Examining the UK Higher Education sector through the network of institutional accounts on Twitter

Katy Jordan

Institute of Educational Technology, The Open University, Walton Hall, Milton Keynes, UK, MK7 6AA

katy.jordan@open.ac.uk

Abstract

Web link mining has been previously used as a way of gaining insight into how the Internet may be replicating or reshaping connections between institutions within the Higher Education sector. Institutions are increasingly active on social media platforms, and these connections have not been studied. This paper presents an exploratory analysis of the network of UK Higher Education institutional accounts on Twitter. All UK institutions have a presence. Standing in recent university rankings is found to be a significant predictor of several network metrics. In examining the communities present within the network, a combination of ranking and geolocation play a role. Analysis of a sample of tweets which mention more than one UK Higher Education institution provides an indication of why the topics of tweets would reinforce prestige and location in the network structure.

Introduction

The Internet and World Wide Web have facilitated new ways of communicating and collaborating on an unprecedented scale. These processes underpin work in the Higher education sector, and the availability of online data has provided an opportunity to try to examine the phenomenon empirically. Webometrics-based approaches (Thelwall, 2009) have seen the application of social network analysis to mapping links between academic web pages.

The assumption that web links can be used as a measure of quality underpins the Google PageRank algorithm (Brin & Page, 1998), and initial work in the Higher Education sector focused upon understanding the correlation between web link counts and research quality (Thelwall, 2002). Networks of academic web links have been shown to share characteristics with generalised social networks, such as power law-based degree distributions (Thelwall & Wilkinson, 2003).

In addition to their potential utility as a proxy metric for research quality and impact, web links between institutional webpages has also been shown to reflect geographic location. Ortega and Arguillo (2007) demonstrate this in the case of the Nordic Higher Education sector. In their study of the Chinese academic web, Yang, Liu and Meloche (2010) found that while influence was the principal factor, geographic location also played a role.

Links to academic web pages have also been examined as a way of capturing collaborative relationships. Stuart, Thelwall and Harries (2007) analysed links between UK Higher Education institutions and other bodies. While the findings showed some potential for web link analysis as an approach to measuring collaboration, most links did not reflect a collaborative relationship (ibid.).

However, the web has evolved in recent years. While web-link mining is rooted in a ‘web 1.0’ paradigm of static web pages as a source of information, universities also now make use of so-called ‘web 2.0’ or social media tools (O’Reilly, 2005). Founded in 2006, Twitter has become one of the worlds most popular social networking services. A growing body of research exists about the role of Twitter in relation to academia at the level of individual academics (Segado-Boj, Chaparro Domínguez & Castillo Rodríguez, 2015; Veletsianos, 2011; Veletsianos & Kimmons, 2016) or disciplinary communities (Holmberg & Thelwall, 2014; Mahrt, Weller & Peters, 2014).

The use of institutional accounts by universities is analogous to the institutional webpage although has received little research focus. Kimmons, Veletsianos and Woodward (2016) recently reported an analysis of tweets from institutional accounts in the US Higher Education sector. The findings from the analysis report that the institutional tweets are “1) monologic, 2) disseminate information (vs. eliciting action), 3) link to a relatively limited and insular ecosystem of web resources, and 4) express neutral or positive sentiment” (ibid.), and do not fulfil the full potential of the medium for more meaningful interactions (such as collaboration).” (Kimmons, Veletsianos & Woodward, 2016, p.1)

However, the nature of connection between institutional accounts has not been examined, and there is a question of whether the same influences of quality and location apply as seen in the case of the academic web. This exploratory study seeks to address this gap. The research questions in this study are therefore:

1. What are the structural characteristics of the network of UK higher education institutional accounts on Twitter?

2. Do university prestige or geographic location play a role in defining the network structure?

Method

Data was collected from Twitter, and subsequently analysed using a social network analysis-based approach. Social network analysis conceptualizes individuals as nodes, which will be connected by edges if a relationship exists between two nodes (Kadushin, 2012; Prell, 2012; Wasserman & Faust, 1994). In this case, the nodes are Twitter accounts, and a directed edge exists if one account is currently following another.

The first step required a list of Twitter IDs associated with the 147 current Higher Education institutions in the UK. In order to ensure that the IDs were associated with the institution as a whole (rather than a department, student society or parody account, for example), the Twitter IDs were selected on the basis of being the one featured on the institutional homepage. Three institutions did not have accounts linked on their homepages although a Twitter search found unambiguous institutional accounts for the remaining institutions.

Due to restrictions placed on the Twitter API, follower lists were manually collected by the researcher. The Twitter API only allows the first 5,000 friends or followers can be collected; as institutional accounts are highly followed, the sample was at particular risk of incomplete data collection because of this. To collect the data, a new Twitter account was registered and set to follow the institutional accounts intended to form the sample. Follower lists were created for each institutional account by viewing each in turn and recording the accounts listed under followers in common. Data collection took place in the week beginning 10th October 2016.

The data was then prepared and imported into Gephi for network analysis (Bastian, Heymann & Jacomy, 2009). Geolocation data (latitude and longitude) and positions in the most recent Times UK University rankings league table (ukuni.net, 2016) were added to be included in the analysis. Network metrics were subsequently exported and imported into SPSS for statistical tests (Field, 2009).

Results

The degree distribution of the network of UK Higher Education institutional Twitter accounts is shown in Figure 1. The distribution does not clearly show or rule out a heavy tailed distribution, a classic hallmark of social networks where a substantial proportion of nodes have a low degree.

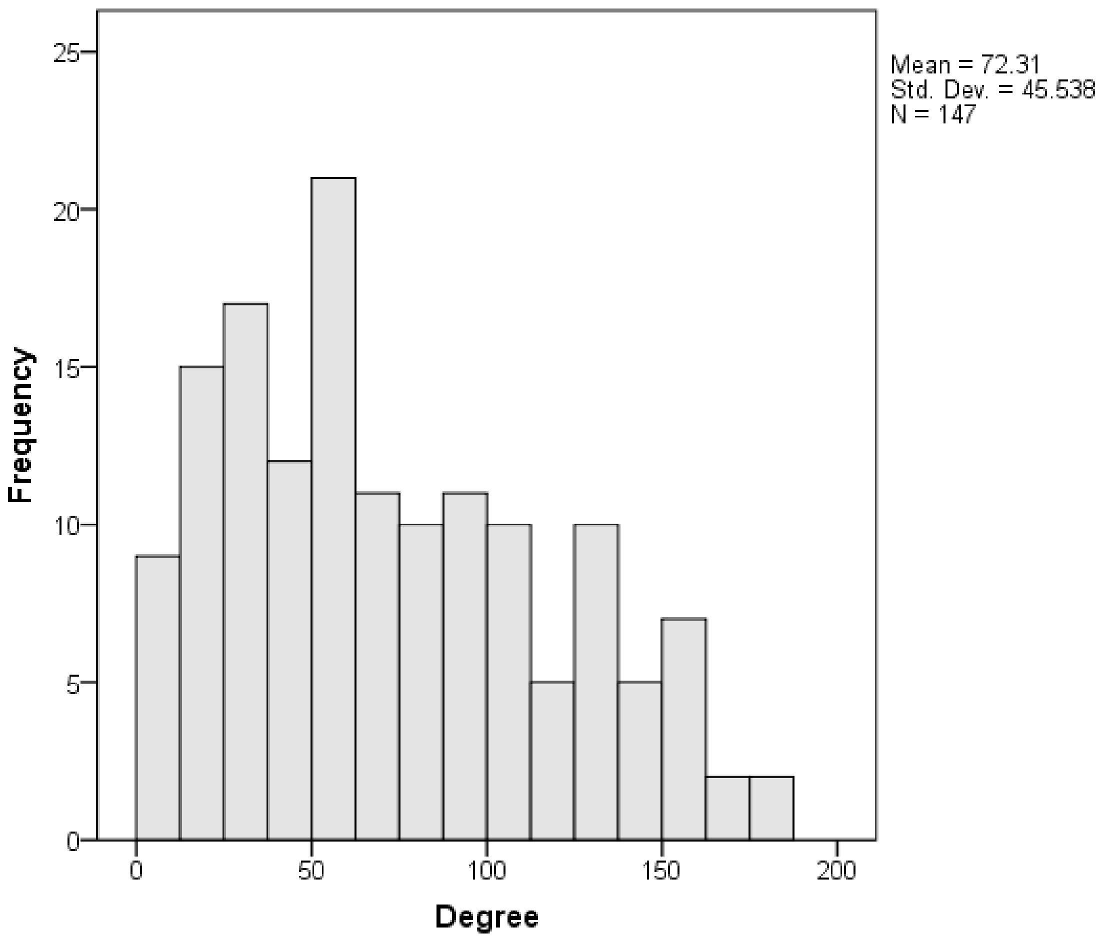

Figure 1: Degree distribution of the UK Higher Education institutional Twitter accounts network.

As the network is directed, it is possible to also consider degree distribution in terms of in-degree (number of accounts which are following an account) and out-degree (number of accounts followed by an account). An interesting asymmetry emerges when the data is considered in these terms. While in-degree follows an approximately normal distribution (Figure 2), the distribution of out-degree is much more unequal (Figure 3), with a substantial proportion of institutional accounts following few or zero other accounts, and a few accounts following nearly the whole network.

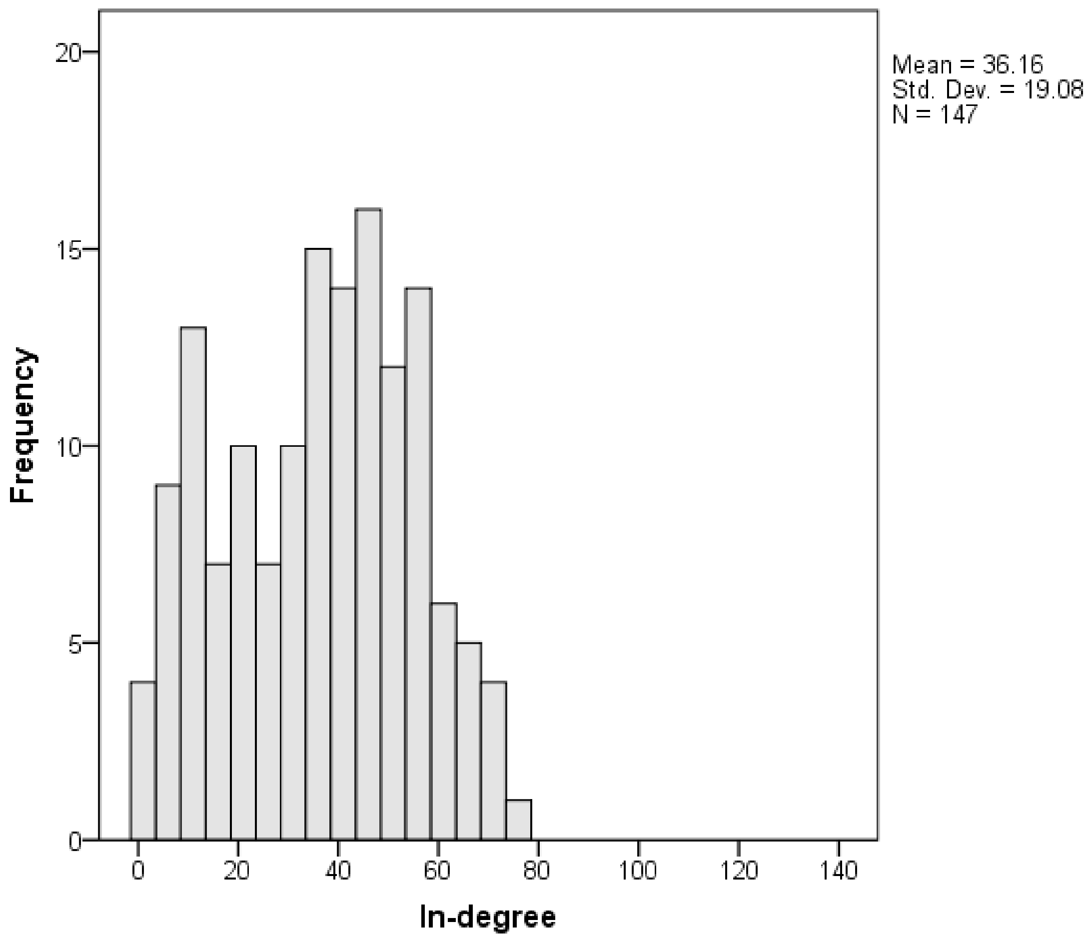

Figure 2: Distribution of in-degree of the UK Higher Education institutional Twitter accounts network.

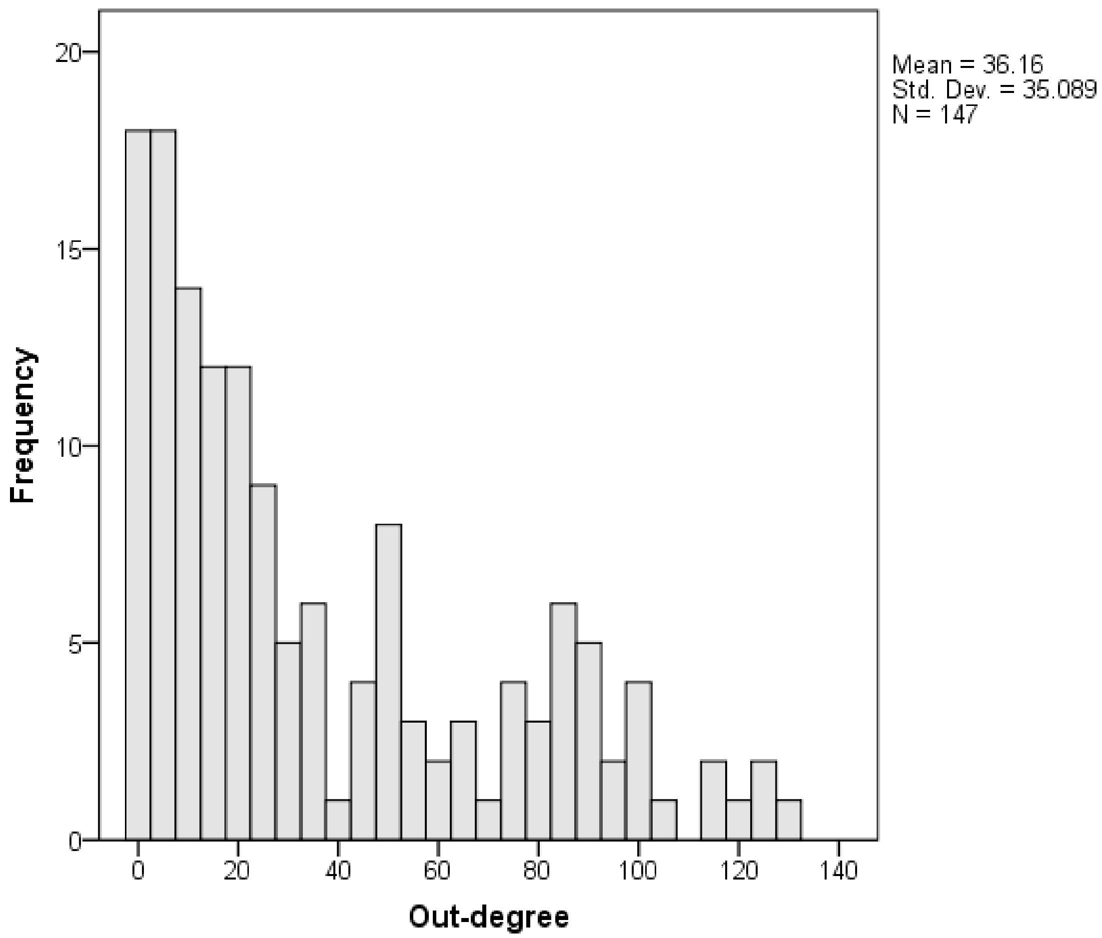

Figure 3: Distribution of out-degree of the UK Higher Education institutional Twitter accounts network.

Whether the network reflects the UK Higher Education sector more generally was examined by using institutional rankings as a proxy for relative standings of the institutions. University ranking is a good metric to use for this purpose as it lends itself readily to statistical tests, and is a metric itself based on a range of different attributes of Higher Education institutions.

The following metrics were tested for correlation with university ranking using linear regression: Degree, in-degree, out-degree, betweenness centrality, closeness centrality and eigenvector centrality.

Degree is indicative of the relative popularity of a node within a given network; it is the number of connections that node has to others (Prell, 2012). University ranking significantly predicted degree, *b* = -.467, *t*(109) = -3.787, *p* < .001. University ranking also explained a significant proportion of variance in degree, R^2 = .116, *F*(1, 109) = 14.338, *p* < .001. The correlation was negative; that is, lower ranked institutions have lower degree (Figure 4).

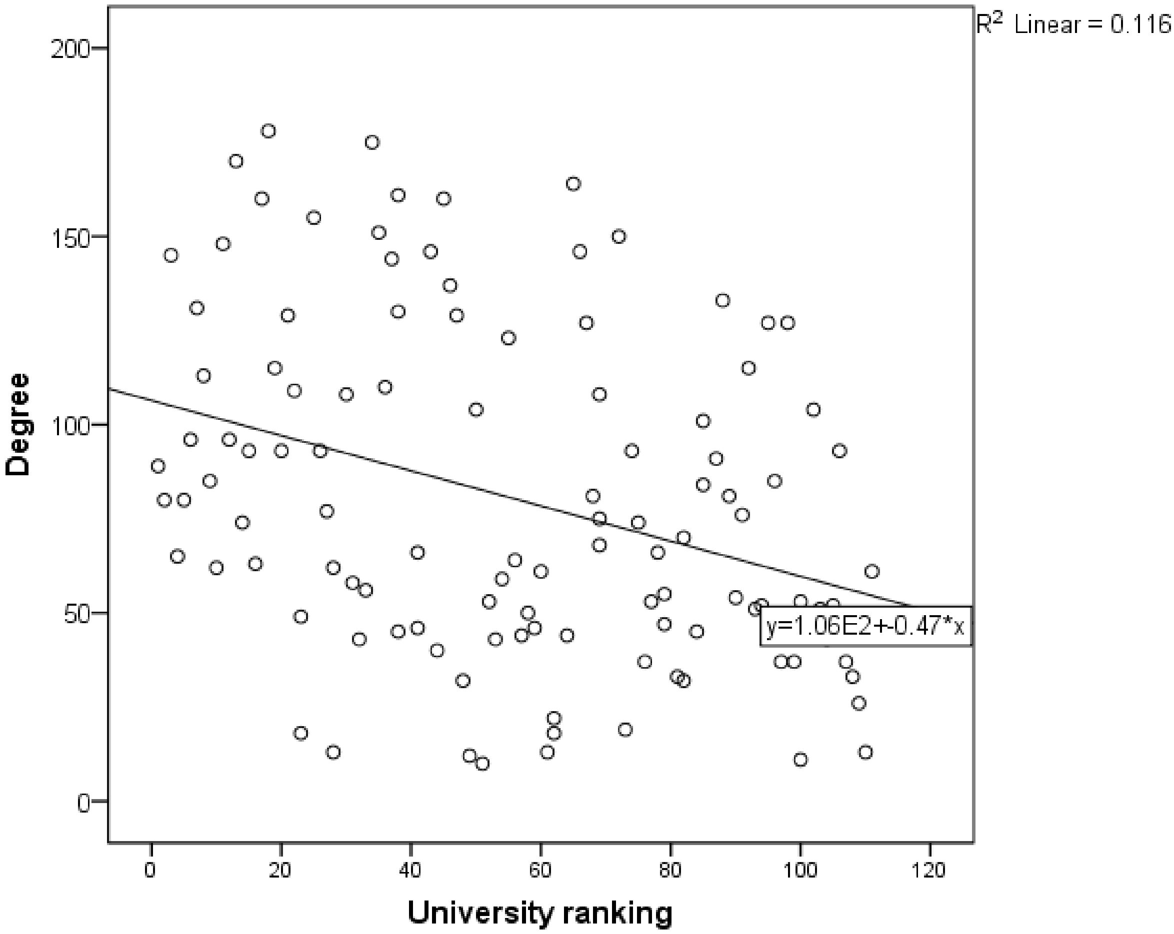

Figure 4: Negative correlation between degree and university ranking in the UK Higher Education institutional account Twitter network.

As connections on Twitter are directed in nature (that is, that links are not mutual but of a follow-following relationship; Wasserman & Faust, 1994), the number of connections an account has can also be considered in terms of in-degree (those who follow an account) and out-degree (the accounts that they follow). It is important to note that in-degree and out-degree here refer to the number of other UK Higher Education institutional accounts, as this is the network that was sampled, not overall follower or following statistics for the account.

University ranking significantly predicted in-degree, $b = -.219$, $t(109) = -4.642$, $p < .001$. University ranking also explained a significant proportion of variance in in-degree, $R^2 = .406$, $F(1, 109) = 21.546$, $p < .001$. Like degree, the correlation was negative; that is, lower ranked institutions have lower degree (Figure 5). There is a significant correlation between out-degree and University ranking (Spearman's coefficient of rank correlation, $r_s = -.23$, $p = .01$). Again, the correlation was negative (Figure 6).

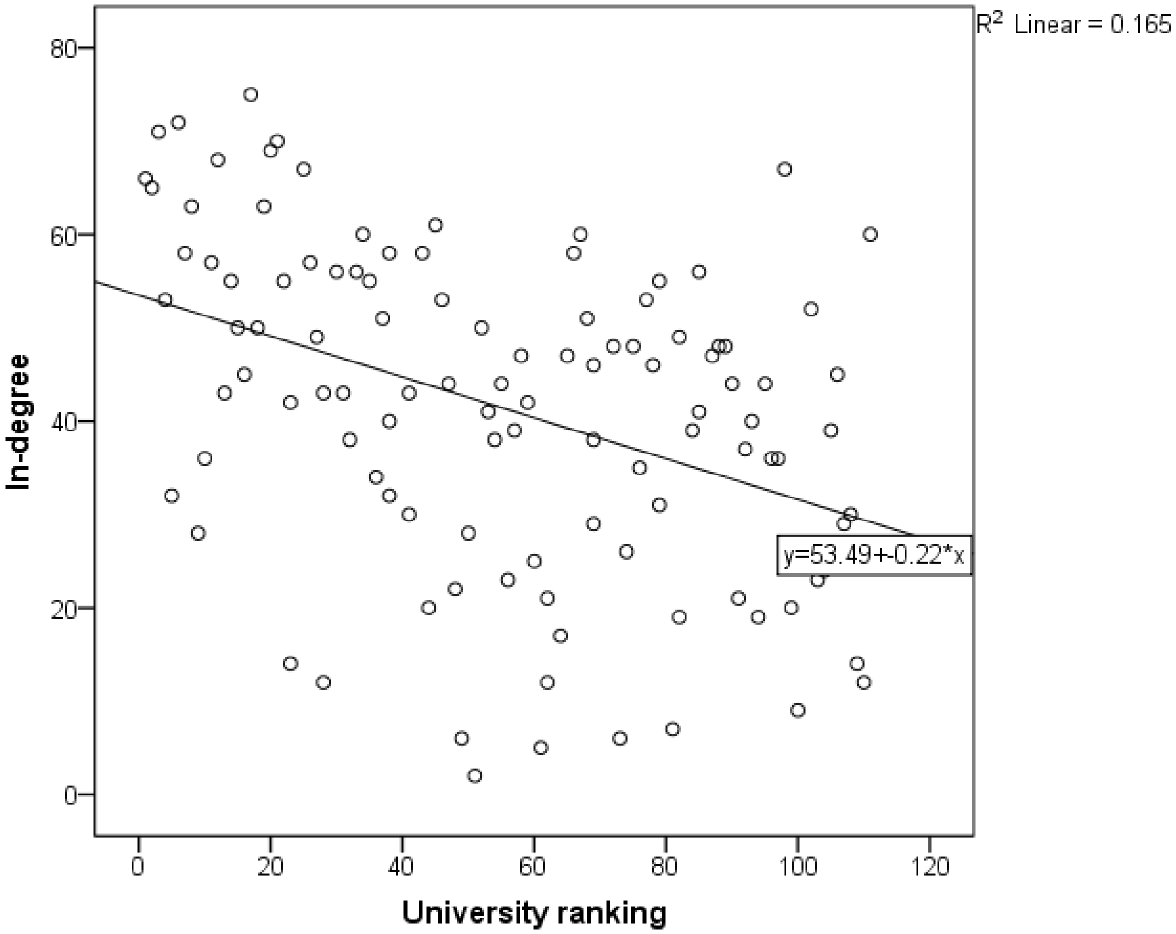

Figure 5: Negative correlation between in-degree and university ranking in the UK Higher Education institutional account Twitter network.

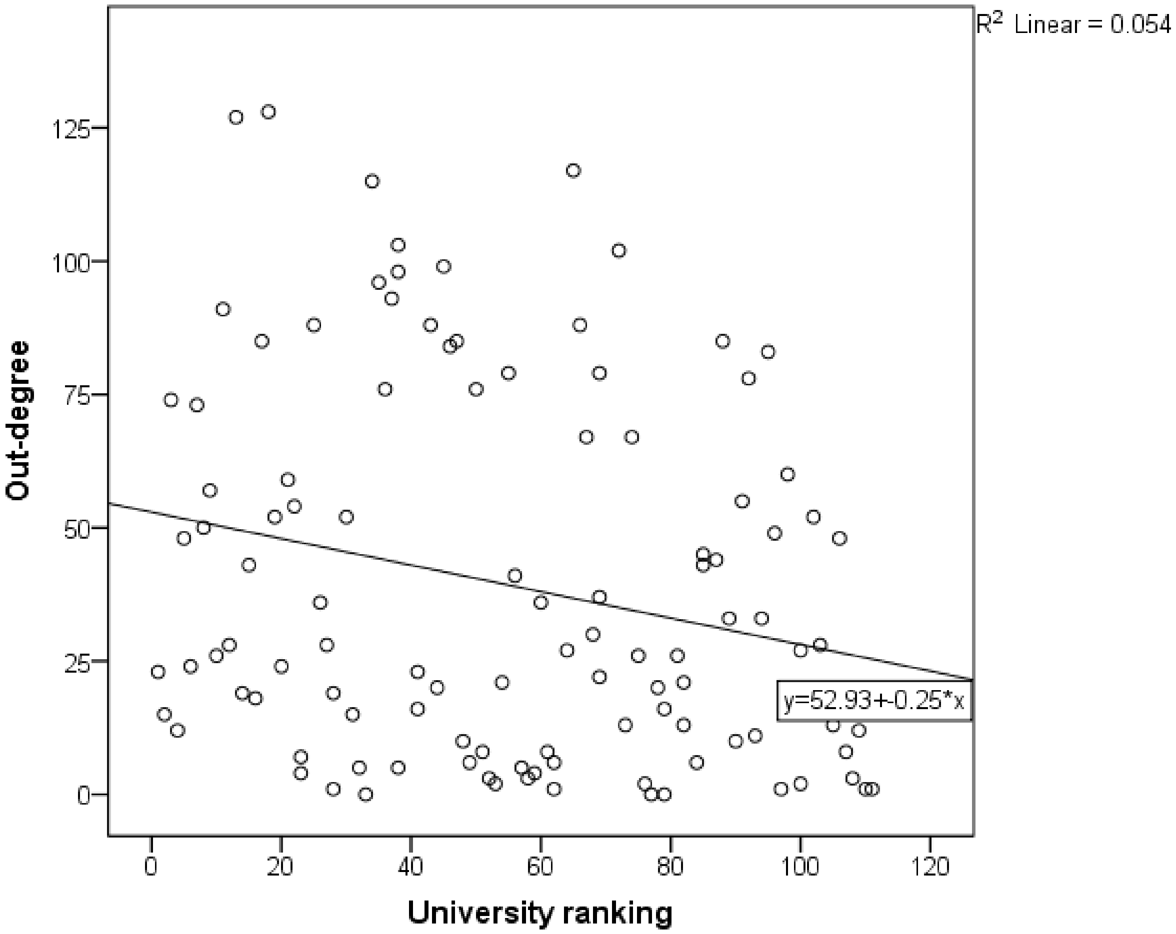

Figure 6: Negative correlation between out-degree and university ranking in the UK Higher Education institutional account Twitter network.

The relative position of different accounts within the network structure was considered through two types of centrality measure: betweenness centrality, and closeness centrality.

Betweenness centrality is a measure of how central a nodes' location is within a network, based on the how frequently a node is located on the shortest path between any two other nodes (Prell, 2012). A high betweenness centrality would imply a more privileged position in terms of controlling the flow of information. In this network, higher betweenness centralities are associated with more prestigious institutions.

University ranking significantly predicted betweenness centrality, *b* = -1.989, *t*(109) = -3.913, $p < .005$. University ranking also explained a significant proportion of variance in betweenness centrality, $R^2 = .123$, $F(1, 109) = 15.310$, $p < .005$. Again, the correlation was negative (Figure 7).

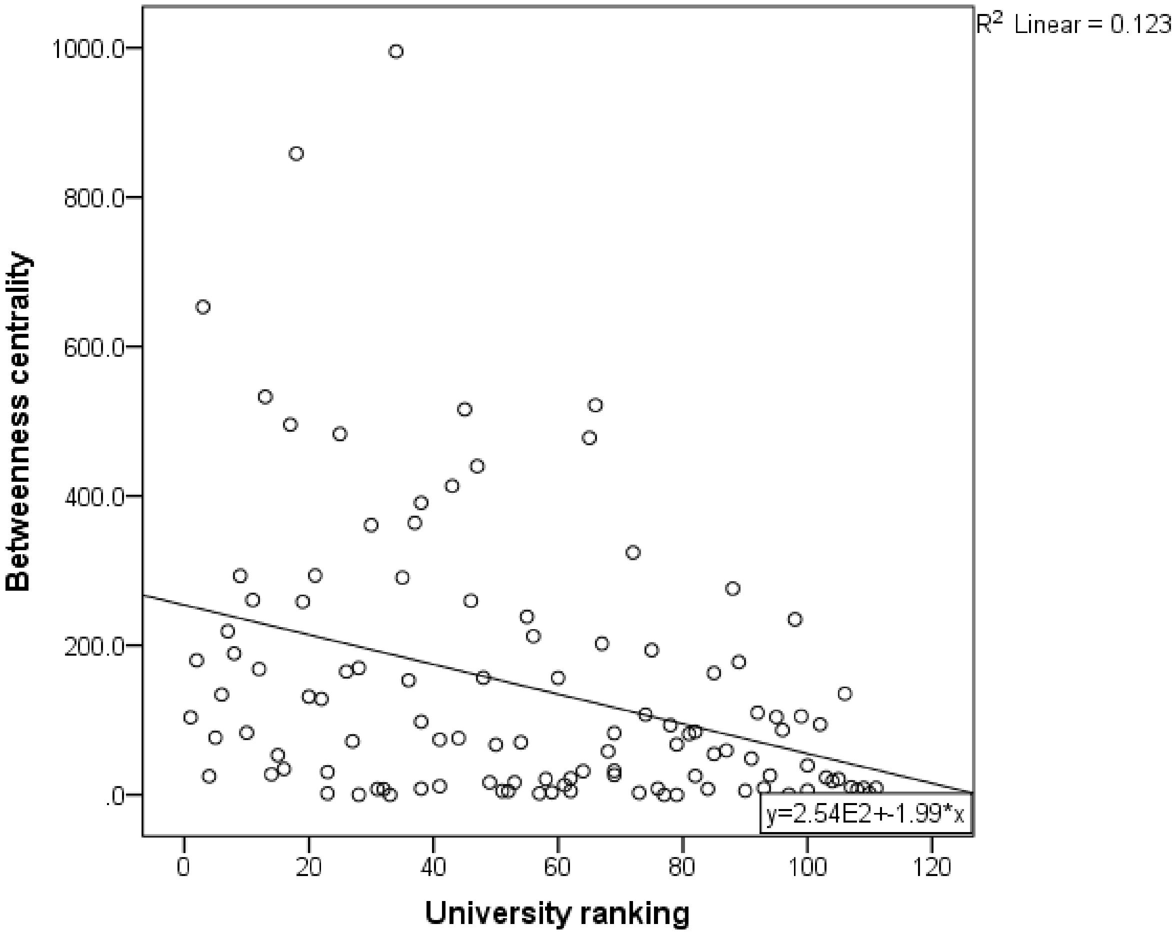

Figure 7: Negative correlation between betweenness centrality and university ranking.

A contrasting measure of centrality was also examined. Closeness centrality is a measure of independence; if a node is not central themselves, they would rely upon other, more central (in the sense of betweenness centrality) actors to provide and disperse information. Closeness centrality is a measure of how well a node is located in relation to other key nodes (Prell, 2012). University ranking does not significantly predict closeness centrality, suggesting that lower ranked universities may be able to take advantage of their position in relation to higher ranked universities in the context of Twitter as an information network.

Turning to the structural properties of the network overall, the network itself had a single connected component. The network has a diameter of 4, in keeping with being a ‘small world’ (Kadushin, 2012). The modularity algorithm in Gephi (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) was applied in order to identify clusters of more highly connected nodes within the network. Five communities were detected, as shown in Figure 8.

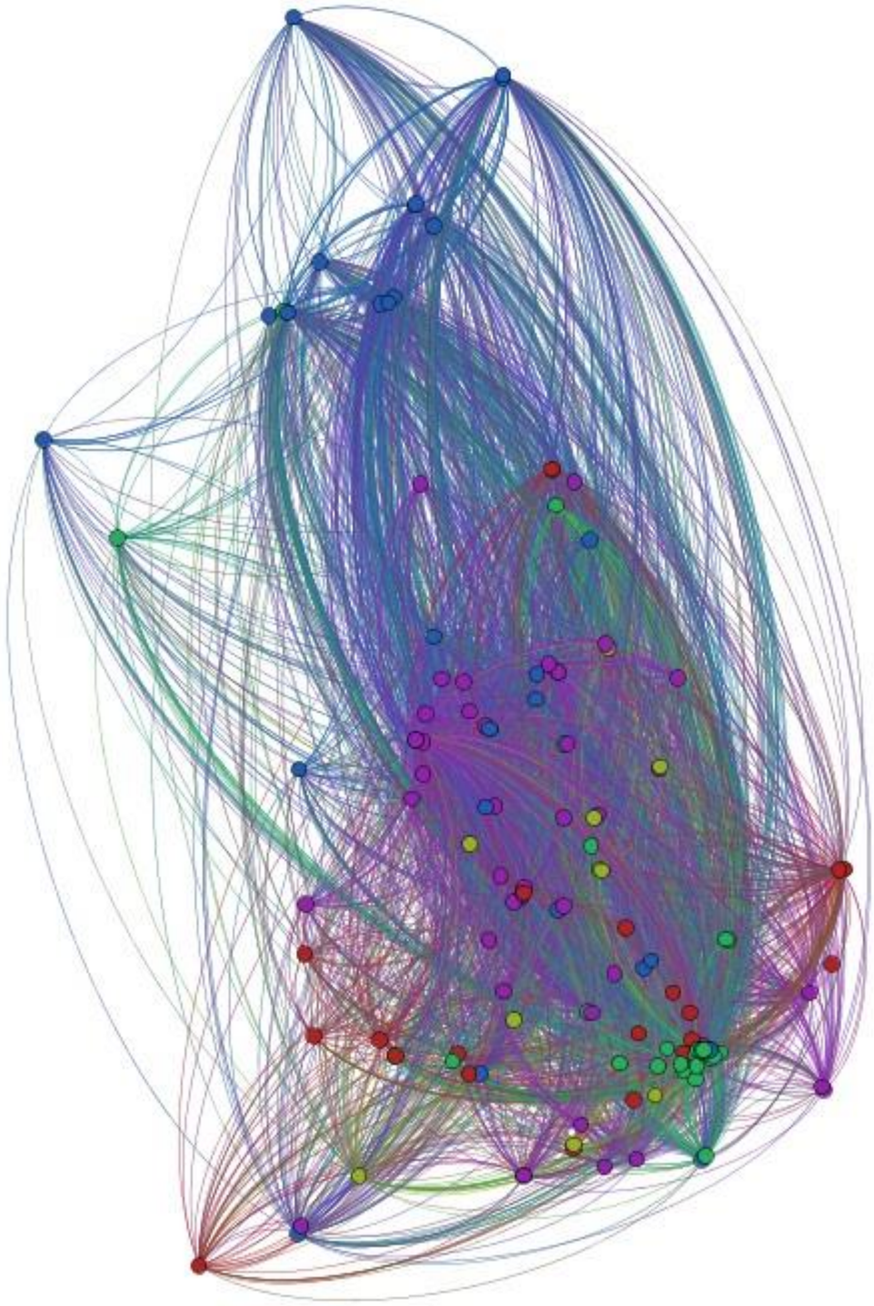

Figure 8: Network structure of the connected component of the UK Higher Education institutional accounts network, arranged according to geolocation. Nodes are colour-coded according to communities.

The network visualisation suggests that there may be both a geographic and prestige component to the communities. The influence of prestige and geolocation were tested statistically through nonparametric tests between the communities as groups based on these criteria.

The role of prestige was examined by testing for statistical differences in ranking associated with each of the communities identified in the network. Median values of university ranking varied significantly according to each of the communities identified (independent samples median test, χ^2 (4, N = 111) = 28.997, median = 56.0, p < .005). Note that the test N is less than the total number included in the network as not all of the institutions are included in the Times ranking. The distribution of ranking scores present in each community is shown in Figure 9.

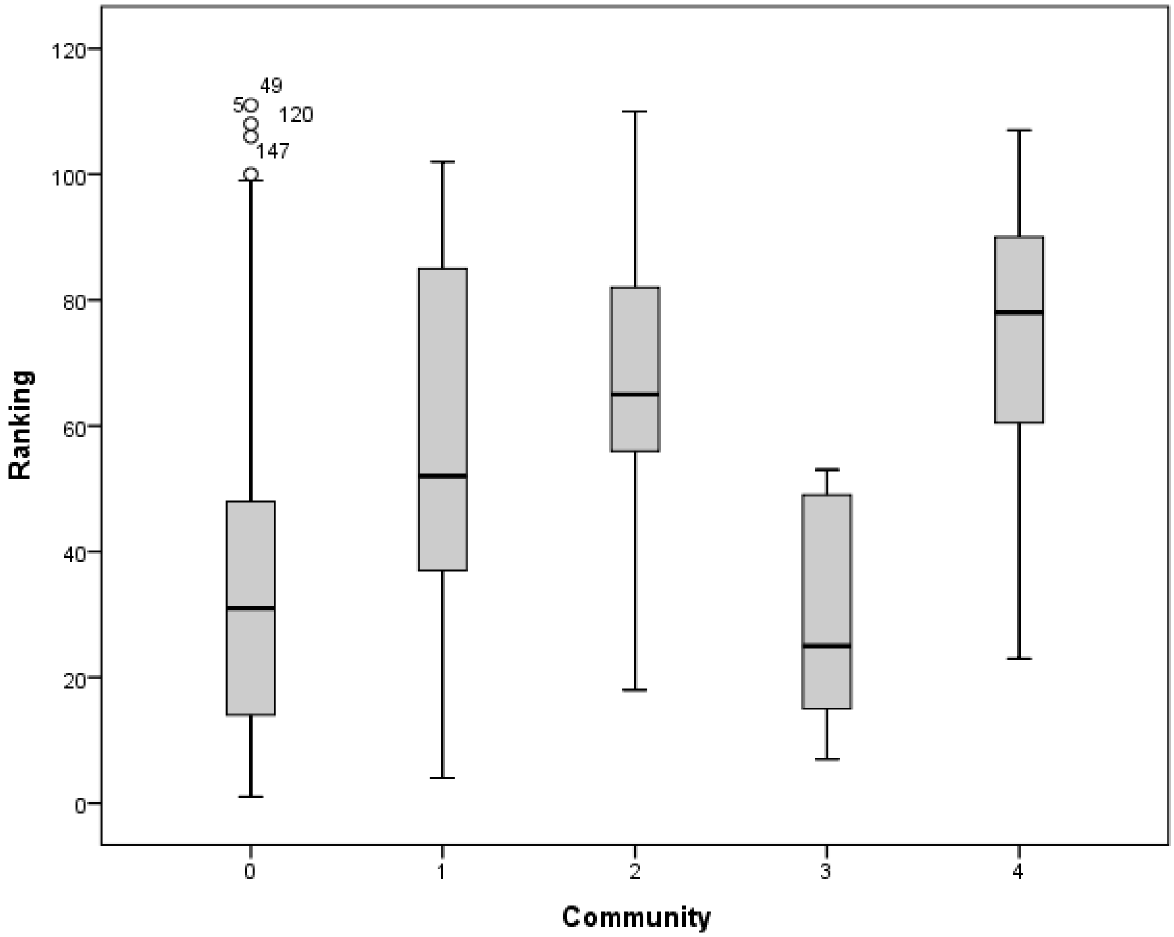

Figure 9: Boxplots showing the distribution of ranking scores for institutions present in each community identified within the network.

While the importance of geolocation in how communities are defined within the network is clearly perceptible from the network visualisation (Figure 8), it was also tested statistically by applying median tests to examine whether latitude or longitude vary significantly according to the different community groupings. Both latitude and longitude were found to vary significantly according to the communities. Median values of latitude varied significantly according to each of the communities identified (independent samples median test, χ^2 (4, N = 147) = 32.472, median = 52.1, $p < .005$); similarly, median values of longitude varied significantly according to each of the communities identified (independent samples median test, χ^2 (4, N = 147) = 31.143, median = -1.4, $p < .005$). The distribution of latitude and longitude values present in each community are shown in Figures 10 and 11 respectively.

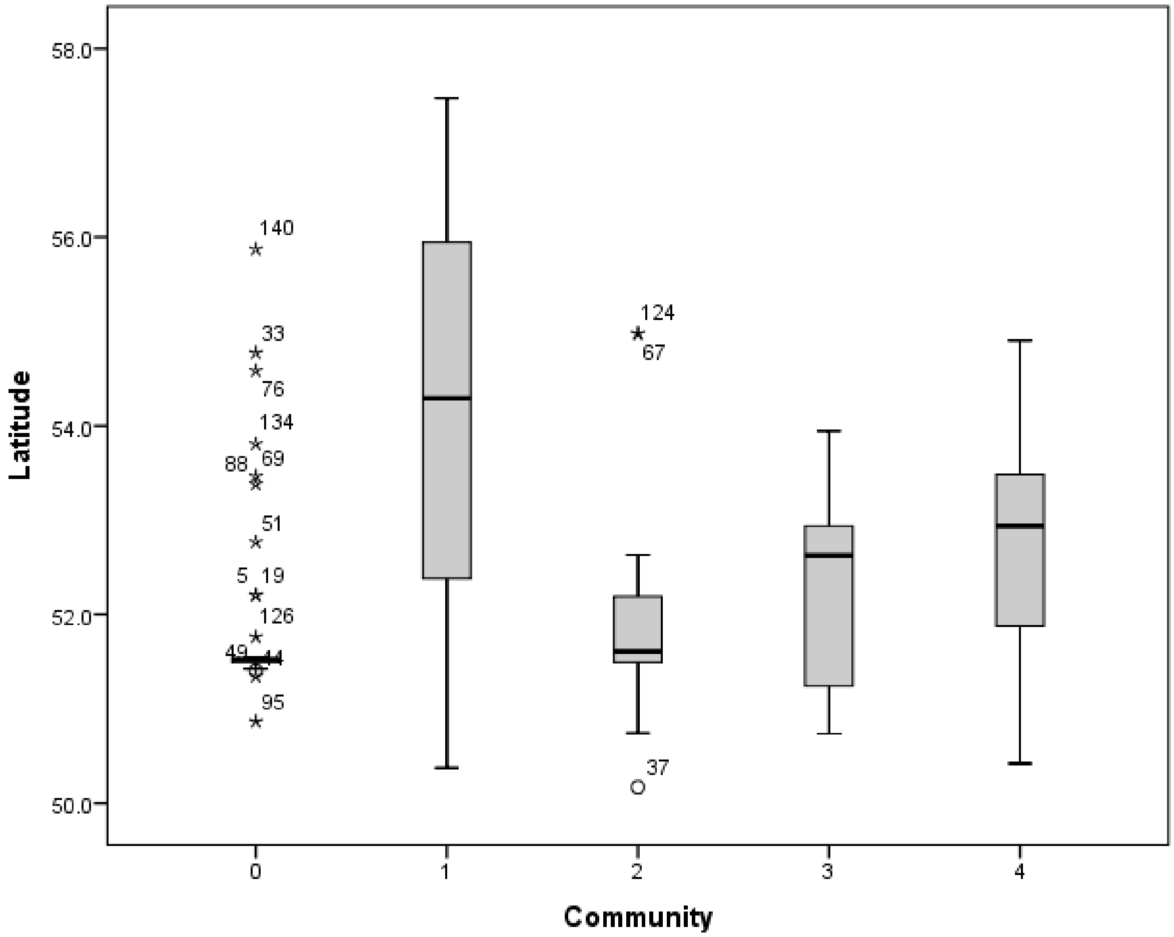

Figure 10: Boxplots showing the distribution of values of latitude for institutions present in each community identified within the network.

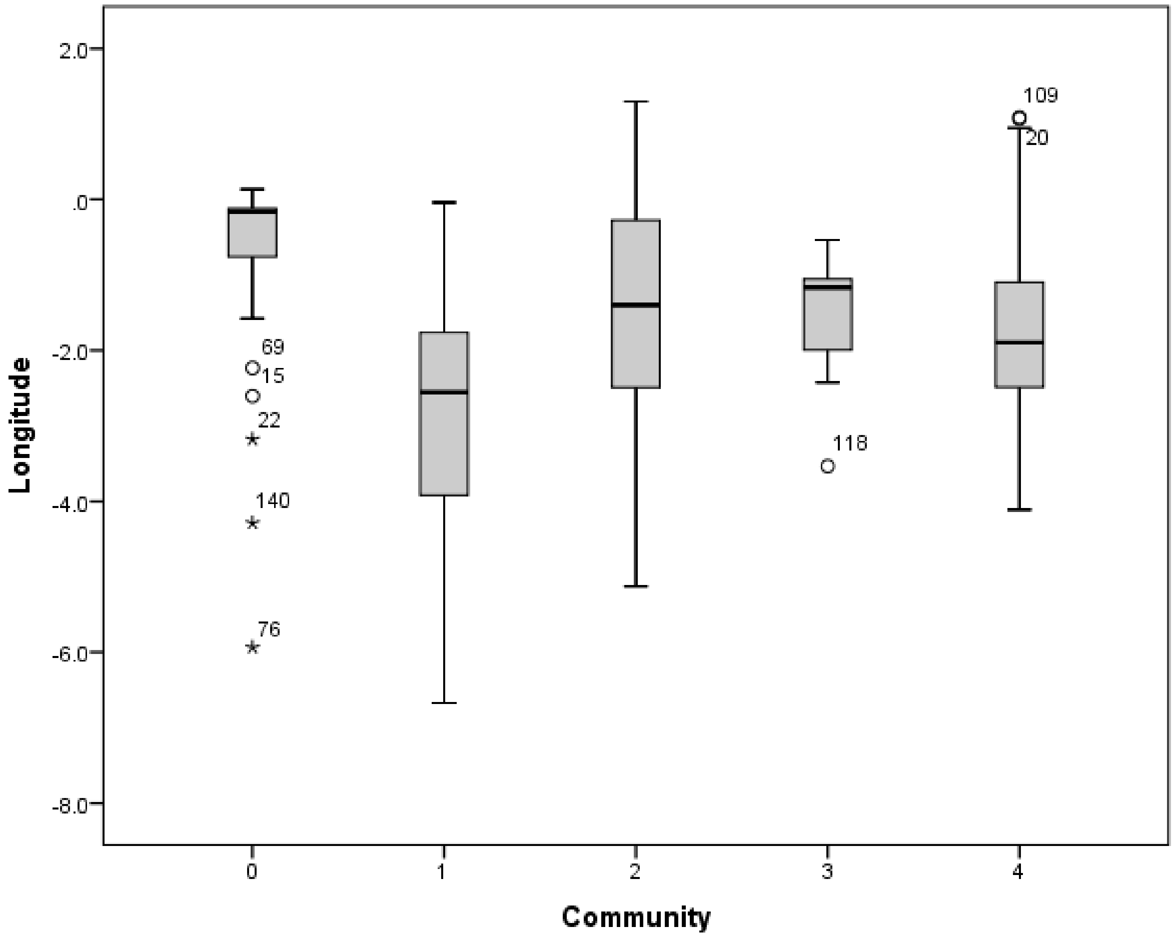

Figure 11: Boxplots showing the distribution of values of longitude for institutions present in each community identified within the network.

Further insight into the role of geolocation and prestige in community definition was sought by looking at incidence of mentions of other institutional accounts in tweet data. This data was collected using NodeXL (Smith et al., 2009). This method of Twitter data collection uses the API, so it is restricted to the previous seven days and is not necessarily an exhaustive set. As the purpose for collecting the data in this case was to explore possible explanations to triangulate findings from the network analysis the restrictions were not problematic, although a fuller analysis of the UK Higher Education institutional tweets over a longer time period would be a valuable follow-up project in its own right. The data were collected on 28th October 2016. As a high proportion of institutional tweets mention only themselves, the focus was upon mentions to look for evidence of multiple institutional accounts interacting and the reasons for this. Duplicate tweets and those which only mentioned a single Higher Education institution were removed; the sample contained 185 tweets once these criteria had been applied. Each tweet was categorised according to its apparent purpose using an open coding approach.

The data revealed a range of purposes for mentioning other institutions in tweets that shed light on the clusters relating to a combination of location and prestige.

Table 1: Emergent themes from analysis of tweets which mention at least two UK Higher Education institutions.

Topic	Frequency	Percentage
Academics mobility	31	16.8
Collaboration	83	44.9
Local events	11	5.9
Prospective students	16	8.6
Redirection	4	2.2
Social	14	7.6
Sporting events	15	8.1
Students mobility	6	3.2
Unknown	5	2.7
Total	*185*	*100*

The principal way in which institutional accounts were co-mentioned in relation to prestige was the frequent tagging of partner institutions in relation to collaborative endeavours, either through announcing new partnerships and initiatives, or promoting the findings and outputs from research projects. Three of the categories may be related indirectly to prestige, being linked by strengths in particular subject areas. These include academics mobility (academics visiting other institutions, or receiving honorary degrees), students mobility (student visiting other institutions), and prospective students (where institutional twitter accounts encourage potential applicants). In relation to location, universities located in the same city or wider locality were frequently mentioned together in tweets about local events or sporting events, either hosted at one of the mentioned institutions, or being co-mentioned in tweets originating from local third parties.

Conclusions

This exploratory study has shed some light on the structure and social dynamics of the UK Higher Education institutions on Twitter. Similar to previous studies focused on university web links, prestige and geographic location play a role in the structure of the network. The study also provides potential avenues for future research.

Having an institutional Twitter account is now standard practice in the UK Higher Education sector; the results show that all UK Higher Education institutions have an institutional account.

The network structure created by the accounts who are connected suggests that the network replicates offline structures to an extent. Relative popularity (degree and in-degree) is predicted by university ranking; that is, there is a trend toward higher ranked universities being more popular and having more followers from within the sector. It is more surprising that the same trend is seen for out-degree, which is the number of institutional accounts that they follow. Although the correlation is weaker, this may suggest a third factor of level of institutional social media use perhaps as also being correlated with ranking.

Position within the network is also related to university ranking, but in more nuanced ways. Higher ranked institutions have greater betweenness centrality, suggesting that higher ranked institutions are in positions of greater control over the flow of information within the network. In contrast, closeness centrality, is not related to university ranking. Closeness centrality is indicative of independence within the network, and has been shown to be associated with influence, power, and access to information (Prell, 2012), so there may be potential for traditional hierarchy to be overcome in this context.

University ranking and geolocation both contribute to the communities observed within the network. Drawing upon a weeks' mentions in tweets, geographically-close institutions are frequently mentioned together in relation to local events. The data suggests that Universities of similar ranking may develop ties through tweeting about outputs from collaborative projects. While these reasons appear to be the most prevalent in relation to explaining the observations in the network data, a range of other reasons were also observed to lesser extents (such as engaging with prospective students, visiting academics and students, which may be related through subjects in common across different universities).

Although this provides some insight into the network structure, the tweet data is limited to seven days (the limit imposed by the Twitter API) and is not the main focus of this study. A more sustained data collection and systematic analysis of tweets would be valuable in this respect. Analysis of tweets would also have potential for a comparative study with the US Higher Education sector, drawing parallels with recent work by Kimmons, Veletsianos and Woodward (2016).

Although the results are statistically significant, a substantial proportion of the variation is unexplained. A limitation of the analysis presented here is that it does not account for factors other than ranking or location; expanding the model to include other factors may provide further insights. By focusing solely upon the links between institutional accounts, the study does not examine the popularity or position of accounts in relation to the Twitter network associated with the Higher Education sector or public more broadly. The variation may also be due to different motivations or reasons for institutions to use institutional Twitter accounts. Surveying those who run institutional accounts would be highly illuminating to address this gap.

References

M. Bastian, S. Heymann, and M. Jacomy, 2009. "Gephi: An open source software for exploring and manipulating networks", *Proceedings of the International AAAI Conference on Weblogs and Social Media*. San Jose, USA, 17-20 May. Palo Alto, USA, AAAI, pp. 361-362.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, 2008. “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics: Theory and Experiment*, volume 2008, number 10, P1000 [Online]. Available at http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta (Accessed 24 April 2016).

Sergey Brin and Larry Page, 1998. “The anatomy of a large-scale hypertextual Web search engine”, *Computer Networks and ISDN Systems*. Volume 30, pp. 107–117.

A. Field, 2009. *Discovering Statistics using SPSS*, 2nd ed. London: SAGE.

Kim Holmberg and Mike Thelwall, 2014. “Disciplinary differences in Twitter scholarly communication”, *Scientometrics,* volume 1010, number 2, pp. 1,027-1,042.

C. Kadushin, 2012. *Understanding social networks: Theories, concepts, and findings*. Oxford: Oxford University Press.

Royce Kimmons, George Veletsianos, and Scott Woodward, 2016. “Institutional uses of Twitter in U.S. Higher Education”, *Innovative Higher Education*, Early view doi:10.1007/s10755-016-9375-6

M. Mahrt, K. Weller and I. Peters, 2014. “Twitter in scholarly communication”, in Weller, K., Bruns, A., Burgess, J., Mahrt, M. & Puschmann, C. (Eds.) *Twitter and society*, New York, Peter Lang, pp. 399-410.

T. O'Reilly, 2005. "What is web 2.0? Design patterns and business models for the next generation of software", at http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html , accessed 27 April 2016.

José L. Ortega and Isidro F. Aguillo, 2007. "Visualization of the Nordic academic web: Link analysis using social network tools", *Information Processing and Management,* volume 44, number. 4, pp. 1624-1633.

C. Prell, 2012. *Social network analysis: History, theory and methodology*. London: SAGE.

Francisco Segado-Boj, Maria A. Chaparro Domínguez, and Cristina Castillo Rodríguez, 2015. "Use of Twitter among Spanish communication-area faculty: Research, teaching and visibility", *First Monday,* volume 20, number 6, at http://dx.doi.org/10.5210/fm.v20i6.5602 , accessed 27 April 2016.

M.A. Smith, B. Shneiderman, N. Milic-Frayling, E. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave, 2009. "Analyzing (social media) networks with NodeXL", *Proceedings of the Fourth International Conference on Communities and Technologies*. Pennsylvania, USA, 25-27 June. ACM, pp. 255–264.

David Stuart, Mike Thelwall and Gareth Harries, 2007. “UK academic web links and collaboration - an exploratory study”, *Journal of Information Science,* volume 33, number 2, pp. 231-246.

Mike Thelwall, 2002. “An initial exploration of the link relationship between UK university web sites”, *ASLIB Proceedings*, volume 54, number 2, pp. 118-126.

M. Thelwall, 2009. *Introduction to webometrics: Quantitative web research for the Social Sciences*. Morgan & Claypool.

Mike Thelwall and David Wilkinson, 2003. “Graph structure in three national academic Webs: Power laws with anomalies”, *Journal of the American Society for Information Science and Technology*, volume 54, number 8, pp. 706-712.

UK Uni.net, 2016. “UK university ranking 2016-Times”, at https://www.ukuni.net/articles/UK-University-Ranking-2016-Times , accessed 26 September 2016.

George Veletsianos, 2011. “Higher education scholars' participation and practices on Twitter”, *Journal of Computer Assisted Learning*, volume 28, number 4, pp. 336-349.

George Veletsianos and Royce Kimmons, 2016. “Scholars in an increasingly open and digital world: How do education professors and students use Twitter?”, *The Internet and Higher Education,* volume 30, pp. 1-10.

S. Wasserman and K. Faust, 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Bo Yang, Zuihui Liu and Joseph A. Meloche, 2010. “Visualization of the Chinese academic web based on social network analysis”, *Journal of Information Science*, volume 36, pp. 131-143.