Understanding Information Centrality Metric: A Simulation Approach

Chintan Amrit¹ and Joanne ter Maat²,

¹Amsterdam Business School, University of Amsterdam, The Netherlands
²Topics, Enschede, The Netherlands

Corresponding author: Chintan Amrit (e-mail: c.amrit@uva.nl).

ABSTRACT Identifying the central people in information flow networks is essential to understanding how people communicate and coordinate as well as who controls the information flows in the network. However, the appropriate usage of centrality metrics depends on an understanding of the type of network flow. Networks can vary in the way node-to-node transmission takes place, or in the way a course through the network is taken, thereby leading to different types of information flow processes. When metrics are used for an inappropriate flow process, the result of the metric can be misleading and often incorrect. In this paper we create a simulation of the flow of information in a network, and then we investigate the relation of information centrality as well as other network centralities, like betweenness, closeness and eigenvector along with the outcome of simulations with information flowing through walks rather than paths, trails or geodesics. We find that Information Centrality is more similar to Eigenvector and Degree centrality than to Closeness centrality as postulated by previous literature. We also find an interesting pattern emerge from the inter metric correlations.

INDEX TERMS Social networks, Information centrality, Information flow processes, Network flow simulation.

I. INTRODUCTION

Social networks and social network metrics are used to analyse the behaviour of actors in various areas such as intelligence [2], mobile ad hoc networks [3] and graph robustness [4]. There are different kinds of flow processes that can occur within a social network [5], and in this paper we concentrate on information flow [6]. Examples of these processes are the way gossip spreads or in the way information is communicated by using e-mail. When central nodes in such networks are located, one can aim their marketing strategies at these nodes to make sure it will reach the greatest number of people. Alternatively, one could conduct surveys on the most central persons in a network, since they are most likely to have the information that one needs. Although one could use centrality metrics to locate these nodes, not every metric is suitable for every kind of network [5]. Betweenness centrality, for example, is a metric that considers only the shortest paths, or geodesics [7]. So, if the betweenness metric is used in a network in which information can flow through paths other than the shortest, it would not provide the right information [5]. Borgatti [5] focuses on flow processes in general, and analyse their fit with a few metrics; namely, the Freeman’s degree, closeness, betweenness and Bonacich’s eigenvector centrality. He however does not consider the information centrality metric [8] in much detail. Stephenson and Zelen’s information centrality metric [8] has been regarded as important to study information communication flows, especially when it is known to not take the shortest path[9-12]. Brandes and Fleischer [13] suggest that information centrality, though highly recognized as an important centrality metric, is not often used by network scientists as its foundations are unclear [13]. In particular, it is unclear what kind of network flows information centrality is most suitable for. In this paper, we demonstrate how one can find the right metrics for the different kinds of information flow processes. In particular, we try to determine the appropriate network flow for applying the information centrality [8] metric. We analyse the correlations among the network metrics and the results of the flow simulations in multiple popular social networks. While previous work have analysed network metric correlations in different network configurations [14, 15], here we analyse the metric correlations with the outcome of flow simulations. We also include the information centrality [8], that earlier papers have largely ignored. We find that Information Centrality is similar to Eigenvector and Degree centrality and quite different from Closeness centrality. Therefore our finding is different from what is postulated by previous research [5].

The rest of the paper is organized as follows; first, we explain the background of the research, after which we describe the literature on information centrality. Next, we conduct a simulation in order to compare the actual values of the information centrality metric with most central places in a
network as indicated by the simulations. We then discuss the results of the simulations and then the implications of our result.

II. Research Background

Borgatti [5] arrived at a typology containing two dimensions along which different flow processes in a social network can vary (Table 2). One dimension considers the node-to-node transmission and consists of broadcasting, or parallel duplication, serial duplication, and transfer. Both parallel duplication and serial duplication copy the information before sending it on. With serial duplication a node sends it to only one peer, while in parallel duplication a node sends the information to all its peers. On the other hand, transfer does not copy the information; it just sends it to one of its peers. The other dimension (the rows in Table 2), considers the course that is taken through the network. This can be the shortest route, called a geodesic, or it can be a path, trail or walk. Paths do not repeat nodes and links, trails do not repeat links and in walks both nodes and links can be repeated. Figure 1 represents an example graph where these concepts can be elucidated.

Both geodesics and paths are not allowed to repeat any nodes or trails, so an information flow from a to b to c to f is a path and so is a → b → d → e → c → f. Only the information flow from a through b and c to f is called a geodesic, since this is the absolute shortest path from a to f. A trail is allowed to pass the same node multiple times, but not a link.

So, if the information flows from a to f again, it is a trial if, for example, after it has passed b, goes to c, e, d, c again and then gets to f. So, the node c is passed twice, but none of the links was repeated, making it a trail. In a walk there are no restrictions at all, the information is, for example, allowed to flow multiple times between b and d before continuing to c and f or can make loops between b and c and d.

Table 1 represents examples of these kinds of information flows in the real world. Gossip is an example of a trail with serial duplication, since it spreads from one person to another, but one usually does not tell the same gossip to a person again. However, it is possible that someone you have not spoken to about a particular rumour tells you the same gossip you have already heard from someone else. With word of mouth marketing on the other hand, the message needs to get through, so one might give the same message to a person again.

The centrality of a network is related to the cohesiveness of the graph it can be represented by [16]. It identifies the nodes that play a prominent role within the network, for example, the nodes that are navigated the most, or the nodes that can reach all other nodes in the shortest way [17]. Centrality can be measured in different ways.

![Figure 1: Example network](image)

Table 1: Examples of Information Flow Processes in Social Networks (Adapted from [5])

| Parallel Duplication | Serial Duplication | Transfer          |
|----------------------|--------------------|-------------------|
| **Geodesics**        |                    |                   |
| Group chat           | Individual e-mail  | Handing down a note through class |
|                      | One-one chat       |                   |
| **Paths**            |                    |                   |
| Internet name-server | Viral marketing    | Mooch             |
|                      |                    |                   |
| **Trails**           |                    |                   |
| E-mail broadcast     | Gossip             | Used Goods        |
|                      |                    |                   |
| **Walks**            |                    |                   |
| Attitude            | Word of mouth      | Money Exchange,   |
| Influencing,         | marketing          | Telecommunication information transfer |
| Wikipedia,           |                    |                   |
| Facebook,            |                    |                   |
| Twitter              |                    |                   |

The node that can reach all other nodes in the shortest way, so the node whose sum of links to all other nodes is the lowest, can be measured with the Freeman closeness metric [5]. Metrics such as Freeman degree [17], Freeman betweenness [17] and Bonacich eigenvector [18], measure whether a node plays a central role within the social network in one way or another. The Freeman degree measures the amount of links to a node. Betweenness counts the number of geodesic paths that lead through a node and the Bonacich eigenvector measures whether a node has neighbours who themselves score high on centrality [18]. Knowing the place of highest centrality in a social network gives the possibility to influence the information flow process in the network. Companies could use the information to find the node or nodes in the network that could best be used for marketing, for example.

The metrics for centrality, however, cannot be used with all types of social network. Whether a metric is applicable to a particular network depends on the two dimensions of Borgatti’s typology [5]. In his paper, Borgatti [5] matches the different kinds of metrics with the varying types of information flow processes within social networks. As a result, we have centrality metrics suitable for some of the social network types (Table 2). Table 2 represents the typology of information flow processes[5], and the classification of popular social network metrics in the Typology. It has to be noted that this Typology is independent of the type of network. In other words, Borgatti [5] suggests that certain metrics are appropriate for certain flow processes.
and this is completely independent of the underlying graph/network structure.

| Table 2 | Different Metrics for Information Flow Processes [5] |
|---------|------------------------------------------------------|
|         | Parallel Duplication | Serial Duplication | Transfer |
| Geodesics | Freeman closeness | Freeman closeness | Freeman betweenness |
| Paths    | Freeman closeness   | Freeman degree     |               |
| Trails   | Freeman closeness   | Freeman degree     |               |
| Walks    | Freeman closeness   | Freeman degree     | Bonacich eigenvector |

Parallel duplication is the process by which information at the node is copied and sent out to all connected nodes. Since every node that receives information thus makes a copy and then sends it to all its peers, information always reaches a target through the shortest path. The target can receive the information via multiple ways, but because of the nature of parallel duplication, the shortest path is always one of those ways. This makes the entry of geodesics under parallel duplication in table 2 obsolete. Closeness is a metric that works with shortest paths, so it is suitable for any network in which parallel duplication takes place. In any other network in which information flows through shortest paths or geodesics. Freeman degree refers to the number of links a node has to other nodes, or to how many peers a node has. That this metric is suitable for parallel duplication makes sense, because degree would be equal to the number of messages a node sends to its neighbors again when it has received and copied a message. Freeman betweenness is a metric that looks at the number of shortest paths that pass through a node; hence it is only applicable to geodesics. The Bonacich eigenvector does not consider the node itself, but the importance of the nodes connected to it. This means that the metric analyses the whole network and how the node is placed in it - whether its peers are central or not. High eigenvector centrality means that a node has peers that also have high eigenvector centrality. The calculation of the eigenvector uses a matrix in which the values stand for walks of a certain length between two nodes, so it counts walks of all lengths. Paths or trails thus, do not restrict the flow of information.

**A. Information Centrality**

Literature that describes metrics for centrality is used to identify metrics. Here we discuss some of the centrality metrics from literature and the role they play in understanding information flows.

Borgatti and Everett make a distinction between metrics that measure volume and those that measure length [16]. Another distinction they make is between metrics that start or end at a given node (radial metrics), or the amount of information that passes through a node (medial metrics). Brandes [19] on the other hand, discusses different variants on shortest path betweenness. They calculate metrics for paths in directed networks, where the source and target are also considered in the calculation. Other metrics pay special attention to the second and penultimate node in the paths, or the proxies [19]. Still other metrics are for networks in which the paths are of a bounded length, or where paths are weighted by the shortest possible path. Brandes (2008) also discusses a betweenness metric that takes into account the distance to the target, metrics that consider edges, or groups of nodes, instead of just a single node, and metrics that can handle networks that have different types of nodes [19]. Dolev, Elovici and Puzis [20] introduce routing betweenness centrality, a metric that can identify the most influential nodes in a network were multiple packets of information are send between a pair of nodes with these packets taking different routes. Tutzauer [21] comes up with entropy centrality to use in networks in which the transfer of information flows through paths. Newman [22] introduces random walk betweenness, a variation on betweenness centrality that assumes that information can flow through any path between a pair of nodes and that each time a random walk is taken. Brandes and Fleisher [13] also discuss this random walk betweenness by equating it to their current flow betweenness centrality. They also discuss current flow closeness centrality, and prove this is actually the same as Stephenson and Zelen’s [8] information centrality. Everett and Borgatti [23] show that a graph invariant, a property that depends only on the structure of the graph, can be used to calculate centrality. After the invariant is calculated, a node or edge is removed, after which the invariant is calculated again and the difference between the two indicates the centrality of the removed edge or node. This is called an induced metric. This induced centrality consists of an endogenous and an exogenous part, so part of a node’s centrality is its own importance, but part is also how it contributes to the centrality of its neighbours. Communicability betweenness [24] tries to measure the betweenness of a node when all possible paths can be taken, instead of the betweenness when only shortest paths are taken like in the Freeman’s betweenness centrality[17].

Now let us look closer at the information centrality metric [8]. The reason we consider this metric is that it is used in many situations where the flow of information along the network is considered important. Some examples of research on networks in which information is important and the information centrality metric is used to find central nodes can be found in for example [9-12, 25]. This metric considers all paths between a pair of points and weights them relative to the information they contain. The metric does not consider only shortest or other type of path between a pair of nodes, but all...
the paths are considered for non-directional networks. The information centrality value of a node is an average of the information of all paths originating from that node. Since information is the reciprocal of the path length (reciprocal of variance, according to the theory of statistical estimation), this metric is related to closeness. So, the information in a path is the inverse of the length of a path. When there are two or more paths between a pair of nodes that contain some of the same incident nodes, the information is calculated with use of a matrix that contains the numbers of incident nodes the paths have in common and then is inverted. The information centrality for a node is given by:

\[
\bar{I}_i = \frac{n}{\sum_{j=1}^{n} \frac{1}{I_{ij}}} \tag{1}
\]

Here, \( n \) is the number of nodes and \( I_{ij} \) the centrality of a path from node \( i \) to \( j \). Brandes and Fleischer [13] prove that information centrality is the same as current flow closeness centrality given by:

\[
\frac{n_c}{\sum_{s \in S} \frac{p_{st}(s) - p_{st}(t)}{p_{st}(s)}} \forall s \in V \tag{2}
\]

Where the term \( p_{st}(s) - p_{st}(t) \) corresponds to the difference in potential between two points \( s \) and \( t \) in an electrical network. [9, 16]. However, though information centrality is found to be the same as current flow closeness centrality, it does not make it clear what type of networks the metric is apt for. However, Koschutzki et al. [26] describe the close relation between flow of current in a graph and random walks around a graph. They show that the potential \( p_v \) can be expressed as

\[
p_{st} = (D_v - A_v)^{-1}b_{st} \tag{3}
\]

Where \( D_v \) is the degree matrix and \( A_v \) the adjacency matrix with the rows and columns of a fixed vertex \( v \) removed, while \( b_{st} \) is the solution of the equation \( Lp_{st} = b_{st} \), \( L \) being the Laplacian matrix [20]. They further show that this equation above is similar to \( v_{st} \) given by

\[
v_{st} = (D_t - A_t)^{-1}s \tag{4}
\]

Where \( v_{st} \) is the probability of finding a message at vertex \( i \) while it is on a random walk from vertex \( s \) to vertex \( t \) [26] and \( s \) is an vector that is 1 at vertex \( s \) and 0 elsewhere. Substituting equation (4) in expression (2), we get that current flow closeness centrality is equivalent to a random walk centrality where one is interested in the normalized inverse of the expectation of finding a particle at vertex \( i \) (over any other neighbouring vertex \( j \)), while on a random walk from vertex \( s \) to vertex \( t \).

In the case of current flow betweenness centrality, it is clear that the metric is suitable for information flow through walks, as it is found to be equivalent to random walk betweenness centrality. Hence, given this link between current flow and random walk, we can hypothesize that information centrality is suitable for network flows through walks. We verify this assumption through conducting a simulation of a network.

In the next section, we describe the simulations we conducted for the different kinds of information flows. We are interested in seeing if the simulations point to the same central nodes as when calculating the information centrality [8] of the nodes.

### III Simulation

To verify for which types of social networks the metrics are suitable, we use NetLogo [27] to conduct multiple simulations. We wrote a simulation script in NetLogo to simulate a network and the information flow through it. By running the code in NetLogo, we generated a network, like the one in Figure 3. This is the network of marriage ties in Italian Renaissance families (in Florence) by Padgett and Ansell [1] (Figure 2), with one family that did not have any ties with other families omitted. We chose the Italian Renaissance families network [1] in order to carry on the research and enable a simple comparison with the results of Borgatti [5], who uses the same network in his paper. This also enabled us to make sure implementation of the NetLogo program was done correct, by verifying the results of Borgatti [5].

With the NetLogo simulation\(^1\) setup in Figure 3, it is then possible to simulate a flow of information through the network, using the buttons in the left side of the figure. This flow can be of any type, as mentioned in the introduction, as there are buttons for geodesics, path, trail and walk and different scripts were written for parallel and serial duplication as well as for transfer. We used a variable called information at each node that was incremented only for duplication and not

\(^1\) The NetLogo simulation can be found at: https://github.com/camrit/informationCentralitySimulation
for a transfer. Thus, we simulated the duplication or transfer of information. By counting the number of times a piece of information passes a node, or how long it takes to reach a node, certain statistics can be collected. We will refer to the number of times information passes a node that is not the final target node as frequency of arrival. The amount of time it takes for a node to be reached by all others will be called the arrival time. When looking at the frequency of arrival, a higher value indicates a more central node, since the node is often reached. For arrival time, a lower value indicates a more central node, since a node will be more central if it takes less time to reach it. A node that takes a longer time to reach is often not a very prominent node. We tried to replicate the setting of Borgatti [5] as a means to check our simulation setting, and also to validate his results further. Similar to Borgatti [5], we also ran the simulation a 1000 times for each combination of source and target nodes, thereby simulating the flow of information in the network (for the algorithm see Appendix A). We first explain the results of this first run. Then to test each simulation we multiplied the iterations by 100, leading to 100,000 simulation runs for each combination of source and target nodes. We then took the average of the results to arrive at the final simulation values.

To see if a metric can be used for the simulated type of information flow, it will be compared to these statistics. Figure 4 shows the program after a simulation is completed. Through the command line at the bottom, the program can be asked to display the values for one of the variables, in the case of Figure 4, one can see the values of the frequency of arrival when only the shortest paths are taken.

After the program was run with information flowing according to the transfer process and with the information allowed to run through geodesics, paths, trails or walks, the results were extracted. Table 3 shows the values for arrival times respectively, of the simulation in which transfer took place. Though we also calculated the frequency of arrival, we initially were only interested in the simulation’s arrival times (Table 3). The reason being that we thought information centrality, much like closeness centrality [5], is concerned with the length of time that it takes traffic to reach the node (the information it receives), rather than the actual frequency of traffic through the node. This is also the reason why information centrality is referred to as a closeness measure [7]. Another perspective is that once a node is reached the information is transferred to the node, and recurring transfers does not add to the information (in our simulation). A strict transfer of information can occur when no copy of the information is kept at any node (one can imagine such a scenario occurring in an electronic transfer of information). From Table 3, one can see that for every information flow process, the Medici family holds the most central place within the network. This is a very well-connected node with a degree of six, so it is directly linked to six others. It is therefore not surprising that it is the most central node in the network.

IV RESULTS

| Node   | Geodesics | Path  | Trail | Walk   |
|--------|-----------|-------|-------|--------|
| Medici | 25        | 41.48 | 25.07 | 115.15 |
| Ridolfi| 29        | 40.15 | 36.14 | 211.33 |
| Albizzi| 29        | 42.41 | 37.93 | 273.43 |
| Tornabuoni| 29     | 40.39 | 35.81 | 206.62 |
| Guadagni| 30       | 39.84 | 34.15 | 199.13 |
| Barbadori| 32     | 47.32 | 39.77 | 326.82 |
| Strozzi | 33        | 40.71 | 39.11 | 250.03 |
| Bischeri| 35       | 41.53 | 39.53 | 260.77 |
| Castellan| 36      | 41.48 | 38.38 | 294.23 |
| Salviati| 36       | 42.89 | 45.29 | 562.72 |
| Acciaiuoli| 38    | 44.92 | 49.69 | 630.67 |
| Peruzzi | 40        | 49.47 | 44.16 | 414.93 |
| Ginori  | 42        | 62.97 | 50.95 | 791.50 |
| Lamberti| 43       | 57.67 | 54.86 | 716.56 |
| Pazzi   | 49        | 55.88 | 58.30 | 1082.45 |
We also calculated the information centrality of the different nodes of the network, and then compared to the results of the simulations. Let’s first look at the actual values that the metric provided. These can be found in Table 4.

| Node    | Information centrality value |
|---------|-------------------------------|
| Medici  | 1.064                         |
| Guadagni| 0.917                         |
| Tornabuon| 0.902                        |
| Ridolfi | 0.901                         |
| Strozzi | 0.878                         |
| Bischeri| 0.832                         |
| Albizzi | 0.830                         |
| Castellan| 0.794                       |
| Peruzzi | 0.779                         |
| Barbadori| 0.763                       |
| Salvati | 0.598                         |
| Acciaiul| 0.554                         |
| Lambertes| 0.511                       |
| Ginori  | 0.483                         |
| Pazzi   | 0.394                         |

To compare the values from Table 3 and 4 with the raw values of information centrality (as done by Borgatti [5]) would not bring up any results, since the information centrality for a node is a relative number. Since the aim of the centrality metrics is to find the most central nodes and not the exact values per se, a comparison was made between the rankings of the nodes with the simulation values and the nodes ranked highest to lowest with information centrality. This was done for all the types of information flow in a network where transfer took place. These comparisons of the results are shown in Table 5.

| Information Centrality | Geodesics | Path | Trail | Walk |
|------------------------|-----------|------|-------|------|
| Medici                 | Medici    | Medici| Medici| Medici|
| Guadagni               | Albizzi   | Guadagni| Guadagni| Guadagni|
| Tornabuon              | Tornabuon | Tornabuon| Tornabuon| Tornabuon|
| Ridolfi                | Ridolfi   | Tornabuon| Ridolfi| Ridolfi|
| Strozzi                | Guadagni  | Strozzi| Albizzi| Strozzi|
| Bischeri               | Barbadori | Castellan| Castellan| Barbadori|
| Albizzi                | Strozzi   | Bischeri| Strozzi| Albizzi|
| Castellan              | Bischeri  | Albizzi| Bischeri| Castellan|
| Peruzzi                | Castellan | Salvati| Barbadori| Barbadori|
| Barbadori              | Salvati   | Acciaiul| Peruzzi| Peruzzi|
| Salvati                | Acciaiul  | Barbadori| Salvati| Salvati|
| Acciaiul               | Peruzzi   | Peruzzi| Acciaiul| Acciaiul|
| Lambertes              | Ginori    | Pazzi | Ginori| Lambertes|
| Ginori                 | Lambertes | Lambertes| Lambertes| Ginori|
| Pazzi                  | Pazzi     | Ginori| Pazzi| Pazzi|

In order to verify this result, we decided to test this correlation through another 100,000 runs of the simulation. Using Kendall’s Tau correlation2 (in R version 3.5.1, with confidence interval = 0.95) and the R corrplot package [28], and also including the ‘frequency of arrivals’ we arrive at Fig. 5 which shows the correlation matrix clustered horizontally, with a rectangle added to improve the readability. The GeoArT, WalkArT, PathArT, TrailArT are the arrival times of the information passed through the nodes during the simulation flows following Geodesics, Walks, Paths and Trails traversals, respectively. On the other hand, GeoFreqAr, WalkFreqAr, PathFreqAr, and TrailFreqAr are the frequency of the arrivals at each node in the network, when the graph traversal follows a Geodesic, Walk, Path and Trail respectively. The frequency of arrival variable is incremented for a node every time the object or information passes through it. On seeing Fig. 5, we notice that the relationship between

---

2 Pearson’s correlation could give misleading results if the association is not linear and the variables are not normally distributed [14].
the centrality metric and the type of flow is not what we see in Table 5. Fig. 5 shows the Information Centrality correlates with Eigenvector centrality metric, Degree centrality and to a certain extent with PathFreqAr, WalkFreqAr and TrailFreqAr, and even a lesser extent with Betweenness centrality. On the other hand, Closeness Centrality, as expected, correlates more with GeoArT (the transfer through Geodesic traversal). This is in line with the finding of Borgatti [5].

We also decided to test the external validity of this result by analysing networks other than the Padget Florentine network [1] used above. We decided to measure the correlation of the simulation values with the values of the various popular metrics from Table 2 for four other popular social networks of different sizes. Our sample of networks were taken from [39]. Table 6 gives an overview of the different networks. We consider small networks with 10 nodes (Knoke Bureaucracies [30]), medium size networks with 15 (Kapferer Mine network [31]) and 16 number of nodes (Padget Florentine network [1], Stokman-Ziegler network [32]) as well as relatively large networks with 34 nodes (Zachary Karate Club Network [33]).

Here we need to note that we do not consider networks with more than 34 nodes, due to limitations in time and computing power required for our complex simulations. The results of running Kendall’s tau correlation on the different networks mentioned in Table 6, are very similar to the correlation matrix seen in Fig. 5, and can be seen in Figures 6, 7, 8, and 9 in Appendix B. The Correlation Plots (corrplots) are identical for the different networks, except for Fig 10. (Zachary Karate Club network simulation [33]), in which the correlations are mirrored. However, we see that in Figure 10, the interrelations among the different centralities are essentially the same.

With these results, we cannot conclude that information centrally metric is suitable for networks in which information flows through walks but rather, that information centrality is more similar to Eigenvector and even Degree centrality and quite different from Closeness centrality.

V. DISCUSSION AND CONCLUSIONS

In Borgatti’s initial work [5], he made it clear that not every metric can be used in any network. He made a start with identifying which popular metrics (Freeman’s and Bonacich’s metrics) are best suitable for which types of networks. This resulted in metrics that are suitable for networks with parallel duplication or shortest paths but for networks with other information flow processes, metrics are still missing. We found papers that provided metrics, but none of the works used simulations to see if the metric was actually suitable for a certain type of network. One of the contributions of this paper was to show how such a simulation environment can be created using a popular simulation environment like NetLogo [27]. However, the main contribution is in determining the relation of the different network metrics for flows in different networks. Previous works have analysed network metric correlations in different network configurations [14, 15], while here we analyse the metric correlations with the outcome of flow simulations. We also include the information centrality [8], that earlier papers have largely ignored. Furthermore understanding the correlation between centrality metrics can help one in determining if one metric can replace another, especially if the calculation is relatively less complex [34].

We initially found that information centrality points to the same central places as the simulation did where the information flow was allowed to take walks (with the exception of two nodes). However, on further analysis we find that Information Centrality is more like Eigenvector and Degree centrality than to Closeness centrality as postulated by previous research [5]. We do see a pattern among the different correlation plots - Closeness centrality is rather different from the other metrics considered, based on simulation flows. Thereby indicating that the networks we considered in Table 6 are not threshold graphs [14].

This is our main contribution of this paper – including the network flow processes while analyzing the relationship among the important network centralities. Future work could try and find suitable metrics for other types of information flow, as well as network structure [14]. Further simulation studies can help in understanding the relationships among the different centrality metrics as well as network flows and thereby filling our gaps in knowledge.

| Name                        | Publication                  | Number of nodes |
|-----------------------------|------------------------------|-----------------|
| Knoke Bureaucracies network | [30]                         | 10              |
| Kapferer Mine network       | [31]                         | 15              |
| Padget Florentine network   | [1]                          | 16              |
| Stokman-Ziegler network     | [32]                         | 16              |
| Zachary Karate Club Network | [33]                         | 34              |

REFERENCES

[1] J. F. Padgett and C. K. Ansell, "Robust Action and the Rise of the Medici, 1400-1434," American Journal of Sociology, vol. 98, no. 6, pp. 1259-1319, 1993/05/01/ 1993.
[2] T. Coffman, S. Greenblatt, and S. Marcus, "Graph-based technologies for intelligence analysis," Communications of the ACM, vol. 47, no. 3, pp. 45-47, 2004.
[3] E. M. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant MANETs," IEEE Transactions on Mobile Computing, vol. 8, no. 5, pp. 606-621, 2009.
[4] M. B. Baig and L. Akoglu, "Correlation of node importance measures: An empirical study through graph robustness," in Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 275-281: ACM.
Appendix A

Algorithm 1: For running the simulation and calculating the metrics

```
Result: Write here the result
1 for i ← 1 to 1000 do
2    /* Independent trials*/
3    for j ← 1 to 1000 do
4        /* For each source node*/
5            for j ← 1 to 1000 do
6                /* For each target node*/
7                    Simulate flow i from j to k;
8                end
9            end
10        end
11 Compute node statistics for this trial;
12 end
13 Average node statistics and compare with centrality measures;
```

As mentioned earlier, the NetLogo simulation can be found at https://github.com/camrit/informationCentralitySimulation.

Appendix B

Figure 6: The Correlation Matrix plot (using corrplot) for the Knoke Bureaucracies network simulation [30]
Figure 7: The Correlation Matrix plot (using corrplot) for the Kapferer Mine network simulation [31]

Figure 8: The Correlation Matrix plot (using corrplot) for the Stokman-Ziegler network simulation [32]

Figure 9: The Correlation Matrix plot (using corrplot) for the Zachary Karate Club network simulation [33]