during 1979-2010), there are 250 authors who have not co-authored a paper, but some of these authors are prolific authors and in the course of

Table I. Topological properties of whole network and giant component

	Whole network		Giant component	
	Average	Median	Average	Median
Degree	2.819	2	5.386	2.5
Clustering coefficient	0.515	0.5	0.562	0.667

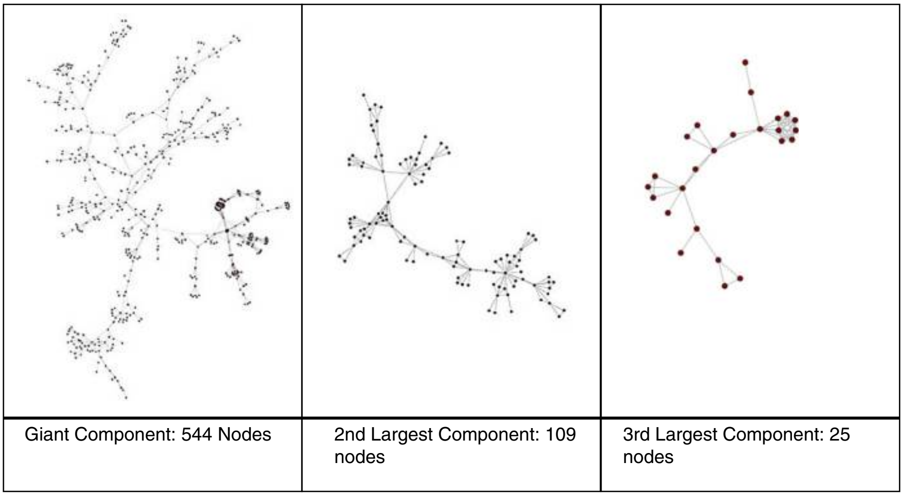
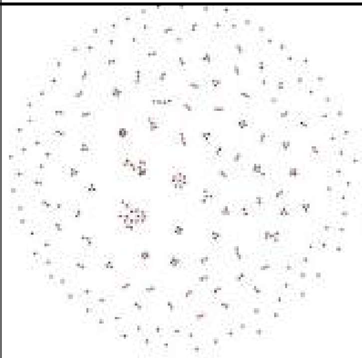
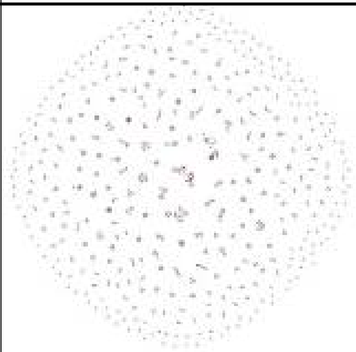
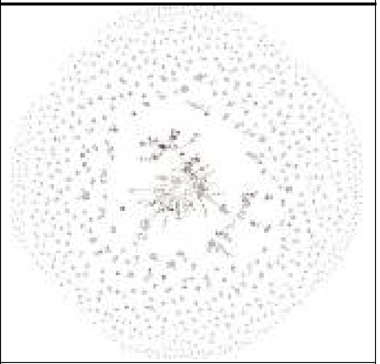


Figure 4. Giant component (or largest component) and the second and third largest components

Time frame/phase	1979-1990	1979-2000	1979-2010
Visualization of the network			
No. of papers	321	714	2,062
Correlation between Degree and Research Productivity	0.32*	0.36*	0.35*
Nodes	321	859	2,681
Isolated nodes	93	168	250
Edges	210	713	3,779
Total Number of Connected Components	171	381	764
Giant component size	15 nodes (6.6% of the total network)	22 nodes (3.18% of the total network)	544 nodes (22.37% of the total network)
Productive authors in GC (top 25 or as per the no. of nodes in GC)	6/15**	5/22**	17/25

Note: *p<0.01; **GC too small

Figure 5.
Dynamics of GC formation

a whole graph study they cannot be ignored. Hence we identified authors who have written five papers or more and determined the component to which they belong or whether they have been isolates. In total, 129 authors have written five or more papers. Out of these 129, four are isolates, which is about 3 per cent of all prolific authors.

70 prolific authors are captured by the GC which is about 54 per cent of all authors with five or more papers to their credit. Among the 129 authors with five articles or more, the trend of the number of articles with respect to degree is listed in Table III.

Degree metrics could be rather misleading since author X may have written just one paper but may have co-authored it with 20 others in a cross-country study. This may convey a distorted sense that author X is highly collaborative. Hence, to remove the noise brought by papers with a large authorship, we filter out those authors who have a high number of papers published, and analyse the number of co-authors with whom they have written articles to determine if they are part of the largest component. When segregated into two layers (one to seven and eight to 30), the degrees of authors suggest that, on average, authors characterized by a larger degree produce more papers (and vice versa). We understand that most of the prolific authors, in terms of number of papers co-authored, are also good collaborators. There is a significant correlation between degree and research productivity in all three time frames (see Figure 5). However, this significance in correlation, which is around 0.35, also implies that there are authors who may be less collaborative but still very prolific (and vice versa). Some low-degree authors also have produced large number of papers (i.e. Fabella, Rv-Vertex code: n1895, see Table IV).

GC captures prolific authors

Table IV lists the top 25 prolific authors across all components, their research productivity, their degree centrality measures and the component to which they belong.

Among the top 25 authors we see no isolates, which suggests that authors who co-author papers write more than do isolates (at least in this data set). Isolates, due to there being no division of labour, must put all their efforts into writing a paper and thus may be able to write a fewer number of papers than those with two or more co-authors. It is pertinent to note that the GC leaves out some of the very prolific authors, including Ang, B.W., Otsuka, K., Tang, T.C., Hoon, H.T. and Rasiah R., among others.

Table II. Average author degree and papers per author in overall network and GC

	Overall network		Giant component	
Time phases	Average author degree	Average no. of papers/author	Average author degree	Average no. of papers/author
1979-1990	2	1.4	3	2.6
1979-2000	2	1.4	3	2.4
1979-2010	3	1.6	5	2.4

Table III. Degree of authors with five papers and above

Degree	No. of authors	Average no. of papers
1-7	85	6.71
8-30	40	11.50

Author name	Vertex	Degree	No. of works	Component[a]
Tse, Y.K.	n446	17	25	G1
Ang, B.W.	n87	14	21	G6
Baharumshah, A.Z.	n55	21	19	G1
Wong, W.K.	n131	18	19	G1
Phillips, P.C.B.	n147	14	19	G1
Zhang, J.	n678	12	19	G1
Otsuka, K.	n508	21	18	G2
Fabella, R.V.	n1895	1	17	G492
Wright, J.	n173	11	16	G8
Lee, M.J.	n395	11	15	G1
Park, D.	n31	9	15	G1
Tang, T.C.	n449	4	15	G25
Sun, Y.N.	n848	8	13	G9
Abeysinghe, T.	n880	5	13	G1
Hoon, H.T.	n874	3	13	G24
Liew, V.K.S.	n56	13	12	G1
Liu, H.M.	n560	12	12	G1
Habibullah, M.S.	n44	10	12	G1
Zeng, J.L.	n680	8	12	G1
Koh, W.T.H.	n1006	4	12	G1
Yang, Z.L.	n1033	9	11	G1
Rasiah, R.	n249	9	11	G19
Hill, H.	n805	8	11	G1
Alba, J.D.	n32	6	11	G1
Huang, W.H.	n187	4	11	G1

Note: [a]G1 is the largest or giant component

Table IV. Prolific authors captured by the giant component

Conclusion

Although the network is sparse, the GC captured a large percentage of the most prolific authors. While a number of co-authorship network studies have examined the GC, this study is one of the first to investigate the efficacy of the GC in its ability to capture productive authors, even when it is not prominent (small size) and is sparse. The maturity of a field (Bettencourt *et al.*, 2009) may be one of the reasons that brings growth in the number of nodes (and thereby in the number of connections) in a network. At some point, this leads to the formation of a largest component. However, there have been instances where a large number of nodes, even in a discipline that is fairly mature, has not resulted in the formation of a prominent GC. For example, Kumar and Jan (2013b) show that some engineering disciplines have maintained less prominent GCs even after decades of activity. There seems to be no single cause for the formation of a GC. Hence, it may be that just the presence of a large number of nodes is not sufficient. The pace of paper production together with other factors (i.e. a discovery in the field or an increasing multidisciplinary or interdisciplinary nature of the field) may be responsible for the faster formation of a GC. As fields or subject areas "nucleate" around shared concepts and techniques, collaboration may become more widespread, and thus enable smaller clumps of nodes to join one another and form a GC (Bettencourt *et al.*, 2009). We have seen that degree correlates with research productivity. A GC likely grows due to the efforts of productive authors.

Our study's significance lies in its examination of one of the most important network topological characteristics, namely, the GC and its efficacy in trapping productive authors. Could a GC, irrespective of its size or density, represent the seat of productive activity in a network? The study suggests that GCs in co-authorship are indeed not just the largest lump of connected nodes; rather they may represent, irrespective of their size or sparseness, the very seat or nucleus of intellectual activity in the community. Our results also show that it is the productive authors who are resident in the GC, and who are also better connected than the other authors in the entire network. The author degree and average number of papers per author has only increased over the time frames studied and this corresponds to the size of the GC. Our guess is that the productive authors may well be responsible for the formation of the GC. It is fairly clear that increasing the number of nodes in a network paves the way for the faster formation of a GC. The addition of nodes may occur due to the maturity of a field in which more authors are entering the field, investigating new ideas or working on existing ones.

The research arena is a stage where researchers either flourish or perish. It is understood that if prominent authors ceased to collaborate, scientific enterprise would be significantly hampered. A prominent GC in a co-authorship network may, for example, indicate faster information flows in this network, and vice versa. A network gradually builds as nodes connect with one another over a period of time. The pace of this connection varies depending on the collaborative activity in the network. In a co-authorship network, this would mean that more authors would collaborate to write papers on "hot" topic areas. It could also mean that, due to the maturity of the field, a large body of authors are already present to contribute papers to the field.

References

Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002), "Evolution of the social network of scientific collaborations", *Physica a-Statistical Mechanics and Its Applications*, Vol. 311 Nos 3-4, pp. 590-614.

Bettencourt, L.M.A., Kaiser, D.I. and Kaur, J. (2009), "Scientific discovery and topological transitions in collaboration networks", *Journal of Informetrics*, Vol. 3 No. 3, pp. 210-221.

Biscaro, C. and Giupponi, C. (2014), "Co-authorship and bibliographic coupling network effects on citations", *Plos One*, Vol. 9 No. 6.

de Souza, C.G. and Barbastefano, R.G. (2011), "Knowledge diffusion and collaboration networks on life cycle assessment", *International Journal of Life Cycle Assessment*, Vol. 16 No. 6, pp. 561-568.

Erfanmanesh, M., Rohani, V.A. and Abrizah, A. (2012), "Co-authorship network of scientometrics research collaboration", *Malaysian Journal of Library & Information Science*, Vol. 17 No. 3, pp. 73-93.

Fatt, C.K., Abu Ujum, E. and Ratnavelu, K. (2010), "The structure of collaboration in the Journal of Finance", *Scientometrics*, Vol. 85 No. 3, pp. 849-860.

Glänzel, W. and Schubert, A. (2004), "Analysing scientific networks through co-authorship", in Moed, H.F., Glänzel, W. and Smoch, U. (Eds), *Handbook of Quantitative Science and Technology Research*, Kluwer Academic Publishing, Dordrecht, pp. 257-276.

Katz, J.S. and Martin, B.R. (1997), "What is research collaboration?", *Research Policy*, Vol. 26 No. 1, pp. 1-18.

Kretschmer, H. (2004), "Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web", *Scientometrics*, Vol. 60 No. 3, pp. 409-420.

Kumar, S. (2015), "Co-authorship networks: a review of the literature", *Aslib Journal of Information Management*, Vol. 67 No. 1, pp. 55-73.

Kumar, S. and Jan, J.M. (2013a), "Mapping research collaborations in the business and management field in Malaysia, 1980-2010", *Scientometrics*, Vol. 97 No. 3, pp. 491-517.

Kumar, S. and Jan, J.M. (2013b), "On giant components in research collaboration networks: case of engineering disciplines in Malaysia", *Malaysian Journal of Library & Information Science*, Vol. 18 No. 2, pp. 65-78.

Kumar, S. and Jan, J.M. (2014), "Research collaboration networks of two OIC nations: comparative study between Turkey and Malaysia in the field of 'energy fuels', 2009-2011", *Scientometrics*, Vol. 98 No. 1, pp. 387-414.

Kuzhabekova, A. (2011), "Impact of co-authorship strategies on research productivity: a social-network analysis of publications in Russian cardiology", doctoral dissertation, University of Minnesota, available at: http://conservancy.umn.edu/handle/108109 (accessed 15 April 2014).

Newman, M. (2010), *Networks: An Introduction*, Oxford University Press, Oxford.

Newman, M.E.J. (2001), "The structure of scientific collaboration networks", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 98 No. 2, pp. 404-409.

Newman, M.E.J. (2003), "The structure and function of complex networks", *Siam Review*, Vol. 45 No. 2, pp. 167-256.

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, Vol. 101*, pp. 5200-5205.

Newman, M.E.J. (2008), "Mathematics of networks", in Blume, L.E. and Durlauf, S.N. (Eds), *The New Palgrave Encyclopedia of Economics*, 2nd ed., Palgrave Macmillan, Basingstoke.

Olmeda-Gomez, C., Perianes-Rodriguez, A., Ovalle-Perandones, M.A., Guerrero-Bote, V.P. and Anegon, F.D. (2009), "Visualization of scientific co-authorship in Spanish universities from regionalization to internationalization", *Aslib Proceedings*, Vol. 61 No. 1, pp. 83-100.

Otte, E. and Rousseau, R. (2002), "Social network analysis: a powerful strategy, also for the information sciences", *Journal of Information Science*, Vol. 28 No. 6, pp. 441-453.

Persson, O., Glanzel, W. and Danell, R. (2004), "Inflationary bibliometric values: the role of scientific collaboration and the need for relative indicators in evaluative studies", *Scientometrics*, Vol. 60 No. 3, pp. 421-432.

Sci2 Team (2009), "Science of Science (Sci2) tool", Indiana University and SciTech Strategies, available at: http://sci.slis.indiana.edu (accessed 16 August 2014).

Smith, M.A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A. and Gleave, E. (2009), "Analyzing (social media) networks with NodeXL", *Proceedings of the Fourth International Conference on Communities and Technologies*, pp. 255-264.

Symonds, Q. (2013), "QS world university rankings", available at: www.topuniversities.com (accessed 14 November 2013).

Uddin, S., Hossain, L., Abbasi, A. and Rasmussen, K. (2012), "Trend and efficiency analysis of co-authorship network", *Scientometrics*, Vol. 90 No. 2, pp. 687-699.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis, Methods and Applications*, Cambridge University Press, New York, NY.

Watts, D.J. and Strogatz, S.H. (1998), "Collective dynamics of 'small-world' networks", *Nature*, Vol. 393 No. 6684, pp. 440-442.

Yan, E.J. and Ding, Y. (2009), "Applying centrality measures to impact analysis: a coauthorship network analysis", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 10, pp. 2107-2118.

Yan, E.J., Ding, Y. and Zhu, Q.H. (2010), "Mapping library and information science in China: a coauthorship network analysis", *Scientometrics*, Vol. 83 No. 1, pp. 115-131.

Yin, L.C., Kretschmer, H., Hanneman, R.A. and Liu, Z.Y. (2006), "Connection and stratification in research collaboration: an analysis of the COLLNET network", *Information Processing & Management*, Vol. 42 No. 6, pp. 1599-1613.

Corresponding author
Dr Sameer Kumar can be contacted at: sameer@um.edu.my