Analyzing research collaborations within the School of Mathematical Sciences, UKM using Graph Theory

Fatimah Abdul Razak*, Faridatulazna Ahmad Shahabuddin and Nik Sarah Nik Zamri
School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia

E-mail: fatima84@ukm.edu.my*

Abstract. A graph (comprised of vertices and edges) mapping out the collaboration between 38 lecturers in the School of Mathematical Sciences is obtained for years 2012-2014. The graph is then studied to understand the nature of collaborations within the school as well as identifying the importance of different vertices in the network. Methods from social network analysis and network theory will also be applied to the graph in addition to classic graph theory measures.

1. Introduction
Graph theory is the study of graphs, which are mathematical abstractions called vertices (also called nodes or actors). Each related pairs of vertices is connected by an edge (also called link or relations). Essentially, graph are structures used to model relations between objects. Typically, a graph is depicted in diagrammatic form as a set of dots for the vertices, joined by lines or curves for the edges. In more recent literatures graphs are also known as networks. Graph theory is utilized in to analyse real networks. One particularly popular form of this analysis is the social network analysis which is the analysis of social structure consisting of a set of players (vertices) and a set of relationships (edges) between the players. Social network analysis examines the structure of social relationships in a group to reveal the structure and key players of the group. The ubiquitous-ness of Facebook, which is a clear social network formed by users (vertices) and their friendships (edges) have resulted in targeted ads for certain groups and even political manipulation in the case of Cambridge Analytica [1]. Facebook itself, uses social network analysis to closely monitor Facebook users and their relationships.

In this article, we shall do a simple network analysis of research collaborations quantified by joint publications (edges) amongst researchers (vertices) within the School of Mathematical Sciences (PPSM), Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM) between years 2012 and 2014.

2. Graph Theory and Networks
A graph or network can be written as $G = (V, E)$, where $V$ is the set of vertices and $E$ is the set of edges. A simple graph is a graph that has neither self-edges nor multiedges (multiple edges between two vertices). Figure 1 is an example of a simple graph. A graph can be represented by a matrix where each row and column represents a vertex. In a simple graph, if two vertices has an edge connecting them, then the corresponding element of the matrix will have value of 1. Otherwise the corresponding element of the matrix will be zero. This matrix is called an adjacency matrix. Table 1 displays an adjacency matrix that represents the graph in Figure 1.
Figure 1. A simple graph

Table 1. The adjacency matrix representing the graph in Figure 1

| Vertices | 1 | 2 | 3 | 4 | 5 | 6 |
|----------|---|---|---|---|---|---|
| 1        | 0 | 1 | 1 | 1 | 1 | 1 |
| 2        | 1 | 0 | 1 | 1 | 0 | 0 |
| 3        | 1 | 1 | 0 | 0 | 0 | 1 |
| 4        | 1 | 1 | 0 | 0 | 0 | 0 |
| 5        | 1 | 0 | 0 | 0 | 0 | 0 |
| 6        | 1 | 0 | 1 | 0 | 0 | 0 |

If Figure 1 represented the network of research collaboration within PPSM, then one could say that researcher 1 (represented by vertex 1) has joint publications with all other researchers in the network. This can also be observed by looking at column or row 1 in Table 1 which has the value 1 for all elements except for the first row. Note that for a simple matrix the diagonals must all be zero since there are no self-loops. Observe that researcher 5 only has a single joint publication with researcher 1 since entry (1,5) and (5,1) are the only nonzero entries in row and column 5 of Table 1. The degree of a vertex in a graph is the number of edges connected to it. We will denote the degree of vertex $i \in V$ by $k_i$. $k_i$ can be obtained from the adjacency matrix in Table 1 by summing the values of any single row or column. By summing up rows in Table 1, we get $k_1 = 5, k_2 = 3, k_3 = 3, k_4 = 2, k_5 = 1$ and $k_6 = 2$, which can easily be verified in Figure 1.

A path in a network is any sequence of vertices such that every consecutive pair of vertices in the sequence is connected by an edge in the network [2]. The distance $d_{ij}$ between two vertices, $i \in V$ and $j \in V$, is defined as the number of edges along the shortest path connecting them. Every two vertices connected by an edge has distance 1, therefore the graph in Figure 1 has $d_{12} = d_{13} = d_{36} = 1$ and so on. However there are no direct edges between quite a few vertices therefore the shortest path between these vertices will be longer. For example to connect vertex 5 to any other vertices excluding 1, we need to go through vertex 1 (since $d_{15} = 1$), therefore the distance becomes 2 such that $d_{25} = d_{35} = d_{45} = d_{65} = 2$. The diameter of a graph is the length of the longest shortest path between any pair of vertices in the graph for which a path actually exists [2-5]. The diameter of the graph in Figure 1 is 2.

The concept of centrality originated from the discipline of social network analysis. Centralities measures rank vertices according to their perceived importance. Perhaps the simplest centrality measure in a network is just the degree of a vertex [2,3]. For the graph in Figure 1, vertex 1 will have the highest ranking since it has the largest degree ($k_1 = 5$), in second place will be vertices 2 and 3 ($k_2 = k_3 = 3$).
This will be followed by vertices 4 and 6 ($k_4 = k_6 = 2$) and the lowest ranking will go to vertex 5 since $k_5 = 1$.

Some centralities depend on the distance between vertices such as the closeness centrality. The closeness centrality of vertex $i \in V$ can be defined as the reciprocal of farness [4], such that

$$l_i = \frac{1}{\sum_{j \in V} d_{ij}}. \quad (1)$$

It uses the inverse sum of distances from a vertex to all other vertices in the graph to rank the vertex. The larger the closeness centrality value the ‘closer’ the vertex is to all other vertices. The betweenness centrality of vertex $i \in V$ can be defined [3,5] as

$$x_i = \sum_{s \neq i \in V} \sum_{t \neq i \in V} \frac{n_{st}(i)}{N_{st}} \quad (2)$$

where $n_{st}(i)$ is the number of shortest paths from $s$ to $t$ that passes through vertex $i$. $N_{st}$ is the total number of shortest paths from $s$ to $t$ (since there may be more than one shortest path between any two vertices). Other definitions and formulas are also used to define closeness and betweenness centrality [2,3,6].

3. Collaboration Network

![Graph representing PPSM researchers’ joint publication for years 2012-2014.](image-url)
Table 2. The adjacency matrix representing PPSM researchers’ joint publication for years 2012-2014

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Note: The adjacency matrix represents the co-authorship network of PPSM researchers for the years 2012-2014. Each entry in the matrix indicates the number of joint publications between two researchers. A value of 1 indicates a joint publication, and a value of 0 indicates no joint publication.
The 2012 to 2014 data of PPSM’s researchers and the number of publications has been obtained from the e-repository UKM. This data was manually tabulated to obtain a combined adjacency matrix for collaborations throughout the years 2012-2014 as displayed in Table 2. The graph or network obtained from the adjacency matrix is displayed in Figure 2. This graph represents the collaboration between PPSM researchers within years 2012-2014. This graph can be represented by $G = (V, E)$, where $V = \{1, 2, ..., 38\}$ is the set of vertices representing researcher 1-38 in PPSM and $E$ is the set of edges representing joint publications between these 38 researchers. The average degree of this graph is 8.4737 implying that on average each researcher in PPSM has around 8 joint publication with others in PPSM. The diameter of this graph is 5.

Table 3. Top 10 vertices ranked by degree $k_i$, closeness centrality $l_i$ and betweenness centrality $x_i$

| Rank | Vertex $i$ | $k_i$ | Vertex $i$ | $l_i$ | Vertex $i$ | $x_i$ |
|------|------------|------|------------|------|------------|------|
| 1    | 7          | 19   | 7          | 0.017544 | 36          | 147.4328 |
| 2    | 28         | 19   | 28         | 0.017544 | 15          | 76.47425 |
| 3    | 27         | 17   | 27         | 0.016949 | 28          | 66.00433 |
| 4    | 9          | 16   | 36         | 0.016667 | 7           | 63.41177 |
| 5    | 24         | 16   | 24         | 0.015873 | 27          | 43.83874 |
| 6    | 36         | 16   | 26         | 0.015625 | 6           | 36.49359 |
| 7    | 26         | 15   | 9          | 0.015385 | 17          | 36.29167 |
| 8    | 33         | 15   | 15         | 0.014925 | 25          | 36     |
| 9    | 35         | 14   | 34         | 0.014925 | 9           | 34.84426 |
| 10   | 34         | 13   | 33         | 0.014286 | 20          | 30.48003 |

We rank the vertices using degree $k_i$, closeness centralities $l_i$ in equation (1) and betweenness centrality $x_i$ in equation (2) using the MATLAB Graph and Network Algorithms toolbox. The top 10 vertices for each ranking is listed in Table 3. The top ranking vertex for $k_i$ and $l_i$ is vertex 7, 27 and vertex 28. However these vertices are ranked below vertices 36 and 15 when using $x_i$. This may be due to the fact that vertices 7, 27 and 28 are researchers that are active both in core mathematics and mathematics education where there exist some large scale publications that have many authors (some up to 15 authors) amongst researchers in the PPSM. Vertices 36 and 15 may also be involved in these publications but they also have diverse publications with various individuals not involved in these large scale publications.

While the top 10 vertices ranked by degree and closeness centrality are rather similar, betweenness centralities highlights a few vertices that are not in the top 10 rank of the others such as vertices 6, 17, 25 and 20. These vertices are researchers that may not have many collaborations by publication but they have a few joint publications that are keeping the network together.
4. Discussions and Conclusion
This paper utilizes graph theory and network science to analyse the collaboration within researchers in PPSM, UKM for years 2012-2014. The graph in Figure 2 depicts the collaboration networks between PPSM researchers and the top ranked vertices are highlighted in Table 3. In general the level of collaboration amongst PPSM researchers within these years is rather high with an average degree of more than 8. Certain researchers are collaborating more than others as highlighted by degree and closeness centrality. However betweenness centrality highlights vertices that serves as vital connections in keeping the network together despite not having as many joint collaborations as others.

Perhaps researchers should not only be judged by how many paper they produce but also by the number of collaborations within their own organization and between groups in the organization which paints them as a good team player. We have recently obtained complete datasets for all publications by UKM Staff which will be the subject of future research. We plan to analyse UKM-wide research collaboration data to highlight the structure of research collaboration within the university and help encourage multidisciplinary studies across different disciplines.

Acknowledgments
We are grateful for the grants FRGS/1/2015/SG04/UKM/02/1 and KRA-2017-014 for funding our research.

References
[1] Ghosh D and Scott B (2018, March 19) Facebook’s New Controversy Shows How Easily Online Political Ads Can Manipulate You. Retrieved from http://time.com
[2] Newman M E J 2010 Networks An Introduction (New York: Oxford University Press Inc)
[3] Otte E and Rousseau R 2002 Journal of Information Science 28 (6) 441-453
[4] Bavelas A 1950 Journal of the Acoustical Society of America 22 (6)
[5] Freeman L C 1977 Sociometry 40 (1) 35-41
[6] Rasidi N M, Bakar S A, Razak FA 2016 AIP Conference Proceedings 1750 030024