On giant components in research collaboration networks: Case of engineering disciplines in Malaysia

Sameer Kumar[1] and Jariah Mohd. Jan[2]
[1]Asia-Europe Institute, University of Malaya,
50603 Kuala Lumpur, MALAYSIA
[2]Faculty of Languages and Linguistics, University of Malaya,
50603 Kuala Lumpur, MALAYSIA
e-mail: sameerkr9@gmail.com; jariah@um.edu.my

ABSTRACT

The purpose of this study was to empirically investigate the size of giant components in the scholarly networks of prominent engineering disciplines in Malaysia. A co-authorship network is constructed by connecting two authors if they have co-authored a research article together. By applying Social Network Analysis (SNA), the size of the giant component of co-authorship networks was investigated in the four prominent engineering disciplines, namely electrical and electronics (EEE), chemical (CHEM), civil (CIVIL), and mechanical (MECH), involving 3675 records of scholarly articles, in which at least one of the researchers per article had a Malaysian address. Results revealed that well-formed giant components (size >50% of all nodes) were already present in EEE and CHEM disciplines, whereas they were at an undeveloped stage in the case of both CIVIL and MECH. All the four disciplines demonstrated small-world properties. However, those with larger giant components also had larger degree of separation (geodesic distance) between the nodes. Density of the nodes was negatively correlated with the size of the giant component. After the mid-90s, both CHEM and EEE had a faster production of articles than the other two disciplines, which corresponds with their well-formed giant components.

Keywords: Co-authorship networks; Social networks; Percolation level; Giant components; Scientific communications; Engineering.

INTRODUCTION

The social and cognitive processes that stimulate scientific knowledge have kept mankind curious for centuries (Racherla and Hu 2010). Patterns of human interaction have remained a topic of significant interest in the field of social sciences during the last 50 years (Wasserman and Faust 1994; Newman 2003). The production and dissemination of scientific knowledge, grounded in cognitive science and psychology, often has a social context (Pepe 2008). The social function through which scientists come together to collaborate contributes to the overall output of research community. Recent decades have seen a phenomenal increase in research publications, which is attributed to increased interaction between researchers through formal and informal channels. The informal channel through which the researchers collaborate is often facilitated by social networks. The success of scientific ties depends to a large extent on the effectiveness of these relationships. An in-depth analysis of informal knowledge networks provides an opportunity to investigate its structure. For example, patterns of these relationships could reveal the mechanism that shapes our scientific communities (Racherla and Hu 2010).

Scientists communicate with one another to convey their point of views, share research results and write research papers (Katz and Martin 1997). Communication between scientists could start with something small, such as discussing ideas. These communications sometimes lead to more serious discussions, to a point where researchers decide to put significant effort into a research project. When the effort is significant, it is termed as a collaboration. On the other hand, scientists may know other fellow scientists in advance, and may decide to collaborate right from the beginning, knowing well the competence of one another in their ability to carry out a research project. Hence, a variety of reasons could bring scientists together to collaborate on a piece of research. It is understood that scientists cannot be co-authors on a paper unless they have conducted the research together. When scientists write a paper, they are co-authors of the paper and their contribution is deemed as the most tangible indicator of scientific production (Glänzel and Schubert 2005). Bibliographic data, thus, is a widely used method to measure scientific achievement (Abrizah and Wee 2011).

A *social network* is an umbrella term that describes a set of people or entities connected through some kind of interdependency (Cross, Liedtka and Weiss 2005; Liu et al. 2005). For example, people could be connected through a friendship relation, researchers could be linked if they have written a scholarly paper together, or trade flows could link countries together. On a graph, entities (or 'nodes' or 'actors' or 'vertices') are represented with dots and connections (or 'ties' or 'edges') are represented with lines connecting those dots. Airline networks, highway networks or organizational networks could all be represented through this concept of nodes and edges. Unlike the conventional individualistic social theory that pays more attention to the personal attributes and little attention to the social circumstances of an individual (Knoke and Kuklinski 1982), network analysis gives prominence to the relationship one individual has with another. Attributes of entity are not ignored, but rather are seen in the context of the relationship.

GIANT COMPONENTS IN RESEARCH COLLABORATION NETWORKS

A component is a set of nodes in the network connected in a way such that any node could be reached by any other node by "traversing a suitable path of intermediate collaborators" (Newman 2004a). In a network, usually there are components of varying sizes. A giant component is the component having the largest number of connected vertices. Giant components have been extensively studied for random graphs (Molloy and Reed 1995). However, there is a fundamental difference between the pattern of degree distribution in random networks (or graphs) and real-world networks. In contrast to random networks, in real-world networks few nodes receive a lot of connections, while most others only a few or none. This property where some nodes are hubs (also known as the 'scale-free' property of the network), causes any two randomly chosen nodes in the network to be at short geodesic distance from each other (also known as the 'small-world' property of the network) (Barabasi and Albert 1999; Watts and Strogatz 1998).

As with random networks, in real-world networks, such as the co-authorship network, most vertices initially exist in isolation or in small clusters (or components) of connected vertices. The network then dynamically grows with the addition of new vertices and edges in the network. There is a percolation transition (or tipping point) at a special value of probability,

$$p = 1/n \qquad (1)$$

where *n* is the number of vertices, above which a giant component forms, the largest group of connected vertices (Newman 2007).

Giant components in research collaboration networks may represent core of mainstream research activity (Fatt, Abu Ujum and Ratnavelu 2010). Studies in the past have determined the size of giant components of co-authorship networks; Newman (2004a; 2004b) found the giant component in large databases of biomedical research database MEDLINE to be 92.6%, high-energy physics database SPIRES to be 88.7%, computer science database NCSTRL to be 57.2%, biology to be 92%, and physics to be 85% of the total number of vertices in the network. While investigating a small co-authorship network of 381 authors, Kretchmer (2004) found the size of the giant component to be approximately 40%. In fact, most studies on co-authorship networks invariably calculate the size of the largest component. As the global and local metrics are measured for the giant component, detecting the largest component and determining its size is crucial to understanding the topological features of the network.

OBJECTIVES

Most previous studies on giant components in co-authorship networks have been specific to subject area. Here, we calculated size of giant component for a country-specific dataset pertaining to Malaysia of four prominent engineering disciplines as per Web of Science (WoS) subject categories, namely chemical engineering (CHEM), electrical and electronics engineering (EEE), civil (CIVIL) engineering and mechanical engineering (MECH). Being country-specific, it is understood that a majority of the authors would represent Malaysia, although there would be international counterparts with whom the Malaysian authors would have collaborated.

This study aims to (a) calculate the size of giant components in the collaborative networks in the aforesaid four engineering disciplines in Malaysia, based on WoS subject categories; (b) examine if there is any correlation between the degree, density, clustering coefficient, degree of separation between the nodes and the size of giant components; and (c) examine if pace of paper production has any relationship with the formation of giant component.

MATERIAL AND METHODS

We followed the WoS subject categories when extracting the data set of each discipline. During the third week of June 2011 we queried all the 5 databases in the WoS, namely, SCI-Expanded, SSCI, A&HCI, CPCI- S, CPCI-SSH for all years in the disciplines of electrical and electronics engineering (EEE), chemical engineering (CHEM), civil engineering (CIVIL), and mechanical engineering (MECH) with Malaysia as the address of at least one of the authors in each article. The following search query was used for EEE, and similar queries were followed for the other three disciplines.

> Address=(Malaysia), Refined by Document Type=(ARTICLE) AND Subject Areas=(ENGINEERING, ELECTRICAL & ELECTRONIC), Time span=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH.

In the WoS database, publication records were found from 1973 to June 2011 (about 38 years). As research articles are prominent artifacts of new research, we have included only

'articles' as the document type in our study. The data from WoS was downloaded in blocks of 500 records (maximum download allowed by WoS at a time). For two subject categories (EEE and CHEM) that had records over 500, we appended the records by removing the 'EF' in the mid files to make one complete file for each category.

Construction of co-authorship network was carried out using Sci2 (Sci^2 2009) and visualization using GUESS, which is inbuilt in Sci2. We imported .graphml file of coauthorship network created in Sci2 into NodeXL for the calculation of network topologies (Smith et al. 2009).

Our study follows the WoS category for subjects. EEE, CHEM and MECH are the top categories based on the number of papers published. Although environmental engineering had more number of papers than civil engineering (CIVIL), we chose the latter, as CIVIL is one of the more common engineering departments at universities in Malaysia and often environmental engineering is taken as a subset of civil engineering itself.

WoS subject category are non-heirarchial and based on journal title and its citation patterns (Leydesdorff and Rafols 2009). They reflect the overall content of the journals pooled into them. A journal may be categorized into multiple categories depending on its multidisciplinary material. All articles get tagged with the categories at the journal level. For example, *Journal of Hazardous Materials* is categorized in environment engineering., civil engineering and environmental sciences subject categories. Hence, all articles published in this journal, irrespective of its content, will be categorzied in all these three categories. By categorizing the journals based on 'relevance' (type of journals citing the journal) and not with 'heirarchy', WoS subject category handles the multi-disciplinary issue of the journals (and articles) quite effectively. JISC (2012) has a good explanation on how WoS categorizes journals.

Construction of a Co-authorship Network

A network of researchers can be constructed if two researchers co-author a scholarly paper together. In this case, scholars would form the nodes and the paper they have co-authored would represent the link between them. For example, if four authors,

$V_1 = \{a,b,c,d\}$, co-write a paper, the co-authorship links they form is,

$E_1 = \{\{a,b\},\{a,c\},\{a,d\},\{b,c\},\{b,d\},\{c,d\}\}$.

Again when *c* co-writes a paper with *f*, $V_2 = \{c,f\}$, the link is represented as $E_2 = \{c,f\}$. Similarly, when *d* co-writes a paper with *b* and *i*, $V_3 = \{d,b,i\}$ the links are represented as $E_3 = \{\{d,b\},(d,i\},\{b,i\}\}$.

The lines between the nodes in a co-authorship network are undirected, symbolizing mutual relationship. This could be graphically depicted as in Figure 1.

Although, to be termed a giant component, it is not mandatory for the largest component to have a certain percentage of size of total *n*; for this study, the largest component was considered *well-formed* giant component only if it contained a majority (>50 percent) of the total *n* of the network (see Figure 2).

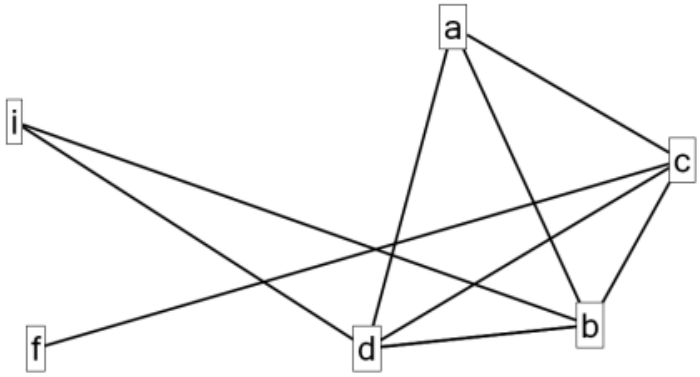

Figure 1: An example of Co-authorship Network

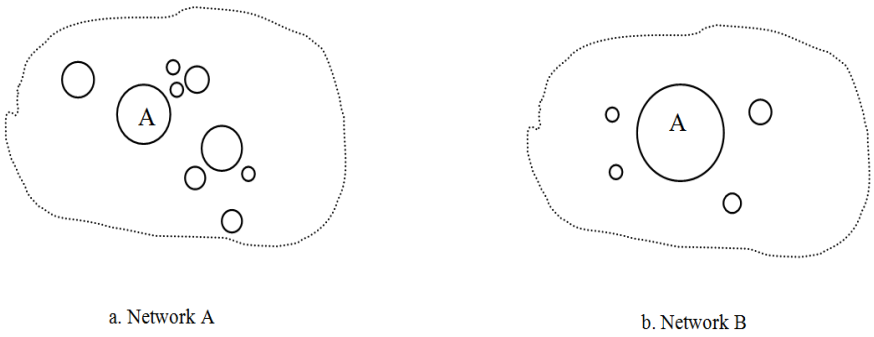

Figure 2: In (a) component A is the largest component, but not a *well-formed* giant component as per our classification. In (b), component A is the largest component and a *well-formed* giant component (component possessing majority of vertices).

RESULTS AND DISCUSSION

Summary of analysis of the four engineering subject categories is given in Table 1. EEE has the maximum number of papers, followed by CHEM, MECH and CIVIL. The ratio between the number of distinct authors and number of papers produced is in the range of 1.59 to 1.73 for EEE, CHEM and MECH, but a good 2.12 for CIVIL, which means that although CIVIL had relatively more authors in the network, they have produced lesser number of papers. Number of authors per paper and author productivity (average number of papers per author) is fairly consistent across the four subject categories. Authors wrote about 2 papers each and average paper had about 3 co-authors each. Figure 3 and 4 show the distribution of papers per author and authors per paper, respectively.

In the co-authorship network of CHEM, a total of 1247 research articles had 1985 authors, who had 4710 collaborative links between one another. There were only 14 isolates or authors in CHEM who have never collaborated with any other authors in the dataset. Similarly, the number of articles, nodes, edges, and isolates of EEE, MECH and CIVIL are given in Table 1.

Table 1: Summary of the Analysis of Four Engineering Subject Categories

	CHEM	EEE	MECH	CIVIL
No. of Papers	1247	1560	466	402
Average papers per author	2.16	2.22	1.76	1.51
Average authors per paper	3.44	3.17	3.05	3.21
Average degree of Collaborators per author	4.74	4.28	3.71	3.75
No. of Nodes (number of distinct authors)	1985	2210	809	855
No. of Edges	4710	4759	1502	1604
Isolates	14	24	10	12
Number of components	163	215	132	173
Average Geodesic Distance	5.52	6.39	3.69	2.67
Maximum Geodesic Distance (Diameter)	14	17	13	9
Average Clustering Coefficient	0.791	0.739	0.756	0.755
Density (Disregarding weights)	0.0024	0.0019	0.0046	0.0044
Nodes in the Largest component	1269	1338	107	57
% Size of Largest component.	63.93	60.30	13.27	6.66

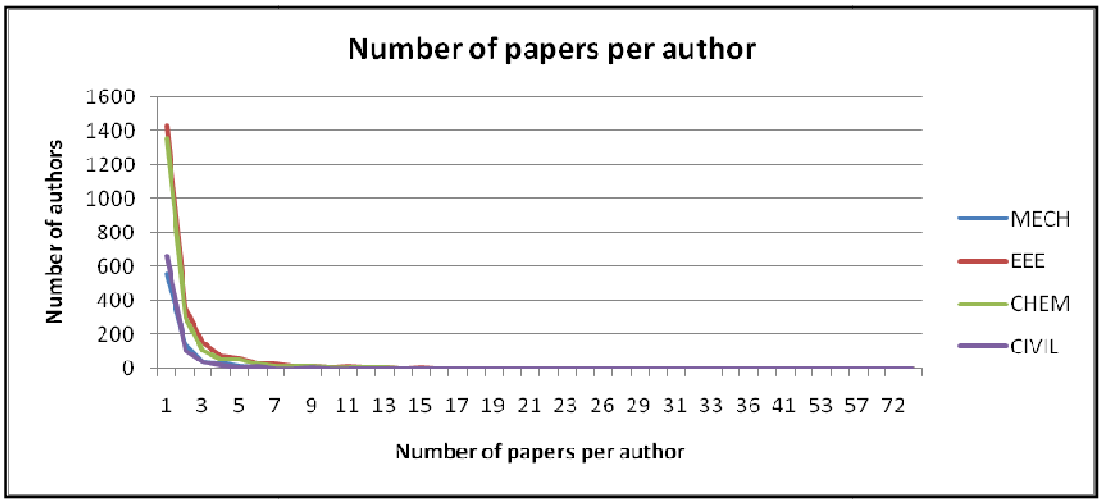

Figure 3: Number of Papers per Author (or Research Productivity) resembles a *Power Law* with Majority Publishing just 1 Paper (Mode Is 1)

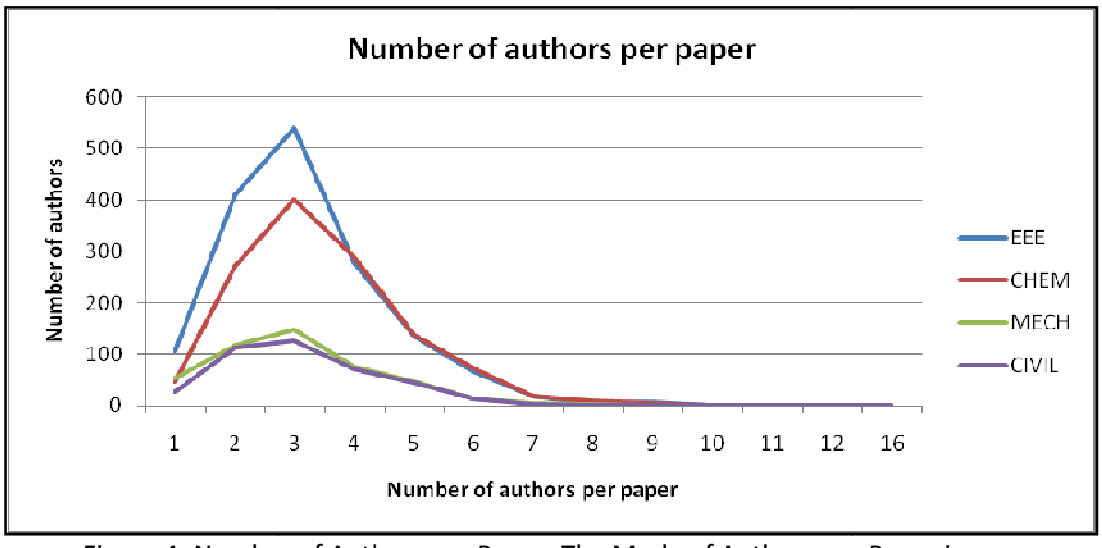

Figure 4: Number of Authors per Paper. The Mode of Authors per Paper is 3 for all the Categories

We then calculated the total number of components, average degree of each node, the density of network, clustering coefficient of the network and average and maximum geodesic distance of nodes in the network.

A degree of a node is the number of direct connections a node has. Degree is a prominent centrality measure. Degree k_i of a node is

$$k_i = \sum_{j=1}^{n} g_{ij} \qquad (2)$$

whereg_{ij}= 1 if there is a link between vertices *i* and *j* and g_{ij}= 0, if there is no such connection (Newman 2007).

The authors of all four engineering disciplines had an average of 4 collaborators each. A long tail depicts skewed degree distribution; majority of the authors had between 2 to 4 collaborators and few authors had a large degree of collaboration. An author in EEE had as high as 107 collaborators. Figure 5 shows the chart of degree of collaboration in the four disciplines.

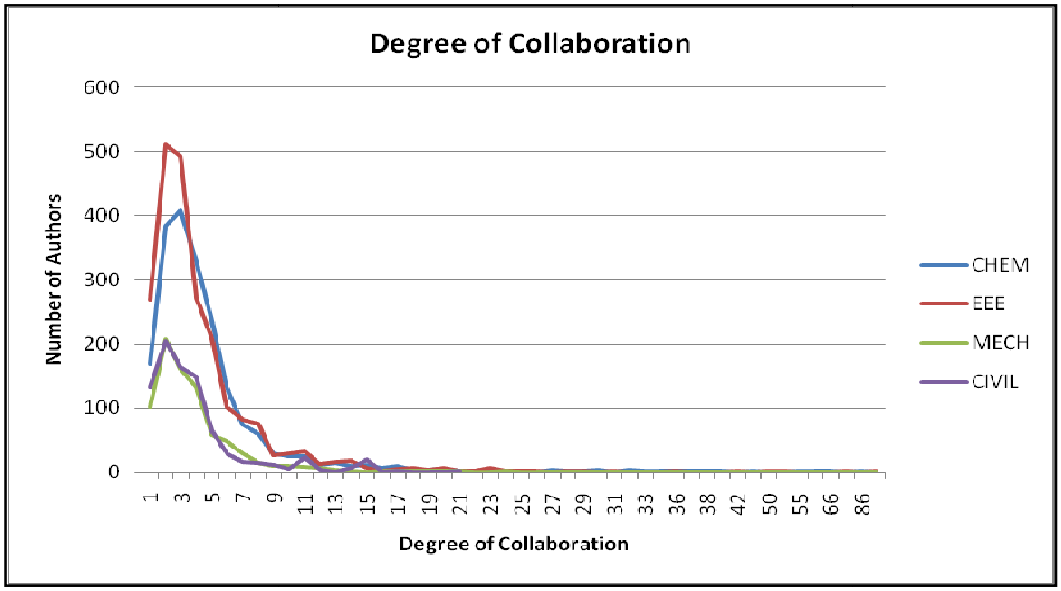

Figure 5: Degree of Collaboration of Authors. Mode for CHEM is 3 and a Total of 2 for EEE, MECH and CIVIL respectively

The degree of separation between any two random authors in the largest component had an average distance of about 6, confirming their 'small world' character. In a 'small world' model, any two random nodes are at shorter distance from each other (Watts and Strogatz 1998). Interestingly, MECH and CIVIL had average degree of separation as 3.69 and 2.67 respectively, when compared to their bigger counterparts EEE and CHEM, which had average degree of separation at 6.39 and 5.52 respectively (see Table 1).

Clustering coefficient of the network is the average of clustering coefficient values of the vertices. Also known as 'transitivity', clustering coefficient is defined as

$$C = \frac{3 \times no.of\ triangles}{no.of\ connected\ triples} \qquad (3)$$

where triangles represent trios of vertices (Newman 2004).

In simple terms, clustering coefficient determines the probability of A connecting to C, if A and B and B and C are already connected. We found the clustering coefficient of all subject categories to be fairly similar, hovering around 0.7 (see Table 1), which means that there is about 70% chance, in all these disciplines, for the nodes to form a clique.

The density of a network, G, indicates the number of links in the network in ratio to the maximum possible links. The density, D, of an undirected network P (cooperation network in which the relationship is mutual) with *n* vertices is expressed as (Otte and Rousseau 2002),

$$D = \frac{2*(\#L(P))}{n(n-1)} \quad (4)$$

The density was found to be low for larger networks (EEE 0.0019 and CHEM 0.0024) and relatively higher for small networks (CIVIL 0.0044 and MECH 0.0046) (see Table 1). The average degree and density of a network are indicative of connectivity of the network. Higher connectivity would result from more collaboration between the actors, thus causing faster diffusion of information through such networks.

Giant components of well-formed size have been formed (see Table 1) in CHEM (63.3%) and EEE (60.30%) disciplines. In the MECH and CIVIL disciplines, the size of largest component is at 13.27% and 6.66% respectively, hence still small to be considered a well-formed giant component (see Table 1). The dense central part of the network explicitly reveals giant components of EEE and CHEM disciplines. Visualization of the four co-authorship networks is presented in Figures 6a-6d.

Interestingly, there is a negative correlation between density of a network and the size of giant component (see Table 2). Networks of CIVIL and MECH are denser than the other 2 networks, yet their giant components are smaller in size (see Table 1). One possible explanation for this is that as the network grows the number of possible connections increase proportionately, thus, making the network sparser. There is a positive correlation between the average degree and the size of the giant component (see Table 2).

However, when it comes to clustering coefficient, we see a weak, yet positive, correlation with the size of the giant component. The number of nodes and edges has a positive correlation with the size of the giant component. The average degree of separation (average geodesic distance) positively correlates with the size of giant component (see Table 2). When the network is small, the average degree of separation between any two random nodes is also small due to high fragmentation and smaller giant component. As the network grows, the formation of giant component, which has large number of nodes inter-connected in a single component, increases the distance of separation between nodes.

Figure 6a: Visualization of co-authorship network of Electrical and Electronic engineering (EEE) WoS subject category. Large connected component in the middle shows the presence of *well-formed* giant component

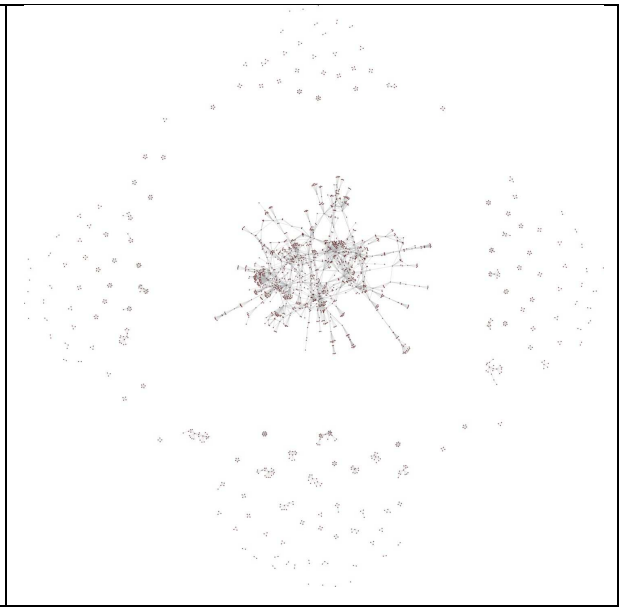

Figure 6b: Visualization of co-authorship network of Chemical Engineering (CHEM) WoS subject category. Large connected component in the middle shows the presence of *well-formed* giant component

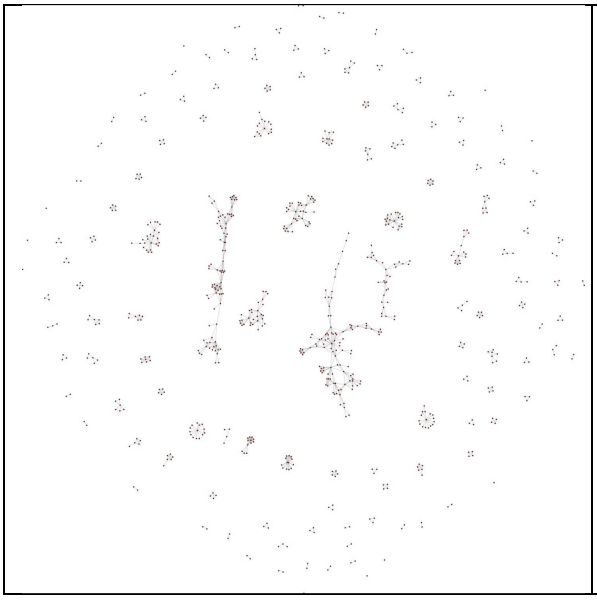

Figure 6c: Visualization of co-authorship network of Mechanical Engineering (MECH) WoS subject category. There is no distinct *well-formed* giant component seen as yet.

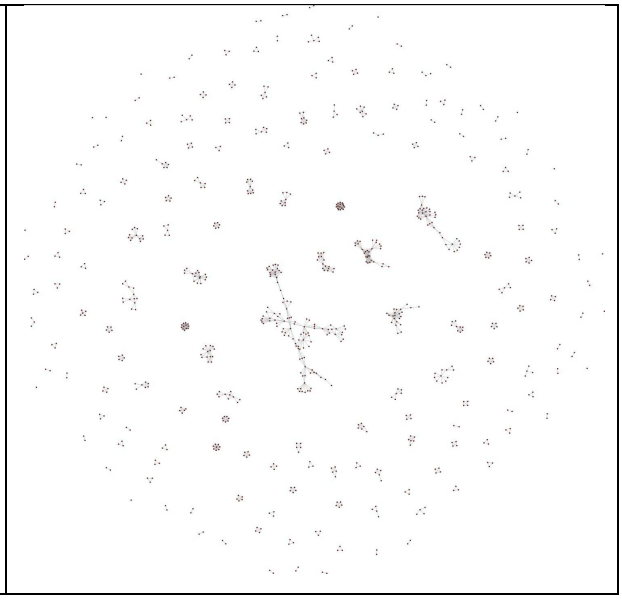

Figure 6d: Visualization of co-authorship network of Civil Engineering (CIVIL) WoS subject category. There is no distinct *well-formed* giant component seen as yet.

Table 2: Correlation Matrix of Various Graph Metrics

	Number of Nodes	Number of edges	Average Degree	Average Clustering Coefficient	Average geodesic Distance	Density	Size of giant component
Number of Nodes	1						
Number of edges	0.99	1					
Average Degree	0.87	0.92	1				
Average Clustering Coefficient	0.13	0.24	0.60	1			
Average geodesic Distance	0.95	0.94	0.78	0.04	1		
Density	-0.99	-0.99	-0.86	-0.10	-0.95	1	
Size of giant component	0.98	0.99	0.93	0.30	0.95	-0.97	1

Over the years there has been clear increase in the number of articles across all four disciplines (see Figure 7).

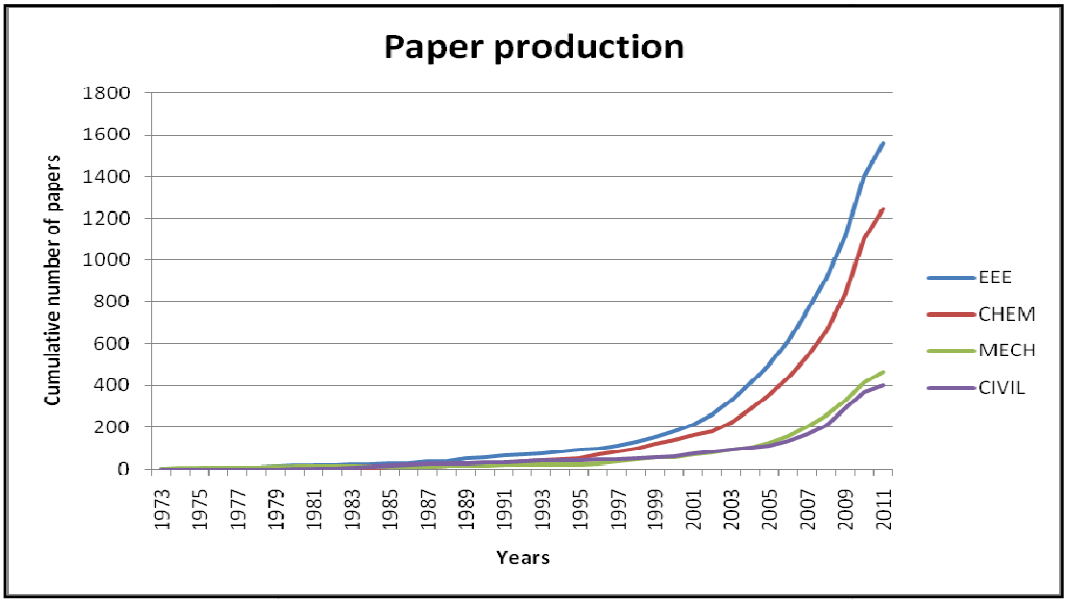

Figure 7: Cumulative Increase in Number of Research Articles in the Four Engineering Disciplines over Time.

(Till around mid-90s all disciplines were having similar paper production, after which EEE and CHEM have added papers faster than MECH and CIVIL. Corresponding to this faster paper production, giant Components have formed in EEE and CHEM, whereas they are still not evident in MECH and CIVIL)

All disciplines were almost in the same position until about mid-90s, after which both EEE and CHEM added articles faster than the other two. This greater proportion of increase in EEE and CHEM networks corresponds to the formation of giant components in these networks. There are authors who repeatedly write papers with their existing co-authors in

addition to the large proportion of new players who enter the scene. Increase in paper production, thus, directly increases in the number of nodes and edges in the network.

As stated earlier, we see a positive correlation between the number of nodes (and edges) in the network and the size of the giant component, within the context of these four engineering disciplines. However, looking from another perspective, just the existence of a large number of nodes (authors) in a network cannot be the sole reason for the formation of a giant component. For example, MECH has 809 nodes; yet, the largest component is just at 13.27% even after over three decades of activity. Even a very small network of just 48 researchers of COLLNET (Yin et al. 2006), a dedicated research forum of scientists studying scholarly collaboration networks, had a largest component possessing 32 nodes or 66.6% of the total network. Hence, just the presence of large number of nodes is no guarantee that a giant component would exist in such networks.

It may be that scientific network possessing a large number of nodes, but nodes working separately in diverse sub-disciplines, would still keep the network fragmented for a long time. Engineering disciplines have dedicated sub-disciplines. For example, mechanical engineering may have 'complex mechanics' and 'micro-mechanical science' as two separate divisions or sub-disciplines. In universities, these sub-disciplines are sometimes enshrined as separate departments within the faculty. Such categories within a discipline can lead to fragmentation as researchers generally have favorable circumstances to collaborate with fellow researchers within their research divisions.

One way to see faster formation of giant component is by fostering collaboration between these sub-disciplines. After all, it takes just one edge to bring two components or clusters of researchers together. Additionally, unlike random networks, collaboration in real-world networks, such as, co-authorship network, follows a certain pattern, also known as *preferential attachment* (Newman 2002). As such, some nodes attract connections by virtue of these nodes being already well connected or due to some other kind of *assortative mixing* (Newman 2002).

There seems to be no particular cause for the formation of giant components. Although, rise in the number of research articles or increase in collaboration among researchers might play an important role, they cannot be standalone reasons for the formation of giant components. Rather, a variety of causes working in tandem may be responsible for the formation of giant components.

Limitations

This study had a couple of limitations. Firstly, it is limited in its scope by including only four engineering subjects based on the number of papers in the WoS subject category. By incorporating more engineering disciplines we could have increased the breadth of this study. However, by limiting it to prominent few we were able to make a more in-depth comparative study.

Our second limitation is author name disambiguation. Author name disambiguation is a difficult and unresolved issue in bibliometrics (Garfield 1969; Tang and Walsh 2010). In bibliometric records, due to similarity of author names, there is a possibility that two or more authors may be represented as one author. Additionally, author name variations can make one author to be represented as two or more authors.

There have been several approaches to resolve this issue but they all suffer from some drawbacks or the other (*for a review* - Smalheiser and Torvik 2009). Tang and Walsh (2010) state that some studies simply avoid micro-level analysis, some indicate a method without elaborating on how author names issue is dealt with and others show results and analysis, but keep the authorship identification in the black box. Manual cleaning seems to be a solution to a certain degree, however, even manual disambiguation is a surprisingly hard and uncertain process, even on a small scale, and is entirely infeasible for common names (Smalheiser and Torvik 2009). Moreover, hand cleaning relies on institutional affiliation and full names, which is always a challenge. Even while using a standardized bibliometric database such as WoS, this is a perplexing issue. This is because, prior to 2007, WoS did not have a 'full author name' field and identification of author was based on only last name and initials. Also, while identifying authors with their institutional affiliations in WoS, one can never be certain if they exactly match, except for the correspondence address (Tang and Walsh 2010).

In this study, we have retained the data quality of WoS. One of the biggest advantages of using WoS database for bibliometric study is that it is pre-cleaned for errors and redundancy (Thomson Reuters 2012). WoS product mentions that it has "*met the high standards of an objective evaluation process that eliminates clutter and excess and delivers data that is accurate, meaningful and timely*". Regarding author identification, it mentions that "*… eliminating the problems of similar author names or several authors with the same name*".

Thomson Scientific, the publishers/aggregators of WoS, has their own internal disambiguation efforts on a massive scale (Smalheiser and Torvik 2009). Quality of WoS database at least ensures that we are dealing with clean database. However, it still does not completely eliminate the problem of name variations or other issues related to the dynamic nature of WoS records. But since ours is a macro study we believe these errors would be at the minimum and randomly distributed and would not significantly affect the overall picture of the network.

CONCLUSION

In this study, we empirically investigated the size of giant components in the collaborative networks of four prominent engineering disciplines in Malaysia. Our study found CHEM and EEE networks to already possess well-formed giant components, whereas MECH and CIVIL networks had not yet formed one. All four networks demonstrated small-world properties with networks with larger giant components having longer distance of separation between the nodes. Density of the nodes was negatively correlated with the size of the giant components. Although both degree of collaboration and clustering coefficient showed positive correlation with the size of giant components, the former showed a much stronger correlation than the latter.

Using temporal data, we found that till around the mid-90s, all the four disciplines had similar paper production. However, after this period, CHEM and EEE added papers faster than MECH and CIVIL. Corresponding to this activity, CHEM and EEE show well-formed giant components. Nonetheless, we also point out that just the presence of large number of nodes cannot be the only criteria for the formation of giant component. Rather a multitude of factors (i.e. addition of nodes and these nodes working in related sub-disciplines), may be instrumental in the faster formation of the giant component.

That research collaboration accrues quantifiable benefits is largely understood now. Giant component in a co-authorship network may represent core research activity within the academic community. With this study as an indicator, governments, universities and other bodies can make further efforts to foster, interdisciplinary, inter-sub-disciplinary, inter-institutional and multi-sector collaboration to bring researchers together.

REFERENCES

Abrizah, A. and Wee, M. C. 2011. Malaysia's Computer Science research productivity based on publications in the WoS, 2000-2010. *Malaysian Journal of Library & Information Science*, Vol. 16, no. 1: 109-124.

Barabasi, A. L. and Albert, R. 1999. Emergence of scaling in random networks. *Science,* Vol. 286, no.1: 509-512.

Cross, R., Liedtka, J. and Weiss, L. 2005. A practical guide to social networks. *Harvard Business Review*, Vol. 83, no.3: 124-132.

Fatt, C. K., Abu Ujum, E., and Ratnavelu, K. 2010. The structure of collaboration in the Journal of Finance. *Scientometrics,* Vol 85: 849-860.

Garfield, E. 1969. British quest for uniqueness versus American egocentrism. *Nature*, Vol. 223: 763.

Glänzel, W. and Schubert, A. 2005. *Analysing scientific networks through co-authorship*. Handbook of quantitative science and technology research, 257-276.

JISC. 2012. *ISI Web of Knowledge - Content Classification*. Available at: http://www.jisc-adat.com/adat/adat_plat_details.pl?ns_ADAT:PLAT_ID=1188923740-93:1188924071-36

Kang, I. S., Na, S. H., Lee, S., Jung, H., Kim, P., Sung, W. K. and Lee, J. H. 2009. On co-authorship for author disambiguation. *Information Processing and Management*, Vol. 45, no.1: 84-97.

Katz, J. S. and Martin, B. R. 1997. What is research collaboration? *Research Policy*, Vol 26, no. 1: 1-18.

Knoke, D. and Kuklinski, J.H. 1982. Network analysis. *Sage University paper Series on Quantitative Applications in the Social Sciences, nr. 07-028*.

Kretschmer, H. 2004. Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. *Scientometrics*, Vol.60, no.3: 409-420.

Leydesdorff, L. and Rafols, I. 2009. A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, Vol. 60: 348-362.

Liu, X. M., Bollen, J., Nelson, M. L. and Van de Sompel, H. 2005. Co-authorship networks in the digital library research community. *Information Processing and Management*, Vol. 41, no. 6: 1462-1480.

Molloy, M. and Reed, B. 1995. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*,Vol. 6, no. 23: 161-180.

Newman, M. 2002. Assortative mixing in networks. *Physical Review Letters*, Vol. 89, no.20.

Newman, M. 2003. The structure and function of complex networks. *Siam Review*, Vol. 45, no.2: 167-256.

Newman, M. 2004a. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, Vol 101: 5200-5205.

Newman, M. 2004b. Who is the best connected scientist? A study of scientific coauthorship networks. *Complex Networks*, Vol 650: 337-370.

Newman, M. 2007. *The mathematics of networks*. The new palgrave encyclopedia of economics, Basingstoke: Palgrave Macmillan.

Otte, E. and Rousseau, R. 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, Vol. 28: 441-453.

Pepe, A. 2008. Socio-epistemic analysis of scientific knowledge production in little science research. *tripleC*, Vol. 6, no.2: 134-145.

Racherla, P. and Hu, C. 2010. A Social Network Perspective of Tourism Research Collaborations. *Annals of Tourism Research*, Vol. 37, no.4: 1012-1034.

Thomson Reuters. 2012. *Web of Science*. Available at: http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/

Sci[2,] T. 2009. *Science of Science (Sci[2]) tool: Indiana University and SciTech Strategies*. Available at: http://sci.slis.indiana.edu.

Smalheiser, N.R. and Torvik, V.I. 2009. Author name disambiguation. *Annual Review of Information Science and T echnology*, Vol. 43: 1-43.

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. *Fourth International Conference on Communities and Technologies*, ACM, 255-264.

Tang, L. and Walsh, J. P. 2010. Bibliometric fingerprints: name disambiguation based on approximate structure equivalence of cognitive maps. *Scientometrics*, Vol 84: 763-784.

Wasserman, S. and Faust, K. 1994. *Social Network Analysis, Methods and Applications* (First edition ed.): Cambridge University Press.

Watts, D. J. and Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature*, Vol. 393, no. 6684: 440-442.

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 and Management*, Vol. 42, no.6: 1599-1613.