The Journal of HISTORICAL NETWORK RESEARCH

SÉBASTIEN DE VALERIOLA

# Can historians trust centrality?

## Historical network analysis and centrality metrics robustness

*Journal of Historical Network Research 6 (2021) 85–125*

**Keywords** Methodology, Centrality, Robustness

**Abstract** In this paper, we consider four measures of centrality (betweenness, closeness, degree, and eigenvalue centrality) in their use for the analysis of historical networks. Since the sources used by historians to construct such networks are by their nature incomplete and imperfect, it is necessary to consider as much as possible the robustness of these metrics, i.e., their stability with respect to the hazards that time has inflicted on historical documents. To study this, we apply a battery of tests to three networks constructed from medieval history data. The first is a political history network, which represents the links between protagonists of the conflict for the episcopal see of Cambrai in the 11th century. The second is a network of socio-economic history, describing the credit relations of merchants in Ypres during the 13th century. The third is a hagiographic network that depicts the connections between the lives of saints that are often compiled together in manuscripts. These tests are designed to simulate the processes of disappearance and degradation of the information contained in sources by imitating as closely as possible the situations that historians face when manipulating graphs. In each of them, we create a large set of new graphs by transforming the original graphs, then observing the effect of these transformations on the centrality metrics. For this, we use a random process, but one that respects the particularities of the considered networks, which are built from historical sources. Our results allow us to assess the general relevance of the use of centrality in historical network analysis, to compare the four metrics studied in terms of robustness, and to identify a set of methodological points to which the historian applying such techniques must pay particular attention.

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

## 1. Introduction*

In 1933, when Henri Pirennne describes *what historians are trying to do,* he compares the historian's research material with "footprints in the sand which wind and rain have half-effaced", arguing that they are "merely the vestiges of events and not even authentic vestiges".[1] This observation by the great medievalist is obvious to modern professional historians: pursuing historical research involves working with incomplete and imperfect sources. Not all the pieces of the puzzle that historians put together to describe the societies of the past are available, and some are damaged, having suffered from the vagaries of time. It is therefore essential to be particularly careful when analyzing historical documents, to consider as much as possible their imperfect condition, and the hazards they have experienced. It is at this price that the conclusions drawn from historical studies can be considered as reliable, as it is a central element (if not *the* central element) of the historical method.

### 1.1 Robustness and Historical Analyses

This rule applies to all types of analysis that historians subject their sources to. Nevertheless, in the case of quantitative analyses, researchers have at their disposal a set of mathematical tools that allow them to assess the reliability of the results obtained with regard to the defects of the documents studied. Among them is robustness, a concept well known to statisticians and more generally to researchers in the exact sciences, but rarely used in the humanities, even when quantitative methods are applied. In his classic book devoted to the question, Peter Huber defines robustness as the "insensitivity to small deviations from the assumptions".[2] He uses the term "assumptions" to cover a wide range of modeling choices, which we will reduce here to the set of constraints that are imposed on the historian carrying out a quantitative analysis by the condition of the sources he handles. Gaps and errors in the available documents are the source of deviations in the results obtained (with respect to the correct depiction of the studied phenomenon), the effects of which can be at least partially mitigated by using a robust quantitative analysis tool.

Before getting to the heart of the matter and presenting the actual framework of our study, let's see how this concept comes into play in the context of a very

* **Acknowledgements:** I warmly thank my colleagues Céline Engelbeen and Étienne Cuvelier from the *Laboratoire Quaresmi* for their valuable feedback on this paper, as well as Antoine and Arthur de Valeriola.
**Corresponding author:** Sébastien de Valeriola, Université libre de Bruxelles (ReSic) and ICHEC Brussels management school (Laboratoire Quaresmi); sebastien.de.valeriola@ulb.be

1 Pirenne (1933), p. 438.
2 Huber/Ronchetti (2009), p. 2.

simple example of a quantitative analysis of historical sources. Suppose we want to estimate the typical amount (statisticians would say the "central tendency") of loans made by a merchant in some medieval city. We have at our disposal a set of credit contracts in which the merchant appears as the creditor, specifying each time the amount of money lent to the borrower. The first obvious way to estimate this typical amount is to calculate the arithmetic mean of the loans granted: for example, based on four amounts of 88, 95, 99 and 118 pounds, we would have an arithmetic mean of 100 pounds. Now let's imagine that a fifth loan contract resurfaces after being misclassified, in which our merchant lent 600 pounds, a very large amount compared to the first four. The arithmetic mean, recalculated on the basis of the five amounts now available, is equal to 200 pounds, double the value it had before the document was rediscovered. This high sensitivity to the addition of extreme data – or in other words, a lack of robustness – is one of the drawbacks of the arithmetic mean as a statistical indicator of central tendency. To overcome this problem, one might consider using the median rather than the mean to estimate the typical amount of the merchant's loans. This second statistical indicator is robust, and changes little when a value is added to the data, whether it is extreme or not: the median rises from 97 pounds to 99 pounds when the rediscovered loan contract is taken into account. The superiority of the median over the mean is clear in such a context. However, it is not necessary to consider adding an extreme additional value to the dataset under consideration to reach the same conclusion. This hierarchy between the two indicators can also be seen by thinking in terms of the quality of the estimate made in the specific context of an analysis of historical data: on the one hand, the median is much more stable than the mean with respect to 'forgotten' data; on the other hand, the dataset we are analyzing is necessarily fragmentary, since it is extracted from historical sources. The choice is therefore quickly made between the two indicators.

This example, although extremely simple, shows that robustness appears as a highly desirable quality when estimating the typical value of a quantity appearing in historical sources. The same conclusion can of course be drawn about any quantitative tool mobilized in any historical analysis. This is therefore also the case for the tools historians use to analyze social networks.[3] The issue of robustness is perhaps even more crucial in the application of these techniques. The fragmentary and imperfect condition of the sources is indeed an objection that is sometimes raised by historians when it comes to making use of these methods.[4]

3 Note that we will use the terms 'network' and 'graph' interchangeably throughout this paper. The same holds for 'node' and 'vertex'.

4 Often because of confutations about the implicit assumptions they presuppose, as noted by Lemercier (2015), p. 296.

### 1.2 Centrality Metrics

In this paper we look at the robustness of a set of metrics often used in historical network analysis to estimate the status of individuals within their network, the measures of centrality.[5] Since the first works devoted to this concept[6] and the seminal studies in which its first rigorous definitions were introduced,[7] numerous versions of it have been proposed in the literature, in order to question the importance of vertices within the graph from different angles.[8] However, historians most often focus on four of them, which we will consider here.[9] Betweenness centrality counts the number of geodesics in the graph (i.e., the shortest paths along the edges of the graph) that pass through a given vertex. Closeness centrality calculates the distances between a given vertex and all the other vertices of the graph, and aggregates them into a synthetic indicator, defined as the inverse of the sum of all these distances. Degree centrality counts the edges of which a given vertex is one of the two ends. Eigenvector centrality assigns a score to a given vertex on the basis of the scores assigned to its neighboring vertices, according to the principle that this score is high when the neighbors themselves have a high score. Linear algebra tools can assign all these relative scores at once.[10]

Each of these four metrics is a tool for estimating the importance within the network of each of the individuals who are part of it. While their general objective is identical, they do not measure exactly the same thing, and therefore differ in terms of interpretation. The interpretation that can be made depends on the context in which they are used, and the choices made to build the graph being considered (what do vertices and edges represent?), but the definitions given above still allow to draw some general principles, which we will mention very briefly here. A vertex with a high betweenness centrality score corresponds to a 'hub' (also sometimes called a 'broker'), a node through which a large number of connections between individuals in the network can pass. A high closeness centrality value indicates that the vertex can easily reach all parts of the network. The eigenvalue centrality measures the prestige of an individual, in terms of the

5 To our knowledge, only one study is devoted to the analysis of the properties of these metrics in the framework of a historical analysis: Düring (2016). This author's point of view is quite different from ours, since his goal is to compare the list of the most important individuals within a network obtained by calculating the centrality metrics with that which the historian obtains manually on the basis of his expertise concerning the dossier in question.

6 Bavelas (1948); Bavelas (1950).

7 Bonacich (1972); Freeman (1979).

8 See for example Das/Samanta/Pal (2018). A list of nearly 300 centrality definitions (at the time of writing) is given in Jalili/Salehzadeh-Yazdi/Asgari/Arab/Yaghmaie/Ghavamzadeh/Alimoghaddam (2015).

9 See e.g. Hammond (2017); Rosé (2011); Riva (2019); Cellier/Cocaud (2012).

10 For a more formal definition of these four metrics, see for example Wasserman/Faust (1995).

number of connections with prestigious individuals. These three metrics are global measures, in the sense that they account for the entire graph. On the contrary, degree centrality is a local measure, which considers only the direct relations of the vertex in question.[11] It estimates the importance of an individual by measuring his activity in the network. These four metrics thus carry quite different meanings, which can be combined to perform precise analyses (which individuals are central to the four measures, which are central only to a subset of them and why, etc.).

They are also used at a different level, that of whole graphs. It is indeed possible to aggregate the centrality scores of all the vertices of a network to calculate its centralization.[12] This indicator estimates the extent to which the graph is globally organized around one or more focal points, i.e. it accounts for the existence of extreme values among the individual centrality scores of its vertices. A star-shaped graph has a very high centralization, while a complete graph (with all vertices connected to all the others) is associated with a low value. The four centrality measures lead to four different centralization concepts.

### 1.3 Robustness and Centrality in Historical Networks

We are interested here in the robustness of these centrality metrics when computed in historical network analyses. However, since they are more complex than central tendency indicators such as mean and median, there is no theoretical result to assess their robustness. It is therefore necessary, in order to meet this objective, to embrace the experimental approach, by observing the impact of "small deviations from the assumptions" on the values taken by these measures of centrality. This exercise has already been carried out in graph theory literature, and has led to several interesting studies.[13] However, these reasoned within a general framework and are therefore not very well adapted to historical analyses. In order to compare the reliability of these network metrics and convince historians, it is necessary to perform these tests on historical networks, especially with "small deviations from the assumptions" that make sense in the context of historical analysis.

This is the task we assign ourselves in this article. We carry out robustness tests on the measures of centrality mentioned above, with three networks constructed from medieval sources. The first is a political history network, which represents the links between protagonists of the conflict for the episcopal see of Cambrai in the 11th century. The second is a network of socio-economic history,

11 This distinction is discussed for example in Scott (2000).
12 For details about this concept and its computation, see Freeman (1979), p. 226–237.
13 See, in addition to the references given in the bibliography of this article, the review Landherr/Friedl/Heidemann (2010).

describing the credit relations of merchants in Ypres in the 13th century. The third is a hagiographic network that depicts the connections between lives of saints that are often compiled together in manuscripts. The test methodology we implement has been designed and constructed to replicate the problems faced by the historian due to the condition of the sources he handles.

The main objective for carrying out these robustness tests is to provide answers to two questions. First, we question the confidence historians may have in the metrics of centrality, given the structurally incomplete and inaccurate nature of the historical documents that serve as the basis for the construction of the networks they study. Are these measures of centrality sufficiently stable when subjected to 'shocks' that replicate the vagaries of historical sources? Can the conclusions drawn from them be considered sound? Second, we ask questions to compare these metrics in terms of robustness. Are some of them more robust than others to these shocks? Are some of them better suited for use in the context of analyzing historical networks?[14]

We have structured the battery of tests we perform into four experiments designed to explore different aspects of the issue of centrality metrics' robustness. Our methodology is developed in detail in Section 3 of the article, following a presentation of the data used in Section 2. Section 4 is devoted to the results that it permits us to obtain, which are commented upon and discussed. A concluding point is given in the last section.

## 2. Data

This section is devoted to the description of the three networks that we consider in this article, and upon which our tests are performed. Our choice of these three examples was not random: as we will see at the end of this section, they have quite distinct profiles, suggesting that test results could differ significantly.

### 2.1 The Cambrai Investiture Conflict

The first network that we consider was built by Nicolas Ruffini-Ronzani to model the Cambrai Investiture Conflict.[15] At the end of the 11th and the beginning of the 12th century, in the context of the Gregorian reform, two politico-religious personalities clash for the episcopal see of Cambrai: on one side is Walcher of Oisy,

14 The search for the most appropriate metric for a particular context has already been undertaken in the literature. See, for example, for the dissemination of information in telecommunication networks, Kiss/Bichler (2008).

15 For a detailed description of the historical background and additional information about this network, see Ruffini-Ronzani (2020).

the candidate of the emperor; on the other Manasses of Eu-Soissons at first, then Odo of Tournai from 1103 onwards, supported by the pope. Since the beginning of the conflict in 1092, the two parties have been fighting a real war, which the Treaty of Aachen put an end to in 1107, to Walcher's disadvantage.

Several chronicles recount this conflict, but these are not the only sources that a historian can mobilize to study it: a fairly large number of charters (of which 176 were kept in good enough condition to be used, dated from 1092 to 1107) testify to legal actions undertaken during this period by members of both sides. This diplomatic corpus is the material used to build the network we consider in this article.

This graph is defined as follows. Its vertices represent the persons who appear in these charters. An edge joins two vertices each time they have one of the following relationships in a charter: Alliance (X enters into an alliance with Y); Consent (X consents to an action of Y); Donation (X gives a property to Y, or confirms such a donation); Notice (X gives notice about an action of Y); Request (X requests from Y to take some action); and Subscription (X appears among the subscribers on Y's charter). The edges of the network originally constructed by Nicolas Ruffini-Ronzani thus bear a *type* attribute, which we neglect in the context of this paper.[16] Note that these links, which are oriented by their nature, are considered to be non-oriented for the purposes of this article. An attribute *source* is also associated with each of the edges, which gives the identifier of the charter that attests to the relationship that the edge represents.

### 2.2 The Ypres Credit Market

The second network that we use in our experiments models the Ypres credit market in the second half of the 13th century[17]. This century is a period of economic prosperity for the Flemish city, at least in part thanks to the then flourishing textile industry. As in most medieval cities during this period, the lively market in Ypres is not without intense credit activity.

A large number of loan contracts are concluded between all kinds of individuals, including wealthy foreigners who come to buy cloth, local entrepreneurs who sell it, and the city's smallest artisans. At least a portion of such credit arrangements are subject to the gracious jurisdiction of the city eschevins, and are recorded in writing in the form of chirographs. Until the beginning of the 20th century, the archives of the city of Ypres held several thousand of these recognizances of debts, which unfortunately almost entirely disappeared during the bombing of the First World War. Around 1900, a local scholar, Guillaume des

16 The question of the simultaneous consideration of these edges of different types is treated in de Valeriola/Ruffini-Ronzani/Cuvelier (2021).

17 Details about this network are given in de Valeriola (2019).

Marez, nonetheless took note of summaries of many of these recognizances in notebooks, which the *Commission royale d'Histoire* recovered and edited.[18] The information provided describes 4,953 usable loan contracts, dated between 1249 and 1291, and includes, among other things, the names of all the parties involved, i.e. the creditors, debtors and guarantors of the corresponding loans.

These allow the construction of a graph in a similar way to that described above for the Cambrai graph. Vertices are individuals involved in at least one loan contract. An edge joins two vertices each time the two individuals in question are related in one recognizance of debt, regardless of the type of relationship involved (creditor-debtor, debtor-debtor, creditor-guarantor, etc.). The direction of links is once more neglected. As in the case of the Cambrai graph, the edges carry an attribute *source,* which gives the identifier of the chirograph from which the information carried by the edge is extracted. Finally, we associate an attribute *amount* to the edges, which gives the amount (expressed in Artesian pounds, the currency most used in our recognizances of debts) of the corresponding loan. This is used in only one of the four experiments we carry out (Experiment 3); in the other three, it is simply ignored.

### 2.3 The Co-tradition of Hagiographic Legends

The last network we study here models the co-tradition relationships of saints during the Middle Ages.[19] Hagiographic narratives are very often the subject of compilations, in which the legends follow one after the other. The precise selection of the texts that are compiled together is not entirely due to chance, and it is legitimate to wonder how the copyists, who are at the origin of the sanctorals, chose the saints whose stories they tell.

To investigate this question, we used the database *Bibliotheca Hagiographica Latina manuscripta.*[20] Created by the Bollandists, then extended by several researchers and now hosted in its new form by the *Institut de Recherche et d'Histoire des Textes,*[21] it lists a set of several thousand Latin hagiographic manuscripts, and gives various information for each of them, including the list of legends it contains.

Our hagiographic graph is built on this basis. Its vertices represent each of the saints for which at least one manuscript contains a legend. An edge connects two vertices each time there is a manuscript that contains a text about each of them. As previously, we associate an attribute *source* to the edges, giving the identifier

18 Wyffels (1991).
19 Further details about this dossier can be found in de Valeriola/Dubuisson (2021).
20 On this database, see Trigalet (2001).
21 The link to the *Légendiers latins* database will be available soon.

of the manuscript in which the co-tradition relation that the edge represents is found. Each edge also bears an attribute *century,* which gives the date of the manuscript in which it is attested. It is used in only one of the four experiments we carry out (Experiment 4); in the other three, it is simply ignored.

Using the entire database leads to the construction of a graph composed of 2,498 vertices and 1,487,563 edges. This huge number of edges makes calculations very long and very difficult to manage in terms of machine resources (see Section 3.4 on this subject). We therefore applied a filter to this gigantic graph: the hagiographic graph manipulated throughout this article is constituted from the manuscripts of the database whose place of conservation is Paris and which date from the 8th to the 15th century. While this means of selecting sources is of course objectionable in terms of historical analysis, it has no impact on the present methodological study.

### 2.4 Comparison of the Three Networks

It is natural, if we want the results of our robustness tests to be representative, to apply them to networks that differ significantly from each other. Indeed, the stability qualities of metrics can be expected to depend on the properties of the graphs from which they are calculated.[22] Table 1 presents information on each of the graphs we manipulated, allowing us to compare their main characteristics. It is to be combined with figures 1 and 2.[23] The first gives the densities of the four normalized centrality metrics (based on their theoretical maximums, see Section 3.1) for each of the three networks, and thus gives an idea of the distribution of individual vertex centrality values. On the second, the normalized centralization values are represented (i.e. the centralization values divided by the maximum value over the three networks), making it possible to compare the internal structure of the three networks.

The numbers presented indicate that the three networks have very different profiles. The Cambrai network is the smallest of the three in terms of the number of sources, number of vertices and number of edges. As might be expected due to the nature of the historical phenomenon it models (a conflict opposing two parties, each of which gathered around one or two individuals, candidates of both parties to the episcopal see), its centralization value is very high for betweenness and degree centralities. The existence of a small number of vertices with high centrality values for these two metrics is clearly visible on the density plot. We also see on this plot (for betweenness centrality) that many vertices can be considered

22 See for example Borgatti/Carley/Krackhardt (2006), p. 124; Frantz/Cataldo/Carley (2009).

23 Note that all centrality values presented here have been computed on the graphs after applying to them the simplification process described in Section 3.1.

| | Cambrai | Ypres | Hagio. |
|---|---|---|---|
| number of source units | 176 | 4,953 | 611 |
| number of vertices | 400 | 4,675 | 1,118 |
| number of edges | 1,419 | 12,012 | 229,672 |
| number of pairs of vertices linked by at least one edge | 685 | 11,050 | 92,667 |
| edge density | 0.008 | 0.008 | 0.143 |
| unweighted diameter | 6 | 14 | 5 |
| average unweighted distance | 2.82 | 5.01 | 1.92 |
| transitivity | 0.02 | 0.15 | 0.54 |

**Tab. 1** Main properties of the three networks we consider

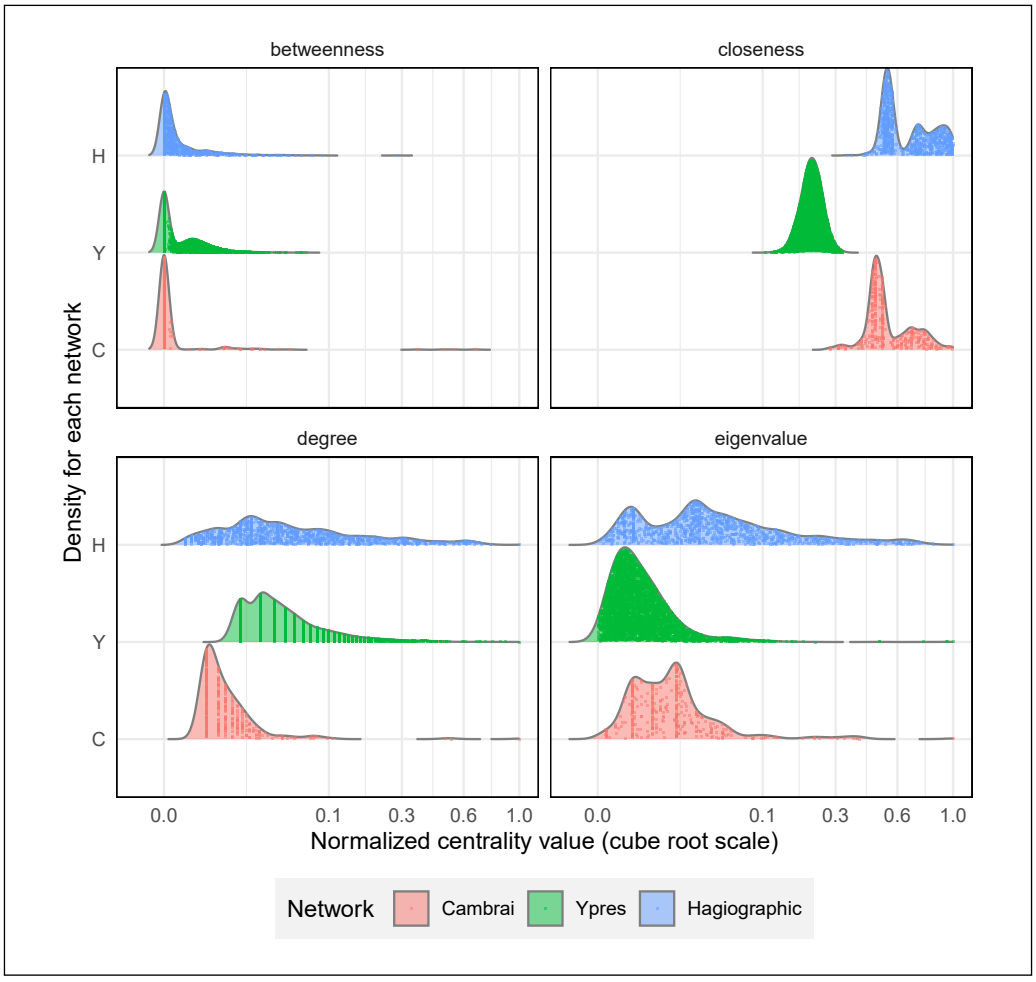


**Fig. 1** Densities of centrality metrics of the three networks

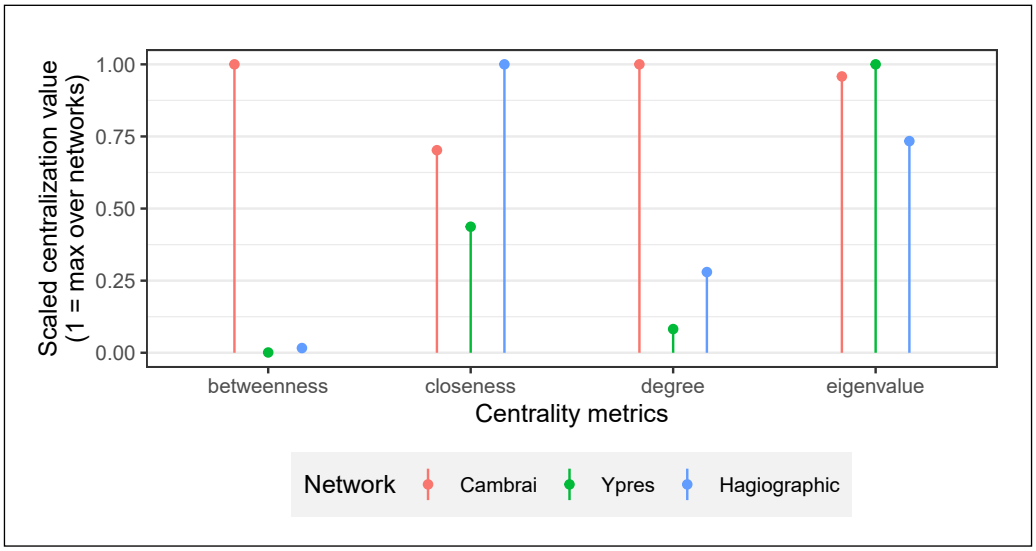


**Fig. 2** Centralization of the three network with respect to the four centrality metrics

as belonging to the periphery of the network, i.e. are not a crossing point of any geodesic.

The Ypres network has the largest number of vertices and is built from the largest number of sources. The moderate value of its transitivity coefficient (the proportion of 3-cliques among the triplets of vertices X, Y, Z such that X is connected to both Y and Z), its large diameter and the high average distance between its vertices suggest that it is composed of a fairly large number of small clusters that are quite well separated from each other. This description is consistent with the way it was constructed, since all the creditors, debtors and guarantors of each chirograph are completely interconnected. Each recognizance of debt therefore corresponds to a clique (the size of which depends on the number of protagonists in the agreement). Note, however, that there is a set of well-connected vertices that act as the 'center' of the graph, as the high level of eigenvalue centralization and the eigenvalue centrality density plot suggests.

The hagiographic network has by far the highest number of edges, and therefore the highest density. Again, the graph construction process explains this: each manuscript is responsible for creating a clique whose size is equal to the number of legends it contains. Among our Parisian sanctorals, 6 include at least 100 legends, and 66 of them at least 50. The resulting graph is therefore the superposition of a set of cliques of rather large sizes. This explains the high value of its transitivity coefficient, its small diameter, the small average distance between its vertices (and consequently its high closeness centralization value), and the 'spread' shape of the densities of degree and eigenvalue centralities.

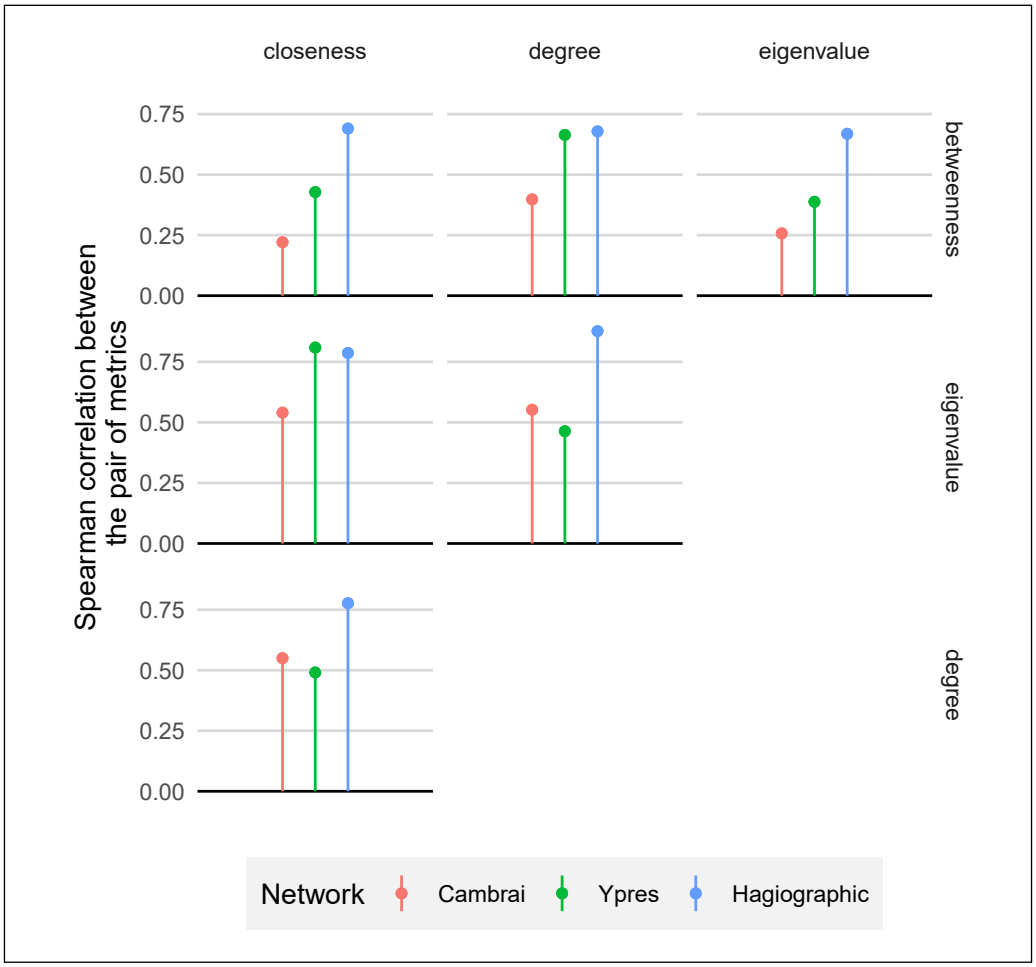


**Fig. 3** Spearman correlation between the six pairs of centrality metrics in the three networks

Finally, Figure 3 gives the value of Spearman's correlation coefficients between the six pairs of centrality metrics.[24] It is interesting to observe that these coefficients change quite a bit from one graph to another, although they are always positive.[25] We can conclude that the amount of information provided by (and

24 Spearman's correlation coefficient measures the dependence between two statistical series by comparing the ranks of each value within the two series. This metric takes values between –1 and 1: a positive value indicates that the two series are 'moving in the same direction', a negative value that they are 'moving in opposite directions'. In the case we consider, it therefore compares to rank each centrality metrics assigns to the network's nodes. For a justification of this choice, see Section 3.3.

25 As the literature has already noted, for example Oldham/Fulcher (2019).

therefore the interest of) the joint use of several metrics of centrality also varies with the graph's properties: the correlations are far from perfect (with coefficients close to 1), and the metrics are therefore not completely redundant. The high values of these dependency indicators for the hagiographic network probably come from its high density.[26] For the Cambrai graph, the correlation between betweenness centrality and the other three metrics can probably be explained by the large number of zeros among the values of the first.

## 3. Methodology

Let us now describe the experiments we perform to estimate the robustness of centrality metrics.

### 3.1 General Remarks

Before doing so, a series of general remarks should be made about the way metrics and networks are handled in this paper. First of all, let us note that the comparison of the four metrics applied to the three graphs described in the previous section imposes two technical constraints upon them. On the one hand, the betweenness centrality score of almost all vertices of the Cambrai graph is equal to 0 if we consider its edges as directed. The comparison with other metrics loses part of its interest in this case; this is why the Cambrai and Ypres graphs, although in principle oriented, are considered in their non-oriented version.[27] Note moreover that comparing the centrality metrics of two oriented graphs with those of a non-oriented graph would make little sense.

On the other hand, closeness centrality is well defined only for connected graphs, i.e. those where there is a path (a succession of edges) starting from any vertex and arriving at any other vertex (since it implies the calculation of the distance between one vertex of the graph and all the others). The three base graphs described above are connected. Nevertheless, since some of the experiments we perform involve the disappearance of edges within them, the result may no longer be connected. It is therefore necessary to transform these networks into connected ones before performing our calculations. In this case, we simply restrict ourselves to the largest connected component of the graph.[28]

---

26 This positive relationship has already been observed by Valente/Coronges/Lakon/Costenbader (2008), p. 6.

27 This way of making the adjacency matrix symmetrical is classical, see for example Costenbader/Valente (2003), p. 289.

28 This is also the solution adopted in Platig/Ott/Girvan (2013), p. 2. Other solutions exist, for example, we could simply not consider simulations in which the graph becomes unconnected, as in Costenbader/Valente (2003), p. 290.

Second, as we will see below, we must note that our methodology is based on operations that require the selection of all the edges of the graph considered that correspond to a given subset of sources. This is why a *source* attribute is associated to the edges of the three studied graphs. Its importance for the tests we perform prevents us from directly aggregating the multiple edges that are present in the three graphs. By 'multiple edges', we mean here the situation in which two vertices X and Y are directly connected by several 'parallel' edges (as are vertices a and c in Figure 4), a situation that is observed many times in the three networks. The parallel edges that connect X and Y in this way are attested in different sources, which is information that we lose if we aggregate them directly.

The aggregation of these multiple edges is nevertheless a step that must be taken to calculate the centrality metrics. This operation is performed just before this calculation, so as not to interfere with the attribute *source*. It simply consists of replacing multiple edges with a single edge bearing an attribute *weight* that counts the number of multiple edges it replaces. For example, if vertices X and Y are connected by 4 edges before aggregation, they will be connected by a single edge of weight 4 after aggregation. In the special case of the Ypres graph in Experiment 3, the role of the attribute *weight* is played by the attribute *amount*. The aggregation phase thus does not simply count the parallel edges, but calculates the sum of their amounts. If the 4 edges in our example are 5, 7, 12 and 21 pounds, the single edge representing them will have a weight equal to 5 + 7 + 12 + 21 = 45 pounds after aggregation. Finally, let us note that the other attributes are dropped during this step.

These two operations of aggregation and restriction to the largest connected component together form the operation of graph simplification. In the rest of this article, when we say that a graph is simplified, we must therefore understand that its multiple edges are aggregated and that it is replaced by its largest connected component.

Figure 4 shows the application of this process to a very simple example graph. The original graph (on the left) is constructed from two historical sources (n° 1

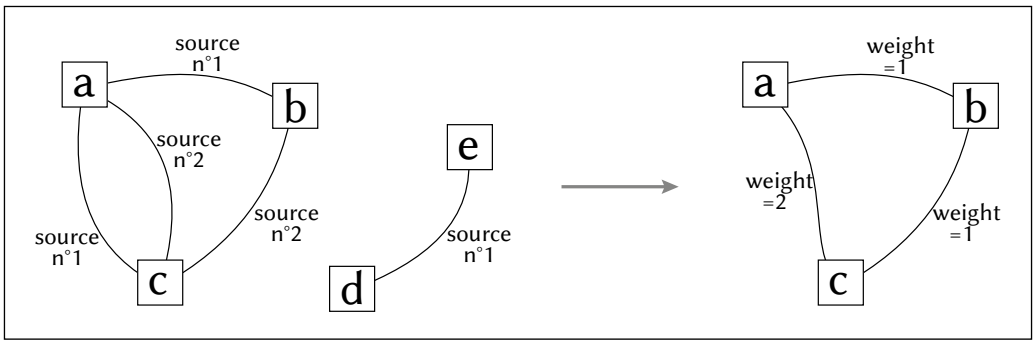


**Fig. 4** Application of the simplification process to an example graph

and n° 2), information which gives rise to the edge attribute *source*. The two parallel edges linking vertices a and c are aggregated within the simplified graph (right) into an edge with weight 2. The two edges joining vertices b and c on the one hand, and a and c on the other, give rise to edges with weight 1. The vertices d and e, as well as the edge connecting them, are deleted in the process, since they form a non-maximal connected component of the original graph.

Third, and for the reasons we have just presented, we use alternative versions of the centrality metrics that are adapted to weighted graphs. The definition of betweenness and closeness centralities is easy to modify for this purpose, since it is sufficient to include the weights into how the distances are calculated.[29] For degree centrality, we sum the weights of the edges incident on the vertex, rather than simply counting them. Because the eigenvector centrality is somehow only a mathematical property of the adjacency matrix of the graph, adding weights to the edges is not a problem (it simply modifies the matrix).[30]

Fourth, since the networks being compared do not necessarily have the same structure (i.e. the same number of edges and vertices), it is necessary to use standardized versions of these metrics. Therefore, for each of these, it is a matter of dividing the result obtained by the theoretical maximum that the metric can reach given the structure of the graph. Standardized metrics take values between 0 and 1.

These remarks concern all centrality computations, whether they are applied to unshocked or shocked graphs.

### 3.2 Experiments

In this article we carry out four different experiments on the basis of the networks defined above. The general idea of these experiments is as follows: we subject our three graphs to shocks intended to replicate the hazards that historical sources undergo, then compare the centrality metrics of the graphs resulting from these shocks (which we will henceforth call the shocked graphs) to those of the original graphs (the unshocked graphs). We are speaking here about defects in the data manipulated by the historian in two ways: the incompleteness of the data on the

29 Note that the edge weights in our simplified graphs correspond to 'link strengths' (i.e. a high value corresponds to a strong relationship between the two vertices connected by the edge) and not to 'link costs' as expected when calculating a distance. They must therefore be transformed before being used in this way. We use the inverse function to do this: cost = 1 / weight.

30 For a description of the adjacency matrix of weighted graphs, see Wasserman/Faust (1995), p. 153.

one hand (Experiment 1), and the inaccuracies in the data on the other (experiments 2, 3, and 4).

*Dealing with Incomplete Data*

To estimate the robustness of the network metrics with respect to the first aspect, we will simulate the incompleteness of our datasets by removing some of the available data. This type of robustness test is relatively common in studies devoted to centrality metrics.[31] Indeed, historical networks are not the only ones to suffer from the phenomenon of incompleteness: in many fields of application, the data used to build networks are fragmentary.[32] In most of the studies in the literature on this issue, "removing some of the available data" means randomly (and uniformly) removing some of the vertices and/or edges of the graph.[33] A set of stochastic scenarios is thus generated: in each of them, a different part of the set of vertices or edges is removed, thus simulating a different set of shocks. The metrics values are then recalculated on the shocked graphs that result from these random deletions (one for each stochastic scenario), and compared to the metrics values for the unshocked graph. Conclusions about stability can then be drawn from the observed differences. In probabilistic formalism, this process can be seen as a Monte Carlo simulation: to estimate the expected value of the robustness of a metric when the corresponding graph is subjected to a random perturbation, a large number of realizations of this perturbation are generated and the arithmetic mean is calculated.

Nevertheless, this method of randomly deleting vertices and/or edges is not suitable for all application contexts, as some authors have already noted.[34] In particular, it is not at all suitable for historical networks. Indeed, they are built in a specific way: by accumulating information from historical sources, each of which brings a 'cluster' of information to the network. The incompleteness of the infor-

31 See for example Bolland (1988), which, to the best of our knowledge, is the first study to evaluate the robustness of centrality measures in a systematic experimental manner, as well as the other references cited below.

32 Landherr/Friedl/Heidemann (2010), p. 373, col. 3.

33 It is also frequent to see another operation on the graphs considered, although one which makes less sense in the context of a historical network: the addition of random vertices and/or edges (note however that in our Experiment 3 some vertices are added to the graph, see below). For these four operations of deleting and adding vertices and edges, see for example Bolland (1988); Borgatti/Carley/Krackhardt (2006); Galaskiewicz (1991); Costenbader/Valente (2003); Tsugawa/Ohsaki (2015).

34 See, for example, concerning networks constructed from interviews of community members, the remarks made in Costenbader/Valente (2003), and concerning transport networks during Roman antiquity, Groenhuijzen/Verhagen (2016). Note also that some authors do not select the edges and vertices to be deleted according to a uniform draw, but by associating deletion probabilities to the edges and vertices which depend on their properties within the graph: Platig/Ott/Girvan (2013).

mation used to build such a network stems from the incompleteness of the available sources. It is therefore natural, rather than randomly removing edges and/or vertices in these networks, to randomly remove elements from the document set used to construct the graph. For example, each Cambrai charter makes it possible to add to the network several edges (when one takes into account the relations between the author, the disposer, the witnesses, etc., learned from the document), and possibly one or more vertices. Randomly deleting edges and/or vertices would thus be tantamount to deleting only certain parts of these information clusters, a process that does not suit the issue of the absence of certain sources with which the historian is confronted. It is thus necessary, if one wants to decrease the number of edges and/or vertices of the graph of Cambrai, to do so charter by charter.

We therefore apply to the studied graphs shocks that are distinct from those applied in previous studies, but we keep the randomness of the process, a point that deserves a brief comment. The methodology we implement aims to imitate the vagaries of historical sources, some of which are preserved while others are not. These hazards, whose succession constitutes the path of the documents through time, can certainly be considered as deterministic. Events such as theft, misclassification, destruction and other disappearances of sources do not really happen by chance, but are the consequences of particular circumstances that can probably be explained. It is therefore natural to ask why the use of random draws is relevant in the context of a historical analysis. Although the course of the sources is deterministic, it is not known to us in its entirety, and in the vast majority of cases, we know nothing of the documents that have not been preserved (their number, content, etc.). Randomness is used here to model those elements that cannot be known or predicted. This way of using a set of stochastic scenarios to model the unknown and the unpredictable is very common in exact sciences, and has led to intense epistemological reflections on the notion of randomness.[35] Finally, it should be noted that the opposite approach, which would consist of not selecting the sources to be deleted randomly but on the basis of deterministic criteria, does not provide a meaningful estimate of the robustness of the metrics. Consider for example the Ypres graph, and a process of deleting the chirographs that would be based on their date. To estimate the robustness of the centrality metrics, one would, for example, remove all recognizances of debts written over a certain period of time (e.g. we would remove all the acts of 1290–1291, keeping only those of 1249–1289) and compare the results with the measures in the unshocked graph. The downside of this method is obvious: as it is very unlikely that the network is homogeneous from a chronological point of view, removing a set

35 For example, Henry Kyburg writes that "the concept of randomness [...] is relative to our body of knowledge, which will somehow reflect what we know and what we don't know". (Kyburg (1974), p. 217). A recent summary of the epistemological discussion is given in Eagle (2005).

of successive chirographs amounts to removing from the graph a part that perhaps has its own characteristics, different from the rest. The comparison with the unshocked graph would thus be meaningless. The same observation could have been made if any other selection criteria had been used, such as the geographical origin of the creditor.

*Experiment 1*

Let us now describe more formally the design of our first experiment, which tests robustness against data incompleteness. The first step is to choose the proportion of sources that will be removed from the set of documents used to build the network. Let's say that it is equal to 10%: in this case, we decide to remove 17 charters for Cambrai, 495 chirographs for Ypres and 61 manuscripts for the hagiographic graph. We then generate 1,000 different stochastic scenarios. For each of them, a set of documents determined by a uniform random draw and corresponding to 10% of the total available documents is erased, and a graph is constructed based on the information contained in the remaining 90% of documents. Figure 5 summarizes this process.

The truncated graph is then simplified and the values of the four centrality metrics are calculated. In this way, we obtain 1,000 shocked graphs and therefore 1,000 sets of values for the metrics. Experiment 1 is finally repeated several times with different proportions of sources removed (1%, 2%, 5%, 10%, 20% and 40%), in order to see how the metrics evolve when the loss of sources is more and more substantial. Table 2 lists the successive steps of Experiment 1.

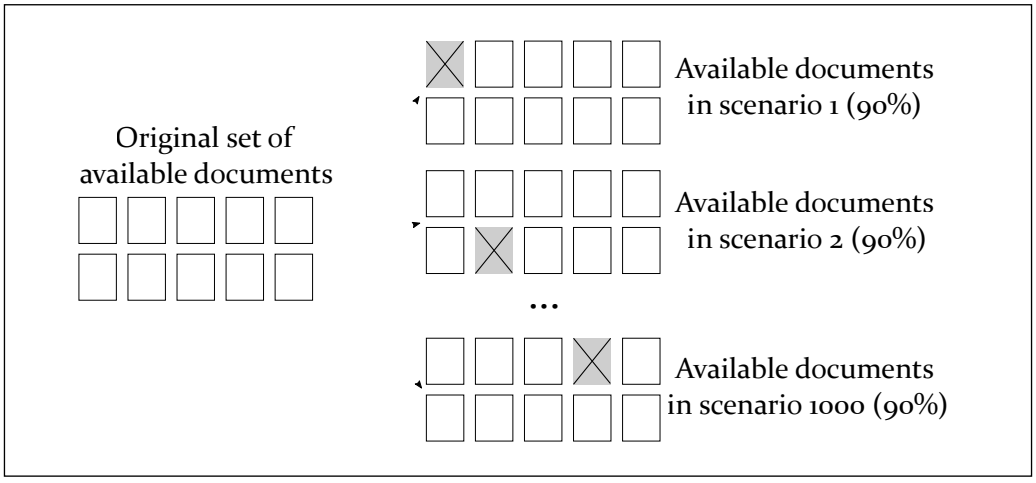


**Fig. 5** Description of Experiment 1

1. Fix a proportion of sources to remove in {1% ; 2% ; 5% ; 10% ; 20% ; 40%}.
2. For each of the 1,000 scenarios:
   a. use a random draw to select the sources to remove,
   b. delete from the graph all the edges whose *source* attribute is among the selected sources,
   c. simplify the resulting shocked graph,
   d. compute the four centralities on the simplified graph.

Output: the values of the four centrality metrics for each vertex of the unshocked graph (except for those which disappear in one of the simulations) in each of the 1,000 scenarios, for each affected sources proportion.

**Tab. 2** Successive steps and output of Experiment 1

*Dealing with Inaccuracies*

Assessing the robustness of network metrics against the existence of inaccuracies in the data or working hypotheses is a larger and more complicated task. Indeed, the potential errors that may appear in the course of the historian's analysis are of very diverse natures, and have their origins in phenomena that are also very diverse. These inaccuracies are sometimes the direct result of archive producers, such as when a scribe makes a mistake – intentional or unintentional – in a document. They can also be made by the historian who transcribes, or by the optical character recognition software he uses, which would transform one word into another. Other inaccuracies are direct consequences of the researcher's working hypotheses, for example simplifications of the reality being studied that do not take into account this or that aspect of the information contained in the sources. Since it is obviously impossible to deal with all eventualities, we have chosen to restrict ourselves to the three inaccuracies that seem to us to be the most important and most frequent in the analysis of historical networks.

The remarks made concerning the incompleteness of the data can also be formulated for the inaccuracy of the data. While the shocks we apply to the graphs in experiments 2, 3 and 4 are different from those in Experiment 1, the same general principles apply.

*Experiment 2*

In Experiment 2, we look at the working hypotheses related to the identification of the individuals composing the network.[36] This problem, which is faced by many researchers using historical network analysis on medieval data, is twofold. Very often, there is no reason to believe that two mentions of the same anthroponym indicate the same individual, and not two individuals with the same

36 This problem, along with possible solutions, is presented in de Valeriola (2021).

name. Conversely, mentions that designate the same individual often appear with variations in spelling that are sometimes difficult to reconcile. An extreme example would be surnames translated from one language to another in some sources, such as "Jean de Neuveglise", who also appears as "Jan van Nieukerke" in the Ypres sources (both forms are equivalent to "John of Newchurch", in French and Dutch).

To replicate problems of this type, we perform two types of operations on graphs, which we will refer to as experiments 2a and 2b. First, we merge pairs of vertices, to replicate the inaccuracy of considering that two anthroponyms designate two different individuals, where in reality they designate only one (Experiment 2a). The vertex resulting from such a fusion has as its neighbors all the neighbors of the two original vertices: in a sense, it gathers their edges. Figure 6a illustrates this operation: vertices *a* and *d* merge into a single vertex called *a + d;*

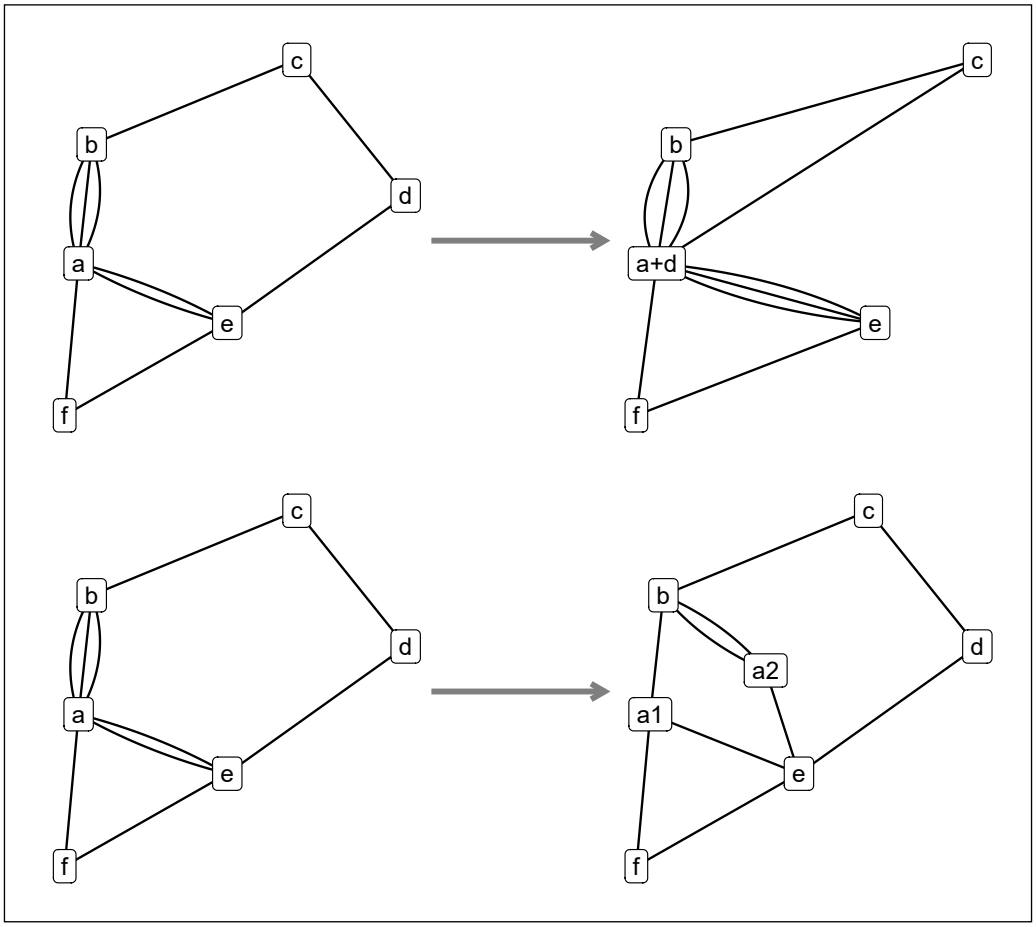


**Fig. 6** Examples illustrating the two operations of Experiment 2: 2a. (above): the two vertices a and d are merged into a new vertex a+d, 2b. (below): the vertex a is split into two vertices a1 and a2.

1. Fix a proportion of vertices to affect in {1% ; 2% ; 5% ; 10% ; 20% ; 40%}.
2. For each of the 1,000 scenarios:
   a. use a random draw to select the vertices to apply the operations to,
   b. for each of these vertices, apply the following:
      - if Experiment 2a is performed (merge), chose another vertex at random and replace these two vertices with a new one,
      - if Experiment 2b is performed (split), replace the vertex with two new vertices,
   c. simplify the resulting shocked graph,
   d. compute the four centralities on the simplified graph.

Output: the values of the four centrality metrics for each vertex of the unshocked graph (except for those which disappear in one of the simulations) in each of the 1,000 scenarios, for each affected vertices proportion.

**Tab. 3** Successive steps and output of Experiment 2

the edges of $a + d$ are those of $a$ and $d$ at the same time. Second, we divide vertices into two new vertices, to replicate the inaccuracy which arises when considering that two anthroponyms designate a single individual, whereas they are actually two different individuals (Experiment 2b). The two vertices resulting from this division split the edges of the original vertex equally. Figure 6b gives an example of such a division: vertex *a* is divided into one vertex *a1* and one vertex *a2,* each inheriting half of *a*'s edges.

After fixing the proportion to impact, we select the vertices to which the two operations we have just described will be applied by a random draw. We then subject each of these vertices to one of the two merge or split operations described above. The graph thus shocked is then simplified, and the centrality metrics are calculated. Table 3 lists the successive steps of Experiment 2.

*Experiment 3*

In Experiment 3, we consider the case where the quantities that are used as weights for the edges are subject to inaccuracies, such as reading or transcription errors. For this, we study the Ypres network and consider the loan amounts as edge weights, as described in Section 2.2. In order to stick as closely as possible to the situation that arises in a historical network analysis, we try to replicate as closely as possible the errors that occur when reading quantities in the sources. As these amounts are expressed in Roman numerals in the Ypres chirographs, this is the form in which we handle them. Two errors are simulated here, in what we will call experiments 3a and 3b. The first one is that of the transcriber who substitutes one letter for another, and transforms for example the number "xxi" into the number "xvi" (Experiment 3a). To do this, the computer randomly chooses one of the letters making up the amount to be processed, and replaces it with another randomly chosen letter from {i ; v ; x ; l ; c ; d ; m}. When this operation produces

a string that does not match the writing of a number in Roman numerals, another random substitution is performed instead, and so on until this is the case. The second error we replicate is that of the transcriber who forgets to copy one of the letters of the number, and transforms, for example, the number "xxi" into the number "xx" (Experiment 3b). Here, the computer randomly chooses a letter to delete, multiplying the attempts if necessary, as for the substitution error.

Since each recognizance of debt is, of course, associated with only one amount (the amount of the loan in question), it is essential to ensure that the weights of the edges of the graph that correspond to the same source undergo the same operations. To do this, we start again from the sources used to construct the graph: we first select by random draw a subset of these sources (e.g. 10% of them). We then subject each corresponding amount to one of the two substitution or deletion operations described above. The weights of all the edges in the graph that correspond to the source in question are then modified. The shocked graph is simplified, and the centrality metrics are calculated. Table 4 lists the successive steps of Experiment 3.

1. Fix one type of error to apply (substitution or deletion), and a proportion of sources to affect in {1% ; 2% ; 5% ; 10% ; 20% ; 40%}.
2. For each of the 1,000 scenarios:
   a. use a random draw to select the sources to apply the errors to,
   b. for each of these sources, apply the error to the amount:
      - if Experiment 3a is performed (substitution), replace one letter of the number in its Roman numeral form with another letter,
      - if Experiment 3b is performed (deletion), delete one letter of the number in its Roman numeral form,
   c. apply these changes to the amounts of all the edges whose source attribute is among the selected sources,
   d. simplify the resulting shocked graph,
   e. compute the four centralities on the simplified graph.

Output: the values of the four centrality metrics for each vertex of the unshocked graph in each of the 1,000 scenarios, for each affected sources proportion.

**Tab. 4** Successive steps and output of Experiment 3

*Experiment 4*

In Experiment 4, we test the stability of the metrics against dating errors. To do so, we restrict ourselves to the hagiographic graph, as it is constructed on the basis of sources extending over a large chronological period: 8 centuries, compared to 12 years for the Cambrai graph and 53 years for the Ypres graph. To this end, we manipulate a quantity that does not correspond to the centrality metrics themselves, but a measure of their variation over time, which we must present before describing the experiment.

When we consider that the edges of this network have an attribute *century,* it can be seen as a dynamic graph, which evolves over time. Here we are interested in how measures of centrality change over the centuries, for example to find out which vertices alter their importance greatly over time. This information is valuable in the context of the historical analysis of sanctorals: it is a question of determining which saints have experienced the greatest shifts in popularity over the centuries (at least in terms of co-tradition).[37] We therefore consider the values of these metrics for the subgraphs made up of the edges corresponding to each of the centuries. For a given vertex and a given measure of centrality, we thus obtain the set of values that the metric takes successively, century after century. This way of producing a time series from a dynamic graph by cutting it into a succession of "snapshot photos",[38] each of which correspond to a static graph, is well known in the literature.[39]

The focus here is on the variation of the metrics over time: we therefore calculate the total variation of the obtained time series divided by their length, i.e., the average of the absolute values of their differences. Let us take a simple example: suppose that the degree centrality of a saint passes, between the 10th and 13th centuries, through the values 10, 15, 40, 30. In this case, the total variation is equal to $(|15 - 10| + |40 - 15| + |30 - 40|)/4 = 10$.

Let us now present the design of Experiment 4. Our goal is to simulate source dating errors. To do this, we randomly select a set of the manuscripts used as a basis for the construction of the hagiographic graph, and modify the century associated with them. To make the errors plausible, we do not make this transformation randomly: each of the drawn manuscripts is ante-dated or post-dated by only one or two centuries. Indeed, for several reasons of historical, paleographic, philological, etc. nature, it is rare for a historian to make a huge dating error. An 8th century manuscript is thus very unlikely to be mistaken for a 15th century

37 de Valeriola/Dubuisson (2021).
38 Lemercier (2015), p. 186.
39 See, among many examples, Uddin/Piraveenan/Chung/Hossain (2013).

1. Fix a shift direction (backward or forward shift) and a proportion of manuscripts to affect in {1% ; 2% ; 5% ; 10% ; 20% ; 40%}.
2. For each of the 1,000 scenarios:
   a. use a random draw to select the manuscripts to apply the operations to, then:
      - if Experiment 4a is performed (backward shift), subtract one century from the *century* attribute of 2/3 of selected manuscripts, and two centuries from the other 1/3,
      - if Experiment 4b is performed (forward shift), add one century to the *century* attribute of 2/3 of selected manuscripts, and two centuries to the other 1/3,
   b. for each century:
      - create a subgraph containing all edges whose *century* attribute corresponds to the selected century,
      - simplify this subgraph,
      - compute the four centralities on the simplified subgraph,
   c. use the centralities values obtained for each century to compute their total variation.

Output: the values of the total variation of the four centrality metrics for each vertex of the unshocked graph in each of the 1,000 scenarios, for each affected sources proportion.

**Tab. 5** Successive steps and output of Experiment 4

manuscript, and vice versa. These two time 'directions' (ante-dating and post-dating) lead to the two experiments 4a and 4b.

The first step in Experiment 4 is to set a proportion of manuscripts to which to apply the transformations. For each of the 1,000 stochastic scenarios, a list of such manuscripts is then drawn randomly, of which two-thirds see their century increase (in Experiment 4a) or decrease (in Experiment 4b) by one unit, and one-third by two units. The *century* attribute of the edges associated with these manuscripts are then transformed accordingly. The total variation of the centrality metrics is calculated as explained above, based on the division of the shocked graph into subgraphs corresponding to each century. These subgraphs are different in each of the stochastic scenarios, since the set of edges that undergo a transformation at the level of the attribute *century* is different each time. Table 5 lists the successive steps of Experiment 4.

### 3.3 Comparison Statistics

We have presented above how we simulate the incompleteness and imperfection of the sources, which result in the random generation of shocked graphs. We now need to explain how we compare these shocked graphs with the unshocked graphs. As we will see, we calculate different quantities for this purpose, which we will call 'comparison statistics' in the rest of this article.

Let's start with the first three experiments, which give similar outputs: the centrality score of each vertex in each of the 1,000 scenarios for each of the four metrics. In this case, we compute four different statistics. First, we estimate the

'deformation' that each network as a whole undergoes. To do this, the global measure of centralization is calculated in each scenario. We then take the average of these results for each network, and normalize it with the value of the centralization of the unshocked graph (which is shown in Figure 2). The same quantity calculated for the unshocked network is, due to this division, equal to 1. This normalized average of centralizations is the first of the four comparison statistics used in the *Results* section below for experiments 1, 2 and 3.

Second, we question whether or not individual centrality scores change significantly. To test this aspect, we calculate the correlation coefficient between centrality scores of the unshocked network vertices and those of each of the 1,000 shocked networks vertices.[40] When the vertices of the unshocked network do not exactly match the vertices of the shocked network, we perform the calculation only for the vertices they have in common. Among the three main statistical correlation indicators, we have chosen Spearman's coefficient, which we believe is the most appropriate.[41] Once these 1,000 correlations are obtained, their average is calculated. The same quantity calculated for the unshocked network is equal to 1, since it then corresponds to the correlation of the series of initial centrality scores with itself. This average of correlations is the second of the four statistics used in the *Results* section below for experiments 1, 2 and 3.

Third, we look more closely at the individuals who can be considered the most important within the network. Indeed, it is often to these individuals that the historian's eye turns when using these metrics. It is therefore natural to wonder whether this list of vertices is stable when the graph is subjected to the perturbations of experiments 1 to 3. To estimate this stability, we calculate, for each of the 1,000 scenarios, the proportion of individuals who are in the top 10% of the unshocked network in terms of centrality (or, equivalently, whose centrality score is greater than or equal to the 10%-quantile of all centrality scores) and who are still in this top 10% in the shocked network.[42] It is therefore the intersection (or overlap) between the top 10% of the two graphs. Once these 1,000 proportions are calculated, we take the average. The same quantity calculated for the unshocked

40 This statistic has already been computed in the literature, see e.g. Borgatti/Carley/Krackhardt (2006); Platig/Ott/Girvan (2013).

41 It seems pertinent, in the context of a historical analysis, to work on the basis of the ranks of the vertex centrality scores, and not on the scores themselves. The ranking of individuals in the network in increasing order of centrality is indeed one of the main interests of the historian. This is why we have not chosen Pearson's correlation coefficient. We chose Spearman rather than Kendall because the latter makes an estimate relatively close to the third comparison statistic, the overlap (see below). It should be noted, however, that all three coefficients were calculated in our exploratory analyses, and the conclusions obtained did not differ significantly.

42 Several authors compute this statistic, see Borgatti/Carley/Krackhardt (2006); Tsugawa/Ohsaki (2015).

network is equal to 1, since in this trivial case we calculate the intersection of a set with itself. This average of proportions is the third of the four statistics used in the *Results* section below for experiments 1, 2 and 3.

Fourth, we are interested in the relationship of the four measures of centrality to each other. We wonder about the impact of the condition of sources on the extent to which the metrics converge or diverge. In each of the 1,000 scenarios, we calculate the Spearman correlation coefficient (for the same reasons as above) between the 6 pairs of centrality measures.[43] We then compare each of these results with the correlation between the same two metrics in the unshocked graph (which is shown in Figure 3). This matrix of correlation coefficients is the last of the four statistics used in the *Results* section below for experiments 1, 2 and 3.

Let us move on to the fourth experiment, which must be considered separately because its output is not the same as that of the other three. Our goal here is to see if the evolution of vertex centrality metrics undergoes large changes when taking into account dating errors. Once the 1,000 scenarios are generated, we obtain a set of 1,000 total variations for each metric of each vertex. For each scenario, we then calculate the correlation coefficient (Spearman) between the statistical series composed of the total variations of the shocked network and that of the total variations of the unshocked network. Finally, we take the average (over the scenarios) of the correlations obtained. The same quantity calculated for the unshocked network is equal to 1, since it then corresponds to the correlation of the series of initial total variations with itself. This average of the correlations is used in the *Results* section below for Experiment 4.

### 3.4 Calculability and Implementation

All our calculations were performed in the R scripting language, using the igraph library.[44]

A remark about the number of scenarios to be generated is worth mentioning here. We chose to set this number at the highest possible value, because of the large variation in the structure of the shocked graph obtained from one scenario to another.[45] Indeed, we affect a fixed number of sources in each scenario, but not a fixed number of edges or vertices, so that the shocked graphs that are constructed potentially present quite disparate profiles. As we have seen in Section 3.3, most

43 The effect of shocks on the correlation between centrality metrics has been studied in e.g. Borgatti/Carley/Krackhardt (2006); Tsugawa/Ohsaki (2015).

44 R Core Team (2020); Csardi/Nepusz (2006).

45 Compared with Bolland (1988), p. 241 (100 scenarios), Platig/Ott/Girvan (2013), p. 5 (500 scenarios) or, in a slightly different context, Borgatti/Carley/Krackhardt (2006), p. 126 (10,000 scenarios).

methods of comparing the values of the centrality metrics of randomly truncated graphs with those of the unshocked graphs are based on averages over all scenarios. By choosing a large number of random scenarios, we ensure that these means are 'stable' (or, in statistical terms, we control the variance of the estimator).

Unfortunately, it is not possible to indefinitely increase the number of scenarios to be simulated. Setting the number of scenarios means making a compromise between the stability of the results obtained and the time the computer needs to perform the calculations. In the case of the Ypres graph for example, the computer calculates 4 measures of centrality for the 4,675 vertices, in graphs generated on the basis of 6 proportions of erasure in 1,000 scenarios each time. No less than $4 \times 4{,}675 \times 6 \times 1{,}000 = 112{,}200{,}000$ calculations are made. It should also be noted that the calculation of a metric value sometimes involves heavy calculations: for closeness centrality, for example, the score of a single vertex is obtained by calculating the distance of this vertex from all the other vertices of the graph. In order to perform this vertiginous number of calculations, we have parallelized the procedure on an eight-core intel i7-7700HQ @ 2.80 GHz processor. More than eight hours of calculations were required to obtain the results of Experiment 1 for the three networks. This duration is multiplied by 10 in the case of Experiment 2, where the graph transformation operations (merging and splitting vertices) themselves take a lot of time.

## 4. Results and Discussion

This section is dedicated to the presentation and discussion of our results. We will first consider them raw, then at increasingly summarized levels.

### 4.1 Raw Results

We will present separately the results concerning the correlation between metrics and those concerning the other comparison statistics, since the output of the computation is different in these two cases: one value per pair of metrics on the one hand, one value per metric on the other. Figure 7 gives the results of the first statistic for two of the three networks.[46]

These plots show curves with fairly gentle slopes, with total value changes (differences between the correlation for the unshocked graph and for a proportion of affected sources equal to 40%) lying between –0.24 and 0.16. The correlation coefficients increase (in 30/66 curves) or decrease (36/66 curves) with the propor-

46 We omit the results for the third network here, for reasons of space. They are very similar to the two presented here.

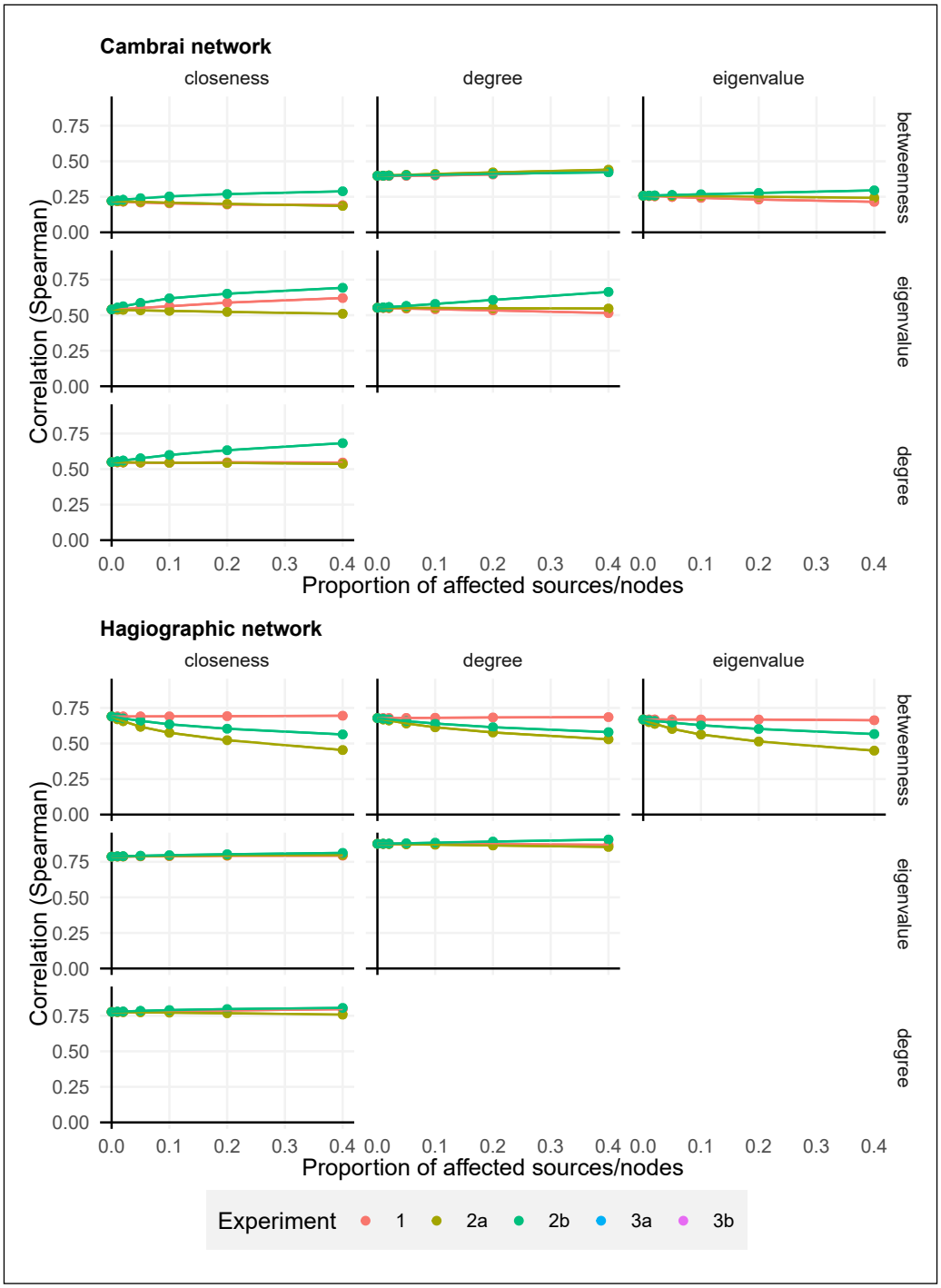


**Fig. 7** Results in terms of correlation between metrics for Cambrai and Hagiographic networks

tion of affected sources depending on the experiments and networks, without any trend seeming to emerge.[47]

This stability indicates that the impact of experiments 1 to 3 on the correlation coefficients between metrics is very limited. We can therefore conclude that the vagaries of the sources do not significantly distort the relationships that our metrics have with each other. The results concerning correlation between metrics do not need to be further summarized.

Let's now look at the four other comparison statistics: centralization, correlation, overlap and total variation. For each experiment, for each proportion of affected sources, for each network, for each centrality metric and for each statistic, we obtain a number to compare with the constant value 1, which corresponds to the unshocked network. These results can be visualized as a matrix of plots, as shown in Figure 8 for Experiment 1.

If we omit the numerical results and the hierarchies it presents (which will be summarized more effectively below), the main interest of this plot lies in the general look of the represented curves.

First, it should be noted that these curves do not show very 'violent' or 'rugged' shapes. This means, on the one hand, that the impact of source removal on centrality metrics depends smoothly (if not continuously) on the proportion of removed sources, an obvious conclusion that has the merit of reassuring us of the quality of the design and implementation of Experiment 1. On the other hand, from the regularity of these curves, we see that the number of scenarios we consider is high enough to extract the trend of the observed phenomena, i.e. to obtain stable results.

Second, it is interesting to note that although the curves are all decreasing, they show differences in terms of concavity. The curves giving the results for centralization (first line of plots) are concave: their slope is more and more negative. On the contrary, those of the overlap (the third line of plots) are convex: their slope is less and less negative. This distinction is due to a major difference between these two comparative statistics. The overlap statistic considers only a small part of the vertices of the graph, and is therefore potentially (remember that we generate a large number of stochastic scenarios) strongly affected by the removal of a small proportion of the sources. When this proportion increases, this effect is gradually mitigated, as if the most important part of the damage was done with the first deletions. On the contrary, the centralization statistic is a met-

47 This contradicts Bolland's observation that the correlation between metrics tends to increase (up to a 90% limit) when edges of the unshocked graph are removed (Bolland (1988), p. 251).

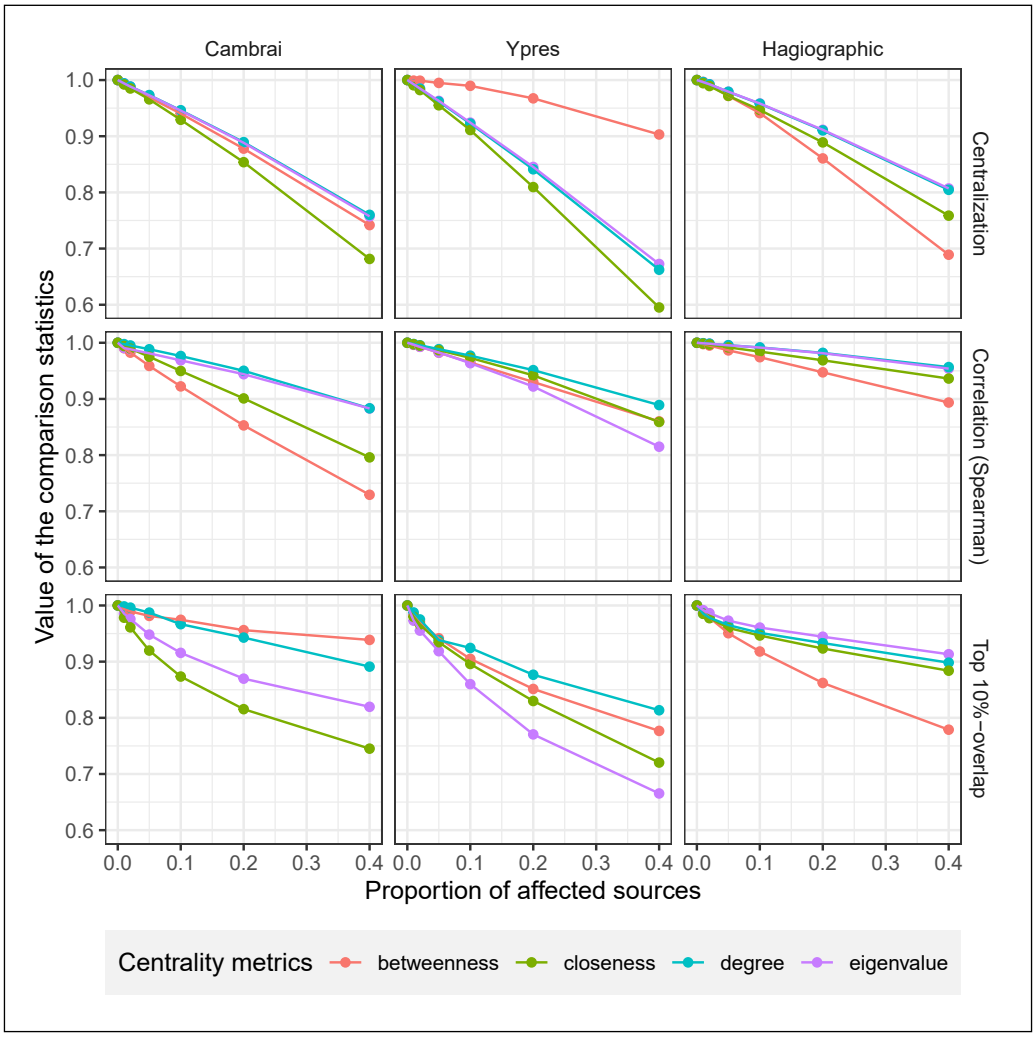


**Fig. 8** Results in terms of centralization, correlation and overlap for Experiment 1

ric that takes into account all the vertices and describes the general internal organization of the graph. Removing a small proportion of the sources therefore has only a limited effect, which is compounded when the proportion of removals increases and the overall structure of the network changes.

Third, note that there is no significant crossover between these curves. The hierarchy between the metrics that their relative positioning indicates is therefore always about the same, regardless of the proportion of sources removed. For this reason, and since similar conclusions can be drawn from similar plots for the other experiments (which we do not present here for reasons of space), we can now stick to the extreme values of these curves, i.e. the values of the comparison

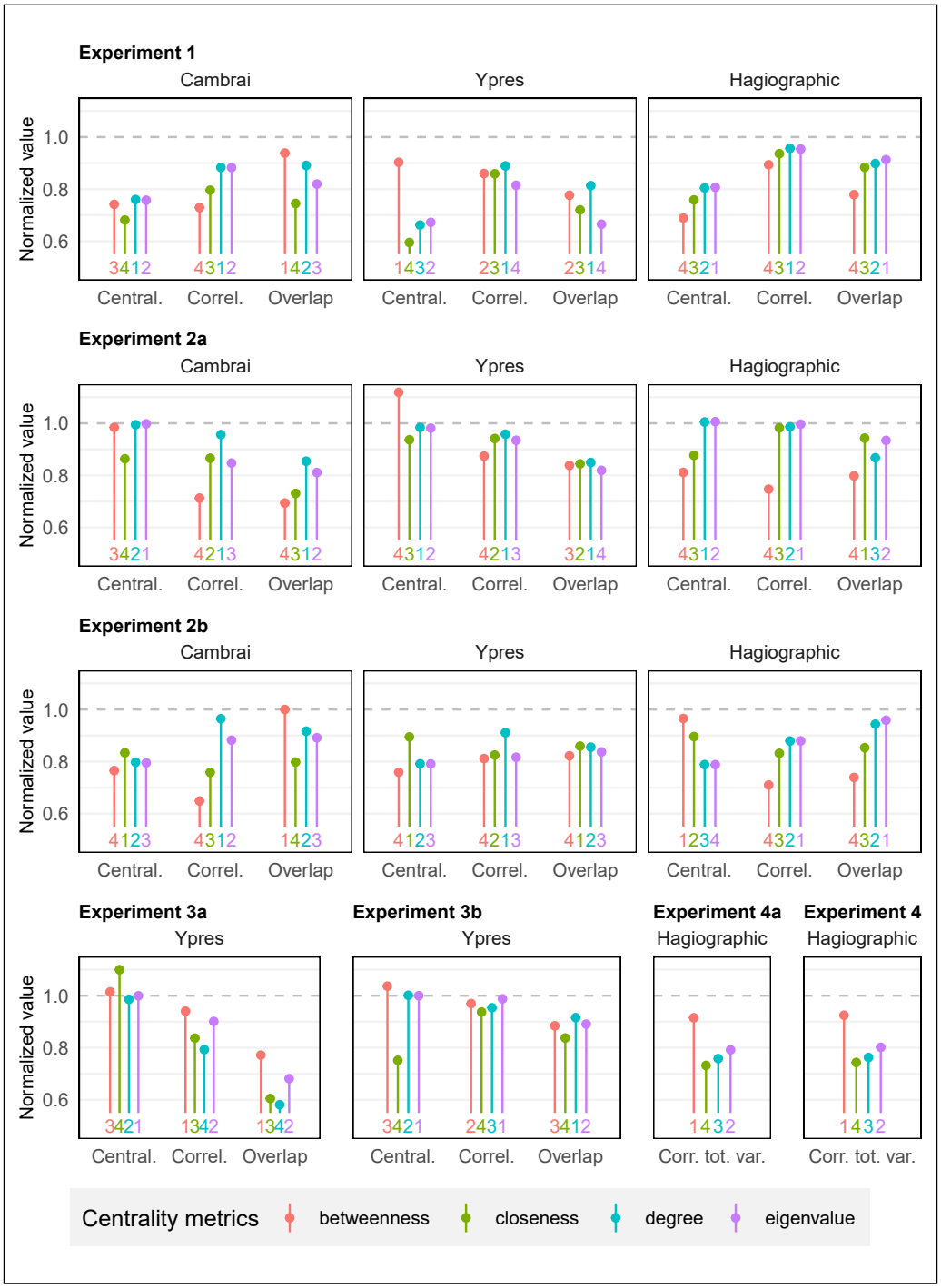


**Fig. 9** Abridged results of all experiments (corresponding to a proportion of affected sources/nodes equal to 40%)

statistics when the proportion of affected sources (or nodes, in the case of Experiment 2) is 40%.

### 4.2 First Level of Summarization

This extreme value is the quantity represented on the plots in Figure 9, which gives an overview of the results obtained for all experiments.

These lollipop plots therefore each represent the worst value (i.e. the furthest from 1) obtained from our robustness tests. For example, the very first plot in this figure (Experiment 1, Cambrai) shows the four levels of the centralization statistic that correspond to the rightmost points of the curves shown in the plot at the top left of Figure 8 (values 0.76, 0.76, 0.74 and 0.68). Below each lollipop is written the rank of each of the centrality metrics, calculated for that particular experiment, graph and statistic, on the basis of their distance from 1.[48] We can thus read in the very first plot that, for the Cambrai graph in Experiment 1 in terms of the centralization statistic, degree centrality is the most robust metric, followed by eigenvalue centrality, then betweenness centrality and finally closeness centrality. A number of observations can be made on this set of plots, before moving on to the next level of summarization of results.

First, let's look at the values obtained as a whole. None of them goes very low, since the smallest is 0.58 (Experiment 3a, Ypres, overlap, degree centrality), and more than two thirds are higher than 0.8, despite the fact that we consider a significant proportion of affected sources/nodes (40%). This observation, combined with the fact that the curves in Figure 8 have a smooth shape, makes it possible to answer the first question asked in Section 1.3. Yes, historians can trust centrality metrics, at least to some extent. The hazards experienced by the sources (at least the replications of these hazards that we have simulated) should have a rather minor impact on the values of these metrics, and therefore do not completely destroy the conclusions that the historian can draw from their use.

Second, note that most of the values obtained are smaller than 1, but not all. This of course occurs only for centralization, since the other two comparison statistics, correlation and overlap, cannot take values greater than 1. This means that the value of centralization is most often underestimated, but it is also possible that it is overestimated. The historian's caution must therefore apply in both directions.

48 For example, in the case of centralization in Experiment 2a for Ypres, the ordered metrics are: degree (value = 0.984, distance to 1 = 0.016), eigenvalue (value = 0.981, distance to 1 = 0.019), closeness (value = 0.937, distance to 1 = 0.063) and betweenness (value = 1.12, distance to 1 = 0.12).

Third, we observe on this set of plots a great diversity of results in terms of networks, comparison statistics and experiments. This brings nuances to the overall conclusion drawn in point 1 immediately above, which should not be misinterpreted. While the overall level of results is quite good, the disparity of these figures calls for caution and further analysis.

The values taken by some statistics in particular cases can be linked to the characteristics of the corresponding objects, and thus better understood. For example, the fact that betweenness centrality shows good performance in terms of overlap in the case of Cambrai for experiments 1 and 2b is rather intuitive. Indeed, as we noticed in Section 2.4, this metric takes very differentiated values for this network, with a large number of vertices having a centrality of 0, but a very small number of vertices with a huge centrality. The objective of the overlap statistic is to account for the movements observed in the top 10% vertices, among which are obviously those of the second category that we have just mentioned. While the number of geodesics passing through these few pivots of the graph is of course impacted by the operations carried out in experiments 1 (edge suppression) and 2b (vertex split), it is unlikely that they will take them out of the leading group. The opposite phenomenon can be observed, however, for Experiment 2a: betweenness centrality is in fourth place in terms of overlap for the Cambrai graph. The operation carried out in this experiment (vertices merging) has a much greater potential impact on this metric. Let's imagine that the computer merges two 'moderately important' vertices, each belonging to a different party. Before the merging, the vast majority of the shortest paths pass through the network hubs, the candidates for the episcopal see; after the merging, it is possible that a significant portion of the shortest paths will pass through the merged vertex and no longer through the hubs. In this eventuality, the vertices whose betweenness centrality is high can 'easily' lose their status.

It is possible to identify some more global trends within this diversity, as we will see below, but these are rather rare. We are far from a perfectly clear-cut situation, with uniform patterns repeating themselves from graph to graph or from statistic to statistic. A notable exception is the hagiographic network, where a hierarchy of metrics occurs almost every time for experiments 1 to 3. Similarly, careful observation of all the plotted values suggests that closeness centrality is quite often assigned a high rank, and eigenvalue centrality a low rank. It is nevertheless difficult, on the basis of Figure 9, to trust this sketch of hierarchy, or to discover other trends in these results. In order to untangle this somewhat chaotic situation, we need to move to higher levels of summarization of the results.

### 4.3 Second and Third Levels of Summarization

We now build two different second level summaries. We first calculate a score that aggregates all the results obtained for each of the metrics in the four experiments, represented in Figure 10.

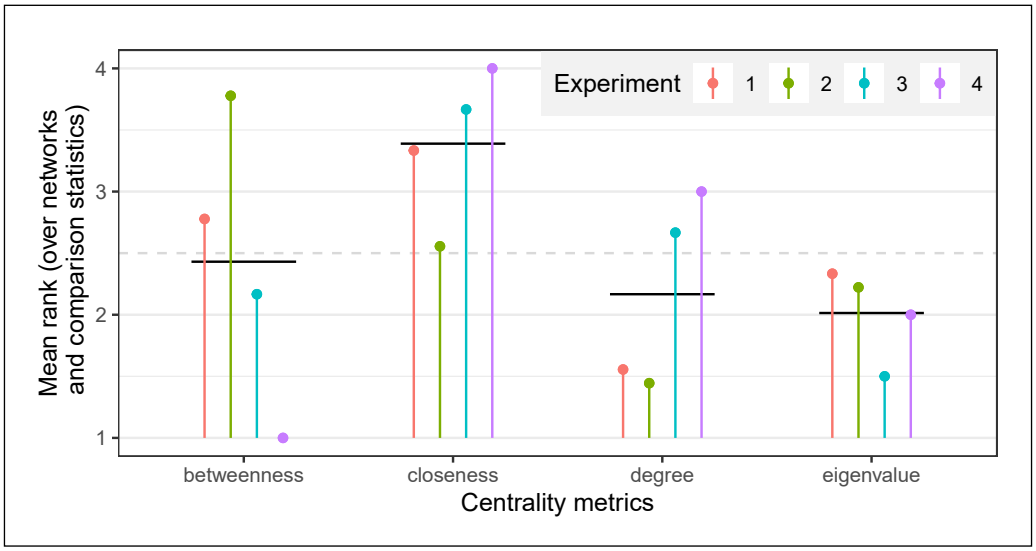


**Fig. 10** Mean ranks of metrics in each experiment (lollipops) and over experiments (black horizontal segments). The horizontal dotted grey line represents the total average rank (2.5)

To do this, we take the average of the ranks of the metrics over the graphs and over the comparison statistics.[49] For example, the average score of betweenness centrality in Experiment 1 (the height of the leftmost red lollipop) is obtained by averaging the 9 ranks indicated below the lollipops in the plots in the first row of Figure 9: $(3+4+1+1+2+2+4+4+4)/9 = 2.78$. In addition, the global average over the experiments is also calculated, which is represented on the plot by a black horizontal segment and corresponds to the third level summary.

We then calculate, in a similar way, a score that aggregates all the results obtained for each of the three graphs in the four experiments, represented in Figure 11.

As in the previous plot, a black horizontal segment gives the overall average over the experiments. This plot allows us to see which networks are, on average, most impacted by the experiments we subject them to. The hagiographic network takes first place on the podium, a conclusion consistent with the observations made above on the basis of Figure 9, and that is thus rather unsurprising. Indeed, as we saw in Section 2.4, this graph is much denser than the other two. Its high number of edges probably makes it less sensitive to the operations we per-

49 It seems far better to calculate the average of the ranks than the average of the results: what would it mean to add a correlation coefficient to a normalized centralization score?

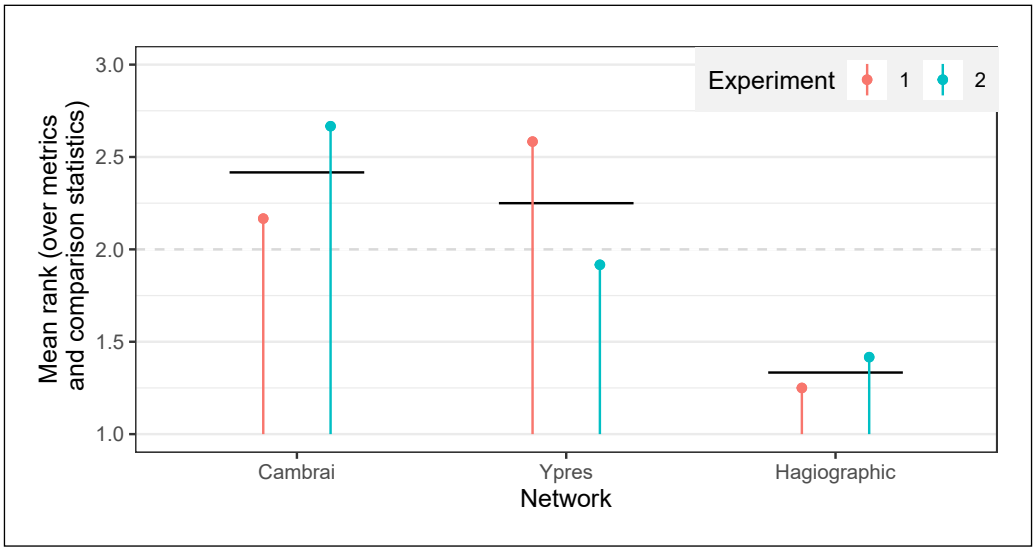


**Fig. 11** Mean ranks of networks in experiments 1 and 2 (lollipops) and over experiments 1 and 2 (black horizontal segments). The horizontal dotted grey line represents the total average rank (2)

form on it.[50] The next network in order of increasing impact is that of Ypres, but it is closely followed by that of Cambrai. It is interesting to note that both have opposite profiles in terms of experiments: the Ypres graph is less affected by Experiment 2 than by Experiment 1, while the opposite is true for the Cambrai graph.

Let us now return to Figure 10, which allows to draw up a hierarchy of centrality metrics and thus provide an answer to the question asked in Section 1.3. Eigenvalue centrality is the most robust under this criterion, followed closely by degree and betweenness centralities. We also observe that, if these three mean ranks are quite close to each other, the same cannot be said for the dispersion (over the experiments) of these values around their mean: the four values of the eigenvalue centrality are much less scattered than those of the degree centrality, which in turn are much less scattered than those of the betweenness centrality.[51] Betweenness centrality is rather extreme from this point of view, since it is the most robust for Experiment 4, but the least robust for Experiment 2. This observation about

50 The literature is divided on this relationship between robustness and density. Some authors observe an inverse phenomenon to ours, in which robustness decreases when density increases (Borgatti/Carley/Krackhardt (2006), p. 134; Galaskiewicz (1991)), while others show that the relationship depends on the metric of centrality considered (Frantz/Cataldo/Carley (2009), p. 319–321).

51 This observation is confirmed by the standard deviations of these three series of four numbers: 0.37 for eigenvalue centrality, 0.78 for degree centrality, and 1.16 for betweenness centrality.

dispersion allows us to be more confident about the very tight order given by the averages for the first three metrics. Indeed, a small dispersion value is an enviable quality, since it means a great uniformity over the experiments, i.e. that the stability of the metric is about the same for each of the experiments. Finally, closeness centrality is the least robust of our four metrics, with some consistency across the four experiments.[52]

Note that looking at the mean rank and the dispersion around this mean rank is not the only way to analyze these results. We can also decide to look at the maximum rank on the experiments (which corresponds to the worst performance of each metric). This other way of comparing the results reflects the idea that a metric is globally robust if it is robust in all situations, that is, in our case, regardless of the type of experiment we subject it to. The hierarchy obtained by comparing the maximum ranks is the same as that obtained by comparing the average ranks.

## 5. Conclusion

In this paper, we have assessed and compared the robustness of centrality metrics in the context of their use in historical network analysis. To do this, we have implemented a battery of tests that simulate the source defects that historians face, both in terms of incompleteness and inaccuracy. Our results show that, from a global point of view, centrality is a sufficiently stable quantity to be used in such a context. However, we have also shown that the hazards experienced by the sources have impacts that differ in magnitude depending on the network studied, the comparison statistic used, and the centrality metric considered. This finding calls for caution in the choice of network tools applied, and for nuance when interpreting the results of an analysis of historical networks using centrality.

We have found throughout the previous section that our conclusions are not always consistent with the literature. It should be recalled here that the experiments we conduct are not the same as those carried out in these studies, since they are directly inspired by the historian's practice. It is therefore legitimate that the results obtained are different. One can even go further, and argue that these divergences carry an interesting meaning. Indeed, they show that it is relevant to study the robustness of centrality metrics in our particular context, that of his-

52 The literature does not give a uniform picture of this hierarchy. For some authors, the four metrics are equally robust (Borgatti/Carley/Krackhardt (2006), p. 134; Tsugawa/Ohsaki (2015), p. 34), while for others the betweenness is the most unstable (Bolland (1988), p. 248–250; Costenbader/Valente (2003), p. 305). Eigenvalue is sometimes considered the most robust (*ibid.*), but sometimes it is closeness centrality that takes the top spot (Kim/Jeong (2007), p. 5). These results are therefore sometimes in adequacy, and sometimes in complete inadequacy with ours (see the commentary on this subject below).

torical analyses. We conclude that the 'one size fits all' approach does not work well, and that the particularities of the field of application need to be considered. This begs for a multiplication of methodological studies on the properties of historical networks.

We were also able to establish a hierarchy among the four centrality metrics considered, from the most to the least robust. We think it is important to take this ranking into account, since it is directly related to the degree of confidence one can have in the conclusions that centrality allows us to draw about a network. In particular, it seems necessary to avoid basing an analysis of centrality solely on the closeness metric, and not to focus on the results it provides when other metrics are used alongside it (and especially if they lead to different conclusions).

These recommendations may seem disappointing at first glance, since they weaken some of the tools historians have at their disposal to carry out their network analyses, thus making these analyses less efficient. However, this is what it's all about: in a sense, robustness is a matter of compromise. By using a robust tool, we trade efficiency for stability, something that is essential for historical studies. The very famous statistician Anscombe presents it as an insurance policy: you pay a premium (part of the efficiency of your process), and in exchange you get protection against accidents (i.e. process deviations).[53]

This type of compromise is, of course, not unfamiliar to historians, since it also occurs when the conclusions of a historical study are nuanced with respect to the sources from which they were drawn. We can even take this reflection further, going back to the reason that led us to analyze the robustness of centrality metrics. We wondered how confident we could be in what these tools teach us about the historical objects that these networks model. This function is nothing other than that of historical criticism, if we take, for example, Paul Veyne's definition.[54] It thus appears that considerations around robustness, an a priori purely statistical concept, are an integral part of this central tool of the historical method, the "common treasure of the corporation"[55] of historians.

The case of robustness is certainly not an isolated one, and many quantitative techniques, when applied to historical data, can play the role of criticism tools. Can we not therefore rethink the place of this set of methods within history as a discipline? According to Pirenne, the various specialized branches of history that he calls auxiliary sciences (epigraphy, diplomatics, numismatics, etc.) arose from the particularization of the process of criticism to particular objects and tech-

53 Anscombe (1960).
54 Veyne (1984), p. 12.
55 Stengers (2004), p. 103.

niques (related to inscriptions, charters, coins, etc.).[56] Following this idea, historical quantitative methods would therefore deserve to be included in the list of history sub-disciplines. Placing them in this way in a new framework would make it possible to reflect on fundamental questions that have so far only been touched upon, on the historical data themselves and their relationship to the quantitative. We can only hope that this time will come soon, and that new paths will then open wide.

## Bibliography

E. J. Anscombe (1960), Rejection of outliers, in Technometrics 2, p. 123–147 (DOI: 10.2307/1266540).

A. Bavelas (1948), A mathematical model for group structures, in Human Organization 7, p. 16–30 (DOI: 10.17730/humo.7.3.f403334485lgl053).

A. Bavelas (1950), Communication patterns in task oriented groups, in Journal of the Acoustical Society of America 22, p. 725–730 (DOI: 10.1121/1.1906679).

J. Bolland (1988), Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks, in Social Networks 10, p. 233–253 (DOI: 10.1016/0378-8733(88)90014-7).

P. Bonacich (1972), Factoring and weighing approaches to status scores and clique identification, in Journal of Mathematical Sociology 2, p. 113–120 (DOI: 10.1080/0022250X.1972.9989806).

S. P. Borgatti/K. M. Carley/D. Krackhardt (2006), On the Robustness of Centrality Measures Under Conditions of Imperfect Data, in Social Networks 28, p. 124–136 (DOI: 10.1016/j.socnet.2005.05.001).

J. Cellier/M. Cocaud (2012), Le traitement des données en histoire et sciences sociales. Méthodes et outils, Rennes.

E. Costenbader/T. W. Valente (2003), The stability of centrality measures when networks are sampled, in Social Networks 25, p. 283–307 (DOI: 10.1016/S0378-8733(03)00012-1).

G. Csardi,/T. Nepusz (2006), The igraph software package for complex network research, in InterJournal (Complex Systems) 1695, p. 1–9.

K. Das/S. Samanta/M. Pal (2018), Study on centrality measures in social networks: a survey, in Social Network Analysis and Mining 8, (DOI: 10.1007/s13278-018-0493-2).

S. de Valeriola (2021), L'ordinateur au service du dépouillement de sources historiques. Éléments d'analyse semi-automatique d'un corpus diplomatique homogène, in Histoire et mesure 35, p. 171–196 (DOI: 10.4000/histoiremesure.13534).

56 Pirenne (1933), p. 438.

S. de Valeriola (2019), Le corpus des chirographes yprois, témoin essentiel d'un réseau de crédit du xiii[e] siècle, in Bulletin de la Commission royale d'histoire 185, p. 5–74 (DOI: 10.3406/bcrh.2019.4382).
S. de Valeriola/B. Dubuisson (2021), L'hagiographie à l'aune du numérique: nouvelles perspectives d'étude des légendiers latins, forthcoming.
S. de Valeriola/N. Ruffini-Ronzani/É. Cuvelier (2021), Dealing with the Heterogeneity of Interpersonal Relationships in the Middle Ages. A Multi-Layer Network Approach, forthcoming in Digital Medievalist.
M. Düring (2016), How Reliable Are Centrality Measures for Data Collected from Fragmentary and Heterogeneous Historical Sources? A Case Study, in T. Brughmans/A. Collar/F. Coward (eds.), The Connected Past: Challenges to Network Studies in Archaeology and History, Oxford, p. 85–101 (DOI: 10.1093/oso/9780198748519.003.0011).
A. Eagle (2005), Randomness Is Unpredictability, in The British Journal for the Philosophy of Science 56, p. 749–790 (DOI: 10.1093/bjps/axi138).
T. Frantz/M. Cataldo/K. Carley (2009), Robustness of centrality measures under uncertainty: Examining the role of network topology, in Computational and Mathematical Organization Theory 15, p. 303–328 (DOI: 10.1007/s10588-009-9063-5).
L. C. Freeman (1979), Centrality in Social Networks. Conceptual Clarification, in Social Networks 1, p. 215–239 (DOI: 10.1016/0378-8733(78)90021-7).
J. Galaskiewicz (1991), Estimating point centrality using different network sampling techniques, in Social Networks 13, p. 347–386 (DOI: 10.1016/0378-8733(91)90002-B).
M. Groenhuijzen/P. V. Verhagen (2016), Testing the Robustness of Local Network Metrics in Research on Archeological Local Transport Networks, in Frontiers in Digital Humanities 3, p. 1–14 (DOI: 10.3389/fdigh.2016.00006).
M. Hammond (2017), Social Network Analysis and the People of Medieval Scotland 1093–1286 (PoMS) Database, Glasgow, https://www.poms.ac.uk/information/e-books/social-network-analysis-and-the-people-of-medieval-scotland-1093-1286-poms-database/ (accessed 15/08/2020).
P. J. Huber (2009), Robust statistics, Hoboken.
M. Jalili/A. Salehzadeh-Yazdi/Y. Asgari/S. S. Arab/M. Yaghmaie/A. Ghavamzadeh/K. Alimoghaddam (2015), CentiServer: A Comprehensive Resource, Web-Based Application and R Package for Centrality Analysis, in PLOS One 10, 2015 (DOI: 10.1371/journal.pone.0143111).
P.-J. Kim/H. Jeong (2007), Reliability of rank order in sampled networks, in The European Physical Journal B 55, p. 109–114 (DOI: 10.1140/epjb/e2007-00033-7).
C. Kiss/M. Bichler (2008), Identification of influencers – measuring influence in customer networks, in Decision Support Systems 46, p. 233–253 (DOI: 10.1016/j.dss.2008.06.007).
H. Kyburg (1974), Randomness, in The Logical Foundations of Statistical Inference, Dordrecht, p. 216–246.

A. Landherr/B. Friedl/J. Heidemann (2010), A Critical Review of Centrality Measures in Social Networks, in Business & Information Systems Engineering 2, p. 371–385 (DOI: 10.1007/s12599-010-0127-3).
C. Lemercier (2015), Formal network methods in history: why and how?, in G. Fertig, Social Networks, Political Institutions, and Rural Societies, Turnhout, p. 281–310 (DOI: 10.1484/M.RURHE-EB.4.00198).
C. Lemercier (2015), Taking time seriously. How do we deal with change in historical networks?, in M. Gamper/L. Reschke/M. Düring (eds.), Knoten und Kanten III. Soziale Netzwerkanalyse in Geschichts- und Politikforschung, Bielefeld, p. 183–211 (DOI: 10.14361/9783839427422-006).
S. Oldham/B. Fulcher/L. Parkes/A. Arnatkevičiūtė/C. Suo/A. Fornito (2019), Consistency and differences between centrality measures across distinct classes of networks, in PLOS One 14, p. 1–23 (DOI: 10.1371/journal.pone.0220061).
H. Pirenne (1933), What are historians trying to do?, in S. A. Rice (ed.), Methods in social science. A case book, Chicago, p. 435–445.
J. Platig/E. Ott/M. Girvan (2013), Robustness of network measures to link errors, in Physical Review E 88, 1–9 (DOI: 10.1103/physreve.88.062812).
R Core Team (2020), R: A Language and Environment for Statistical Computing, https://www.R-project.org/.
G. F. Riva (2019), Network Analysis of Medieval Manuscript Transmission, in Journal of Historical Network Research 3, p. 30–49 (DOI: 10.25517/jhnr.v3i1.61).
I. Rosé (2011), Reconstitution, représentation graphique et analyse des réseaux de pouvoir au haut Moyen Âge. Approche des pratiques sociales de l'aristocratie à partir de l'exemple d'Odon de Cluny († 942), in Redes. Revista hispana para el analisis de redes sociales 21, p. 199–272 (DOI: 10.5565/rev/redes.420).
N. Ruffini-Ronzani (2020), Relire l'histoire des principautés territoriales à travers l'analyse de réseaux: Les cas du Hainaut et de Cambrai (xi$^{e}$–xii$^{e}$ siècles), forthcoming in M. Margue/H. Pettiau (eds.), Principes et episcopi. Dynamiques du pouvoir princier en Lotharingie (seconde moitié du x$^{e}$-première moitié du xii$^{e}$ siècle).
J. P. Scott (2000), Social Network Analysis: A Handbook, London.
J. Stengers (2004), Unité ou diversité de la critique historique, in Revue belge de philologie et d'histoire 82, p. 103–122 (DOI: 10.3406/rbph.2004.4815).
M. Trigalet (2001), Compter les livres hagiographiques. Aspects quantitatifs de la création et de la diffusion de la littérature hagiographique latine (II–XVe siècle), in Gazette du livre médiéval 38, p. 1–13 (DOI: 10.3406/galim.2001.1513).
S. Tsugawa/H. Ohsaki (2015), Analysis of the Robustness of Degree Centrality against Random Errors in Graphs, in G. Mangioni/F. Simini/S. Uzzo/D. Wang (eds.), Complex Networks VI, Cham (DOI: 10.1007/978-3-319-16112-9_3).

S. Uddin/M. Piraveenan/K. S. K. Chung/L. Hossain (2013), Topological analysis of longitudinal networks, in 2013 46th Hawaii International Conference on System Sciences, p. 3931–3940 (DOI: 10.1109/HICSS.2013.556).
T. W. Valente/K. Coronges/C. Lakon/E. Costenbader (2008), How Correlated Are Network Centrality Measures?, in Connections 28, p. 16–26.
P. Veyne (1984), Writing History. Essay on Epistemology, Middletown.
S. Wasserman/K. Faust (1995), Social Network Analysis: Methods and Applications, Cambridge.
C. Wyffels (1991), Analyses de reconnaissances de dettes passées devant les échevins d'Ypres, 1249–1291, Brussels.