ASM Sc. J., 19, 2024
https://doi.org/10.32802/asmscj.2023.1485

An Analysis on Research Collaboration Between Mathematicians in Universiti Putra Malaysia

U.S. Mohd Zu[1] and K.S. Yow[1,2*]

[1]*Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[2]*School of Computer Science and Engineering, College of Engineering, Nanyang Technological University, Singapore*

Social network analysis is commonly used to investigate relationships between a group of objects by modelling them using graphs. Some of its applications include network modelling and sampling, link prediction and social media analytics. In this study, we analyse research collaboration in between 44 researchers in the Department of Mathematics and Statistics, Universiti Putra Malaysia within two periods, from 2015–2017 and 2018–2020, by using social network analysis. We identify the importance of each researcher within the community by using different centrality measures such as degree, closeness, betweenness and eigenvector centralities, based on the collected data. The ten highest values in each measure for both durations are given. The graphs for each centrality measure are plotted, and a comparison on the relevant values across the two periods is made. We also discuss the relationship between each centrality measure, as well as the possible factors that affect the respective values.

Keywords: network analysis; collaboration graph; centrality measure; ranking

I. INTRODUCTION

Graphs (Harary, 1969) are important in our daily life with numerous applications as it can be used to portray the relationship between a group of objects. Its applications can be found in many disciplines, including network modelling, information system development analysis and community-based problem solving (Yow & Luo, 2022).

A graph[1] $G = (V, E)$ consists of a set of *vertices* (or *nodes*) and a set of *edges* (or *lines*), denoted by $V(G)$ and $E(G)$, respectively. Two vertices are said to be *adjacent* if they share one common edge. An *adjacency matrix* of a simple graph G of order n is an $n \times n$ matrix A in which the rows and columns of A are both indexed by the vertices of G, and $A(G) = \left(a_{ij}\right)_{n\times n}$ where $a_{ij} = 1$ if vertices i and j are adjacent, and $a_{ij} = 0$ otherwise. There are various types of graphs such as undirected graphs and directed graphs. For *undirected* graphs (*graphs*), two vertices are connected by an edge that has no direction. For *directed* graphs (*digraphs*), a direction will be assigned on each edge that connects two vertices.

Social network analysis (Wasserman & Faust, 1994; Martino & Spoto, 2006) is a study in investigating social structures and relationships between a group of objects by using concepts in graph theory. Those objects are modelled by using graphs where vertices represent objects and edges indicate the relationship among certain pairs of objects. The objective in studying social network analysis is to observe and interpret the structure of a network that is determined by a group of objects, which can be carried out through visualisation and mathematical analysis. Social network analysis provides one of the most definite representations to figure out the possible communications between objects and the way those objects are connected in the network. Its applications include network modelling and sampling, link prediction and social media analytics.

[1] The term *graph* and *network* are used interchangeably throughout.

*Corresponding author's e-mail: ksyow@upm.edu.my

The relationship between graph theory and social network analysis can be found in many recent studies (Razak *et al.,* 2019). Social network analysis has obtained broad recognition, big growth and outstanding utilisation in various fields since late 1970s (Zaphiris & Pfeil, 2007). In social network analysis, creating model of networks, managing vital character of networks and applying applicable methods in computational analysis of networks are the essential research concerns. There are various types of social network analysis measures (Bavelas, 1950; Freeman, 1977; Newman, 2008) including degree centrality, betweenness centrality, closeness centrality and eigenvector centrality (detailed definitions of these measures can be found in the following section).

Graph theory and network analysis are closely related (Pachayappan & Venkatesakumar, 2018) and broadly identified but the disposition of their relationship is rarely conferred. The development of graph theory was at a slow pace, and the first application of graph theory to network analysis was found in 1953. Graph theory evolved in network analysis including social network analysis in 1983 (Barnes & Harary, 1983). Social network analysis is also a tool to analyse connections and synergy between people by using the concept of graph theory (Kocak, 2014). Apart from the social network analysis measures mentioned earlier, there are also a few parameters that can be found in graph theory in identifying characteristics of contrasting networks such as the diameter, shortest path, clustering coefficient and geodesic. Social network analysis is an effective mean to analyse the changeful trend in all sizes and types of groups (Bandyopadhyay *et al.,* 2011). Tichy *et al.* (1979) published one of the earliest writings regarding the importance of social network analysis in organisations. They proposed that theory and research could be improved critically and suggested to apply social network analysis to organisations in their work since this proposal would assist in analysing the comparative aspects of organisations and also subunits in the organisations (Nunes & Abreu, 2020).

In this study, our aim is to analyse research collaboration in between 44 researchers in the Department of Mathematics and Statistics, Universiti Putra Malaysia (UPM) by using social network analysis. Our focus is on the research collaboration between the researchers within two periods, from 2015-2017 and 2018-2020. Our contributions are summarised as follows: (1) The relevant datasets for the 44 researchers are collected and analysed; (2) We made a comparison between the two datasets and identify the importance of each researcher based on different centrality measures; (3) Graphs represent the relationships between the researchers using different measures are plotted, and relevant discussions and analyses are also given.

II. MATERIALS AND METHOD

In this study, a *research collaboration* between two researchers exists if they both co-authored (other co-authors may present) a paper in a conference, journal, book chapter or book. A survey involving 44 researchers in the Department of Mathematics and Statistics, UPM is conducted by looking up the information through online sources such as ResearchGate and Google Scholar. To improve the accuracy of the dataset, majority of the collected data have been confirmed by the respective researchers.

The collected data are converted into an adjacency matrix that has a dimension of 44×44. Each researcher is assigned with an ID that is used as the row and column indices of the matrix. The matrix is then used to construct collaboration graphs where vertices are used to represent researchers and edges are used to represent collaborations between the researchers.

A. Graph Construction

We construct all relevant graphs using *Gephi*[2] (Bastian *et al.,* 2009), by first importing two files A and B that contain the relevant data. The first file A contains two columns, one for ID and one for label. The IDs are used to represent researchers and the labels are the names of the researchers. The second file B contains three columns, namely *source*, *target* and *type*. The source and the target represent a researcher and his/her collaborator, respectively, and they form an edge. The type of each edge is undirected in our

[2] Gephi is an open-source network analysis and visualisation software.

study. Both files A and B need to be appended in the same workspace, to avoid potential data error in Gephi.

We analyse the graphs constructed based on different centrality measures. In each graph, the size of each vertex associates with the degree of the vertex where a bigger vertex represents a vertex of higher degree. The colour of each vertex associates with the values of the betweenness, closeness and eigenvector centralities. The darker the colour of a vertex, the higher its value is.

B. Network Analysis Measures

We compare and analyse the graphs obtained using four centrality measures, which are degree centrality, betweenness centrality (Freeman, 1977), closeness centrality (Bavelas, 1950) and eigenvector centrality (Newman, 2008). The relevant definitions are given formally, as follows:

Definition 1. *Given a graph $G(V,E)$, and let $v \in V$. The* degree centrality *of v is defined as the number of edges that are incident to v.*

Note that degree centrality is the simplest measure among others, which measures the popularity of a vertex according to the number of its neighbouring vertices.

Definition 2. *Given a graph $G(V,E)$ and let $u \neq v \in V$. The* closeness centrality *(or* closeness*) of v* is defined as:

$$C(v) = \frac{1}{\sum_u d(u,v)} \tag{1}$$

where $d(u,v)$ is the distance from u to v.

Closeness centrality of a vertex is defined as the average length of the shortest path between the vertex and all other vertices in a network, which is the reciprocal of the farness. The higher the value of the closeness of a vertex, the more central the vertex will be. Thus, the vertex will be the quickest to receive any information that appears in the network. For closeness, people usually refer to its normalised form where the numerator in Equation 1 is replaced by $n-1$, where n is the total number of vertices in the graph, as shown below:

$$\boldsymbol{C(v) = \frac{n-1}{\sum_u d(u,v)}}.$$

Definition 3. *Given a graph $G(V,E)$, let $s,t,v \in V$ and $s,t \neq v$. The* betweenness centrality *of v* is defined as:

$$\boldsymbol{B(v) = \sum_{s,t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}}$$

where σ_{st} represents the total number of shortest paths from s to t, and $\sigma_{st}(v)$ represents the number of shortest paths from s to t that pass through v.

The betweenness centrality of a vertex quantifies the count of the vertex lying along the shortest paths between two other vertices.

Definition 4. *Let $A = (a_{u,v})$ be the adjacency matrix of a graph $G(V,E)$ and $u,v \in V$. The eigenvector centrality of v is defined as:*

$$\boldsymbol{x_v = \frac{1}{\lambda} \sum_{u \in M_v} x_u = \frac{1}{\lambda} \sum_{u \in G} a_{u,v} x_u}$$

where λ is a constant and $M(v)$ represents the set of neighbours of v.

Eigenvector centrality measures the influence of a vertex in a network. The concept is that connections to high-scoring vertices give more impact and hence a higher score to a vertex. One variant of the eigenvector centrality is the PageRank algorithm (Page *et al.,* 1999) used by Google in measuring the importance of each webpage.

III. RESULTS AND DISCUSSION

In general, we found that the total numbers of research collaboration between the 44 researchers increase from 80 in 2015-2017 to 133 in 2018-2020, based on the data collected. The distribution of the number of collaborators for the two periods is illustrated in Figure 1.

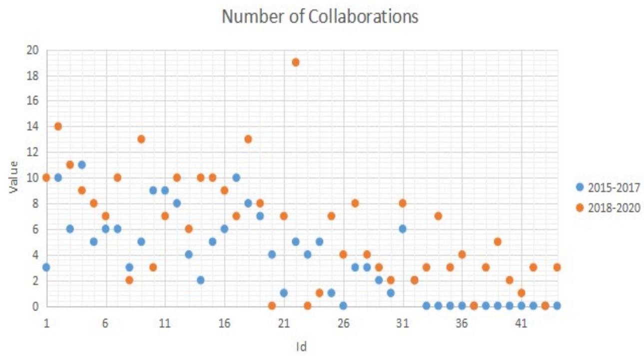


Figure 1. The distribution of the number of collaborators for each researcher during the two periods

By comparing the number of collaborations for each researcher during the two periods by using Figure 1, we can see that 75% of the researchers have more collaborators in the second period compared to the first period, whereas 20.45% of them have a lesser number of collaborators in the second period. Among them, Researcher 22 has the highest increment in terms of the number of collaborators. The number of increments is 12, which is equivalent to 22.64% of the total increment in the second period.

We now present the two tables that show the ten highest values for each centrality measure, during the two periods. We use Gephi 0.9.2 in calculating the values for each centrality measure.

Table 1. Vertices with ten highest values for each centrality measure in 2015-2017

	Degree		**Closeness**		**Betweenness**		**Eigenvector**	
Rank	ID	Value	ID	Value	ID	Value	ID	Value
1	4	11	4	0.5882	4	181.2464	2	1.0000
2	2, 17	10	2	0.5085	28	81.0000	17	0.9296
3	10, 11	9	17	0.5000	18	73.5190	4	0.8912
4	12, 18	8	18	0.4918	11	52.6845	11	0.7883
5	19	7	11	0.4839	17	44.2786	10	0.7616
6	3, 6, 7, 16, 31	6	6	0.4762	2	33.8063	19	0.7314
7	5, 9, 15, 22, 24	5	9, 10	0.4688	12	30.7516	6	0.7120
8	13, 20, 23	4	5, 12	0.4412	1, 27	29.0000	7, 16	0.6236
9	1, 8, 27, 28	3	7, 16	0.4348	10	24.8750	12	0.6200
10	14, 29, 32	2	3, 19	0.4225	5	23.1826	18	0.5053

Table 2. Vertices with ten highest values for each centrality measure in 2018-2020

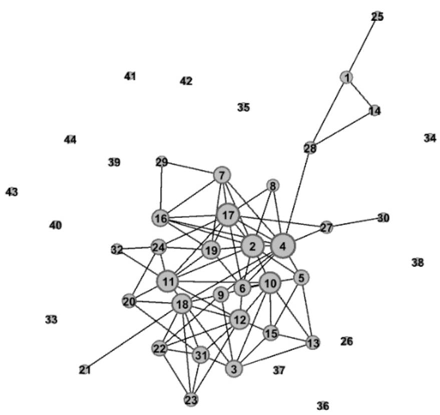
	Degree		**Closeness**		**Betweenness**		**Eigenvector**	
Rank	ID	Value	ID	Value	ID	Value	ID	Value
1	22	19	22	0.6290	22	171.6223	22	1.0000
2	2	14	18	0.5652	9	84.3761	9	0.8826
3	9, 18	13	2	0.5571	14	71.0057	14	0.7225
4	3	11	9	0.5493	18	67.8996	18	0.6448
5	1, 7, 12, 14, 15	10	3	0.5417	3	65.7576	3	0.6435
6	4, 16	9	7, 12	0.5342	2	65.2622	2	0.6340
7	5, 19, 27, 31	8	14	0.5200	11	58.8464	11	0.5726
8	6, 11, 17, 21, 25, 34	7	4	0.5065	7	55.8689	7	0.5477
9	13	6	1, 34	0.5000	15	52.6485	15	0.5292
10	39	5	5, 27, 31	0.4875	27	43.6182	27	0.4990

All the values listed in Tables 1 and 2 can be obtained once the respective graphs are constructed based on the data imported into Gephi. These values are arranged in descending order so that the ten highest values for each centrality measure can be easily identified and compared. The associated ID of the respective researcher is also provided. From the tables, we can see that some researchers shared a common value, especially for degree centrality. This implies that this group of researchers have the same attribute, i.e., the same number of collaborators during the same period.

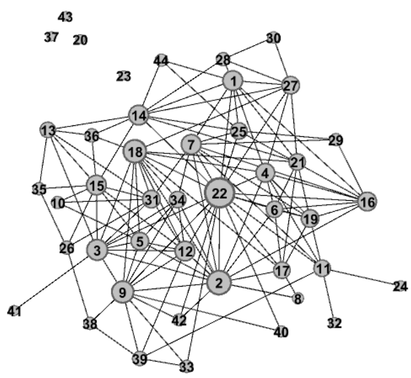
A. Four Centrality Measures

We now provide the graphs constructed according to four centrality measures and give the relevant analyses.

For **degree centrality**, we can see that Researcher 4 ranks the highest in 2015-2017 with 11 collaborators, whereas Researcher 22 with 19 collaborators ranks first in 2018-2020. The rank for Researcher 2 remains throughout the two periods, with 10 and 14 collaborators, respectively. In general, the numbers of collaborations between researchers show an increasing trend, which imply that most of the researchers become more active in participating in various research projects within the department. This can also be evidenced based on the two graphs in Figure 2, where more connections can be found between vertices in Figure 2(b).



(a) 2015-2017

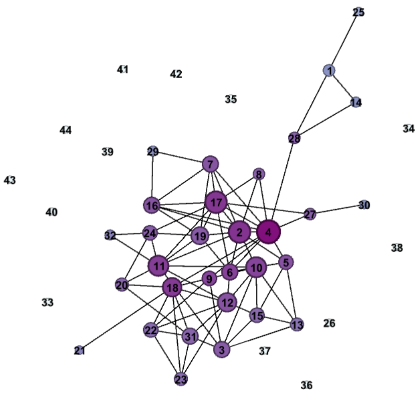


(b) 2018-2020

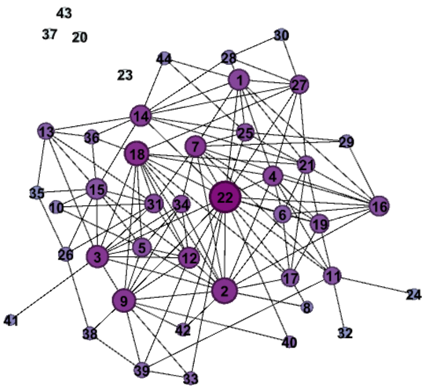
Figure 2. Two graphs that represent degree centrality for each vertex

Based on the two graphs, it is also clear that the numbers of isolated vertices reduce significantly by comparing the two periods. This show that most of the researchers have started to build up collaborations among themselves by working on related problems.

For **closeness centrality**, we can observe that vertices with higher degrees are also among the top in closeness centrality. Specifically, the two researchers, Researcher 4 and Researcher 22, with the highest degree centrality in both periods have the highest closeness centrality in the respective period. This could be due to the structure of the graph itself where vertices with many neighbours are more accessible from others. In comparing closeness for both periods, collaborations between researchers are much stronger in 2018-2020, which also imply that information and resources could be spread more effectively among researchers. The graphs that demonstrate closeness centrality for both periods are shown in Figure 3, where a darker vertex indicates a higher closeness centrality.



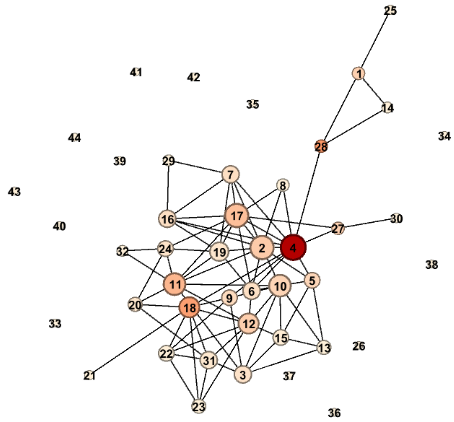
(a) 2015-2017



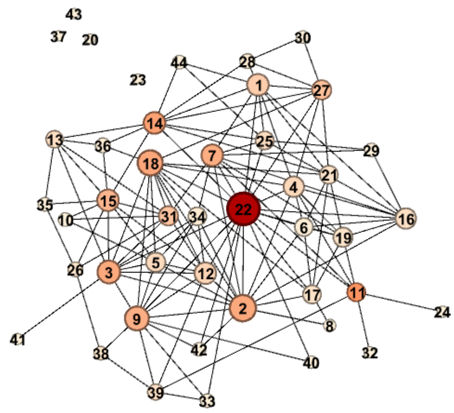
(b) 2018-2020

Figure 3. Two graphs that represent closeness centrality for each vertex

For **betweenness centrality**, we observe that researchers with more collaborators show higher merits. There is however some exception where Researcher 28 that ranks ninth (with three collaborators) in 2015-2017 has the second highest betweenness centrality. The reason could be Researcher 28 plays a crucial role in connecting two different research groups within the department, since it is a cut vertex in the graph (see Figure 4(a) for example) where every shortest path that connects vertices from the two groups must pass through it. We can hence conclude that degree centrality generally affects the value of the betweenness centrality, but it is not the only factor. The two relevant graphs for betweenness centrality can be found in Figure 4, where Researcher 4 and Researcher 22 with the darkest colour give the highest betweenness centrality in 2015-2017 and 2018-2020, respectively.



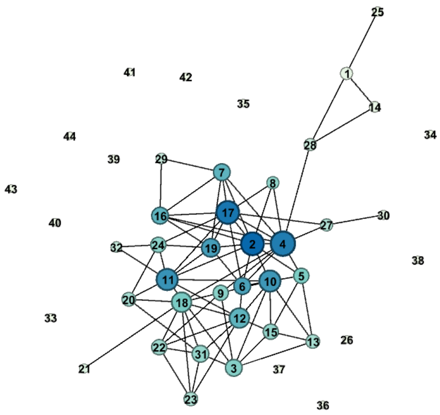
(a) 2015-2017



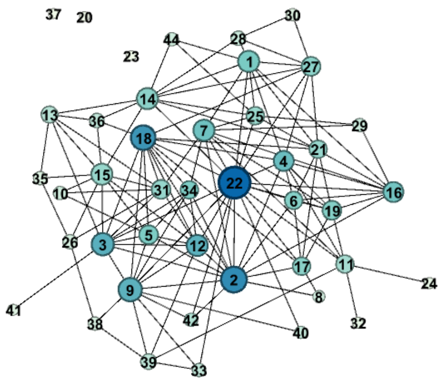
(b) 2018-2020

Figure 4. Two graphs that represent betweenness centrality for each vertex

Lastly, for **eigenvector centrality**, the trends are completely different in both periods. In 2015-2017, the values do not give a clear indication how they are related to other centrality measures. Nonetheless, in 2018-2020, we can see that the values for eigenvector centrality is linked directly to the values of betweenness centrality. In general, researchers with more collaborators during this period tend to have a higher eigenvector centrality given that they are more likely to link with high-scoring researchers. This is somehow important to increase the significance of their roles within a research community. The graphs that illustrate eigenvector centralities for all researchers in both periods are shown in Figure 5.



(a) 2015-2017



(b) 2018-2020

Figure 5. Two graphs that represent eigenvector centrality for each vertex

IV. CONCLUSION

In this study, we analyse and compare the research collaborations between researchers in the Department of Mathematics and Statistics, UPM within two periods, 2015-2017 and 2018–2020. We identify researchers who play an important role in the department, based on various centrality measures in social network analysis. We found that some centrality measures are related to each other, but they are not the only indicators to rank the importance of each researcher. We believe that different centrality measures could be used to strengthen research collaborations and enhance the weaknesses within a community, based on different characterisations. We also hope that these results can be served as a reference in analysing research collaborations in some other areas.

For future work, we could extend this study by including researchers from different disciplines in order to identify the connections and relationships between different fields. The field that plays the most crucial role in connecting different disciplines could also be identified, which helps in boosting research collaborations in university or even national levels. To obtain a comprehensive analysis, some other centrality measures could also be included. Since the structures and polynomials of graphs are correlated for certain classes of graphs (Yow *et al.*, 2021), instead of focusing on centrality measures, one could examine if graph polynomials can be used in achieving the same goal.

V. ACKNOWLEDGEMENT

The authors would like to thank anonymous reviewers for their insightful comments and suggestions. This research was partially supported by Universiti Putra Malaysia under Putra Grant GP-IPM/2020/9684300 and National Research Foundation Singapore.

VI. REFERENCES

Bandyopadhyay, S, Rao, AR & Sinha, BK 2011, 'Model for Social Networks with Statistical Applications', Sage.

Barnes, JA & Harary, F 1983, 'Graph Theory in Network Analysis', Social Networks, vol. 5, no. 2, pp. 235-244.

Bastian, M, Heymann S & Jacomy, M 2009, 'Gephi: An Open Source Software for Exploring and Manipulating Networks', International AAAI Conference on Weblogs and Social Media, vol. 3, no. 1, pp. 361-362.

Bavelas, A 1950, 'Communication Patterns in Task-Oriented Groups', The Journal of the Acoustical Society of America, vol. 22, no. 6, pp. 725-730.

Freeman, LC 1977, 'A Set of Measures of Centrality Based on Betweenness', Sociometry, vol. 40, no. 1, pp. 35-41.

Harary, F 1969, Graph Theory. Addison Wesley Longman Publishing Company, New York.

Kocak, NG 2014, 'Social Networks and Social Network Analysis', International Journal of Business and Social Science, vol. 5, no. 2, pp. 126-134.

Martino, F & Spoto, A 2006, 'Social Network Analysis: A Brief Theoretical Review and Further Perspectives in the Study of Information Technology', PsychNology Journal, vol. 4, no. 1, pp. 53-86.

Newman, MEJ 2008, 'The Mathematics of Networks', The New Palgrave Encyclopedia of Economics, vol. 2, no. 2008, pp. 1-12.

Nunes, M & Abreu, A 2020, 'Applying Social Network Analysis to Identify Project Critical Success Factors', Sustainability, vol. 12, no. 4, pp. 1503.

Pachayappan, M & Venkatesakumar, R 2018, 'A Graph Theory Based Systematic Literature Network Analysis', Theoretical Economics Letters, vol. 8, no. 5, pp. 960-980.

Page, L, Brin, S, Motwani, R & Winograd, T 1999, The PageRank citation ranking: Bringing order to the web, Stanford InfoLab.

Razak, F, Shahabuddin, FA & Zamri, NSN 2019, 'Analyzing Research Collaborations within the School of Mathematical Sciences, UKM using Graph Theory', in Journal of Physics: Conference Series, 1212(1), 012033.

Tichy, NM, Tushman, ML & Fombrun, C 1979, 'Social Network Analysis for Organizations', Academy of Management Review, vol. 4, no. 4, pp. 507-519.

Wasserman, S & Faust, K 1994, 'Social Network Analysis in the Social and Behavioral Sciences', Social Network Analysis: Methods and Applications, pp. 1-27.

Yow, KS & Luo, S 2022, 'Learning-Based Approaches for Graph Problems: A Survey', Preprint. arXiv:2204.01057

Yow, KS, Morgan, K & Farr, G 2021, 'Factorisations of Greedoid Polynomials of Rooted Digraphs', Graphs and Combinatorics, vol. 37, pp. 2245-2264.

Zaphiris, P & Pfeil, U 2007, 'Introduction to Social Network Analysis', in Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK, 2, 231-232.