The impact of missing data on network centrality measures

Joe Chick, University of Warwick
(j.chick@warwick.ac.uk)
Supervisor: Prof. Beat Kümin

From the 1990s onwards, there has been an explosion in the number of history publications making reference to networks.[1] In some, this is little more than a buzz word, but others capitalize on developments in computer technology and specialist software to apply the formal statistical tests of this social science method to historical datasets. Many such studies are with reference to the modern era, for which the sources are most comparable to those of the social sciences. Yet the method's application to pre-modern documents has also grown in recent years. Such sources pose a number of specific challenges. The most obvious of these, and the one explored in this paper, is source survival.

My PhD research has applied social network analysis to Reading in the years 1350-1600 as a way of exploring social relations under monastic lordship and the impact of the Dissolution. In doing so, I have worked with two sets of network data: one based on conveyancing and moneylending transactions, the other on wills. This has been combined with a database of 4,519 people containing information on burgess status, occupation, and officeholding from a wide range of sources. This allowed the actors in the networks to be analysed in terms of shared personal attributes. The social network analysis has enabled me to contribute to scholarly debate on urban society and to challenge aspects of the existing scholarship. In particular, in a recent article I challenged Robert Tittler's claim of a post-Dissolution rise in urban oligarchy by demonstrating continued direct interaction between top-level civic officeholders and lower-status inhabitants after the dissolution of Reading Abbey and the ensuing incorporation of the town.[2] While some statistical tests produce p-values, measures of centrality do not and this inspired me to conduct an investigation into the robustness of the methodology, the results of which I present in this paper.

Testing Robustness

The extant social network analysis literature lacks quantifiable tests of the extent to which statistics are disrupted by incomplete source material. Many researchers have discussed methodological issues, particularly the virtually universal problem of incomplete pre-modern sources, but there has been a lack of quantifiable robustness tests by historians.[3] A few have been undertaken by social scientists, but their applicability to pre-modern datasets could be questioned.[4] Most relevant to the issue of missing data is that of Elizabeth Costenbader and Thomas W. Valente, in which data were randomly deleted from eight fairly complete network datasets to mimic the process of missing data. Their study found a measure of centrality known as *eigenvector* to be the most robust, with betweenness proving unreliable. Yet, their approach sought to mimic social scientists' problem of imperfect sampling methods and participants in a survey not responding. This is very different from the problem faced by pre-modern historians, in which it is often entire collections of historical documents that are lost. Here the data are not missing at random. Another obstacle that distinguishes pre-modern

[1] Innes (2016).
[2] Chick (2019).
[3] For a recent methodological discussion of pre-modern social network analysis, see Goddard (2019).
[4] Costenbader and Valente (2003); Borgatti et al. (2006); Lee et al. (2006).

history from the social sciences is that low-status members of society are often heavily underrepresented in the sources.

This paper makes an original contribution to scholarly knowledge by testing the robustness of the methodology specifically in relation to pre-modern sources. It uses an adapted version of Costenbader and Valente's approach, but with the non-random deletion of an entire source collection. The paper then explores the impact which this has on a group of statistical tests known as measures of centrality. The dataset for this test consists of the 293 surviving Reading wills from the period 1490-1589. The software package UCINET was used to create a network composed of the names of the testators, executors, overseers, and witnesses. The data are treated as a directed network, meaning the arrows point from the testator to the other individuals, reflecting the fact that the testator selected these people.

The wills are located in two archive collections. Wills that people registered with the Prerogative Court of Canterbury are held by the National Archives, while those registered with the Berkshire Archdeaconry Court are held by the Berkshire Record Office.[5] The Canterbury wills represent the richest inhabitants, since there was a wealth requirement for registering one's will with this court and, by consequence, the Berkshire ones represent the less wealthy residents. This paper's test first calculates the measures of centrality of the combined dataset of both courts, then repeats the procedure with all the Berkshire Archdeaconry Court wills deleted. This mimics a historian's problem of entire source collections being missing. Deleting the Berkshire wills is of particular interest, since it removes testators of lower social status, people who are often heavily underrepresented in pre-modern sources. This said, the wealth differences between the two sets of testators should not be overstated, since the very poorest in society simply did not leave a will at all. Rather the Berkshire testators represent the middling sort.

Measures of Centrality

Centrality is one of the most commonly used statistics in studies that utilize social network analysis. There is no single definitive way of measuring centrality so researchers must choose between several options and the four most commonly used ones are tested in this study. The first measure simply counts the number of links that an actor has to other actors in the network, known as degree centrality. If a network is directed, as in this study, then a distinction can be made between in-degree centrality (the number of links pointing towards an actor) and out-degree centrality (the number of links pointing away). The second measure is closeness centrality, which calculates the total number of steps it takes for a node to reach each of the other nodes when following the shortest possible path. Similarly, analysis of a directed network can distinguish between out-closeness, which follows paths along links pointing out from the actor in question, and in-closeness, which looks at links pointing towards the actor. The third measure is betweenness centrality, which looks at how often a node is passed through in the shortest paths between all the possible node pairings in the network. The fourth is *eigenvector* centrality, a calculation in which a node gains a high centrality score by being connected to other nodes with a high centrality. Calculating the centrality of one actor changes the centrality for those around which in turn affects the original actor, making this an iterative calculation.

Deleting half the dataset and reprocessing the figures is a straightforward process. More difficult is to find a satisfactory way of assessing the extent to which they have been disrupted. One approach is to look at the average change in an actor's score for each measure of centrality. This, however, is flawed, since the various measures of centrality are quantified on different scales that are not comparable. As researchers, often we are not actually that interested in the scores themselves, so much as where an actor is in relation to others. For this

[5] TNA, PROB 11; BRO, D/A1.

reason, this study ranks the actors according to their score and looks at the average change in rank as a result of the deleted data. A second way in which disruption is assessed is to look at how often an extreme change of position occurs.

Outcomes

Removing the Berkshire wills substantially reduced the size of the dataset. The full dataset was based on 293 wills and contained 1,248 unique actors, and 1,689 dyads (meaning pairings of connected actors). The deletion of the Berkshire court wills left a dataset of 105 wills, with 552 actors, and 687 dyads.

Tables 1 to 3 show the impact of deleting the Berkshire wills on the measures of centrality. Table 1 gives the average change in score but, as noted, this is of limited value since the figures are not necessarily comparable to one another. UCINET offers an ostensibly standardised version of each measure, used in Table 1's figures, which places all measures of centrality on a scale of 0 to 1 but even this has major problems in terms of comparability.[6] The highest score that anybody actually achieves in practice is 0.27 with *eigenvector* but only 0.004 with out-degree. In assessing changes in score, *eigenvector* will inevitably perform poorly since it produces scores almost 100 times larger than out-degree.

Table 1: *Average change in actors' centrality score*

	ID	*OD*	*IC*	*OC*	*Bet*	*Eig*
Mean change	0.0007	0.0005	0.0141	0.0142	0.0009	0.0117
Median change	0.0010	0.0000	0.0140	0.0140	0.0000	0.0030
Mode change	0.0010	0.0000	0.0140	0.0140	0.0000	0.0000

Notes: Abbreviations used as follows: ID (in-degree), OD (out-degree), IC (in-closeness), OC (out-closeness), Bet (betweenness), and Eig (*eigenvector*).

More useful is Table 2, which shows how many positions an actor moved when the actors were ranked according to their centrality score. All three types of average tell a broadly similar story. The most surprising outcome is that *eigenvector*, which had proved most robust in Costenbader and Valente's study of randomly deleted data, appears a contender for least robust when data are deleted in a non-random way. Yet the situation is not as straight-forward as this. *Eigenvector* produces a particularly broad range of scores compared with closeness and betweenness. Most measures of centrality produce a cluster of people with a score of 0, meaning they are not central at all, but this is a particularly large group in betweenness centrality. In fact, 495 of the 552 actors had a score of 0 and, when the Berkshire wills were deleted, this rose to 519. It is unsurprising that betweenness appears more robust than *eigenvector* when it was doing very little to differentiate between people in the first place.

Table 2: *Average change in actors' ranking*

	ID	*OD*	*IC*	*OC*	*Bet*	*Eig*
Mean change	49.4	14.9	14.3	7.8	22.6	71.1
Median change	59.0	16.0	8.0	5.0	24.0	52.0
Mode change	59.0	16.0	8.0	5.0	24.0	52.0

Table 3 adopts a different approach to prevent measures creating an impression of robustness through having large numbers of actors with scores of 0. This is achieved by looking only at

[6] Table 1 uses UCINET's standardised measures but Tables 2 and 3, being based on rankings rather than scores, use the default scoring system.

those actors who did not have a score of 0 before any data were deleted.[7] This created the need for a standardised ranking system, since *eigenvector*'s broad range of scores meant there were 371 unique scores in the results, while betweenness had only 34. The figures in Table 3 adjusted the ranking system for all measures onto an evenly spread scale of 1 to 80.

Table 3: *Average change in actors' ranking on a standardised scale*

	ID	*OD*	*IC*	*OC*	*Bet*	*Eig*
Mean change	7.7	6.9	16.4	10.5	12.7	8.9
Median change	7.2	3.5	11.7	7.9	11.3	5.1
Mode change	7.2	3.5	25.2	7.9	11.3	0.0
Extreme change	1%	8%	21%	1%	9%	8%

This approach produced very different results. In Table 3, the three most robust measures were out-degree, in-degree, and *eigenvector*. Closeness and betweenness, which had been the most robust under the old procedure, became the worst performing. Yet this approach also has comparability problems. Scaling-up ranking systems like closeness and betweenness means a single change of position counts for more under these measures than under *eigenvector*. What is clear, however, is that *eigenvector* performed poorly in Table 2 as a result of providing a much broader spread of scores than other measures.

In addition to average changes of position, another way of testing the level of disruption is to look at how often an extreme change of position occurs in the rankings. Historians' use of social network analysis often does not cite precise numerical scores. Instead, comments are made on whether an actor's ranking is high, medium, or low. Given this, analysis becomes unreliable if actors are found to be jumping from one end of the hierarchy to the other as a result of incomplete datasets. The final row on Table 3 states what proportion of actors made an extreme jump in position. This is defined as more than a third of the way through the rankings, being sufficient to guarantee that an actor moves between the loose terms of high, medium, and low centrality. It is evident in Table 3 that in-closeness performed particularly poorly in this test.

Centrality and Pre-Modern Datasets

Removing the Berkshire collection of wills from the dataset caused notable changes to the centrality of the actors. Yet the disruption was not so dramatic as to make the outcomes of no analytical value. The evidence presented in this paper supports the use of social network analysis by pre-modern historians so long as the analysis is worded carefully. Under the out-degree measure, actors moved within the standardised scale of 1 to 80 by a median of 3.5 positions and a mean of 6.9. For *eigenvector*, the measure advocated by Costenbader and Valente, this was 5.1 and 8.9 respectively. While not an insignificant number of places, it does suggest that individuals tended to move within the same part of the hierarchy rather than leaping from one end to the other. More importantly, under all measures except in-closeness, extreme jumps in position occurred less than 10 per cent of the time. This suggests that, even with entire source collections missing, centrality will rank the remaining actors in a broadly meaningful position.

As noted, historical analysis often describes an actor's centrality as high, medium, or low, rather than citing centralities as a precise number. This paper's evidence suggests such descriptions are normally accurate, even when based on an incomplete dataset. Yet, while

[7] For closeness centrality, the lowest score achievable was not 0 but 0.111, a score which 439 actors held. Since this serves as the baseline score, it was treated as the equivalent of a 0 score under other measures of centrality.

most people did not jump between extremes, the fact that some did cannot be ignored. Hinging a historical argument on the centrality of a single actor would leave a historian's claims open to question. Rather, it is safer to analyse the average centrality of groups of actors who share an attribute, such as burgesses, officeholders, or members of a particular trade.

Having compared all the major measures of centrality, this paper gives an indication of the most promising ones for researchers. Each measure reflects different qualities in an actor but, given the nature of pre-modern source material, robustness in the face of missing data has to be a major consideration in choosing which measures to use. In terms of the average change of position, out-degree centrality proves most robust, but in terms of the frequency of extreme changes of position, in-degree centrality does. Robustness is not, however, a researcher's only consideration and degree centrality is less sophisticated than other measures in that it treats all the links in the network as equally valuable. Of the remaining measures, *eigenvector* proves most robust.

Conclusion

The loss of whole collections of source material is a problem faced by all pre-modern historians. If historical arguments are to be based on social network analysis, explorations should be conducted of the effect this has on the statistics produced. This paper has taken a first step towards the quantifiable testing that is currently lacking. The evidence presented here supports the use of measures of centrality if analysis is worded appropriately, but the field of historical social network analysis would benefit from further researchers conducting comparable exercises on their own datasets.

Bibliography

Borgatti, Stephen P., Kathleen M. Carley, and David Krackhardt. "On the Robustness of Centrality Measures under Conditions of Imperfect Data." *Social Networks* 28 (2006): 124-36.

Chick, Joe. "Urban Oligarchy and Dissolutioned Voters: The End of Monastic Rule in Reading, 1350-1600." *Cultural and Social History* 16 (2019): 387-411.

Costenbader, Elizabeth, and Thomas W. Valente. "The Stability of Centrality Measures when Networks are Sampled." *Social Networks* 25 (2003): 283-307.

Goddard, Richard. "Trust: Business Networks and the Borough Court." In Richard Goddard and Teresa Phipps, eds. *Town Courts and Urban Society in Late Medieval England, 1250-1500*, 176-99. Woodbridge, 2019.

Innes, Joanna. "Networks in British History." *East Asian Journal of British History* 5 (2016): 51-72.

Lee, Sang Hoon, Pan-Jun Kim, and Hawoong Jeong. "Statistical Properties of Sampled Networks." *Physical Review* 73 (2006): 016102-1-7.