University of Cyprus
Department of Economics

Working Paper 04-2016

On decay centrality

Nikolas Tsakas

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Tel.: +357-22893700, Fax: +357-22895028, Web site: http://www.ucy.ac.cy/econ/en

On Decay Centrality*

Nikolas Tsakas†

Department of Economics, University of Cyprus

April 19, 2016

Abstract

We establish a relationship between decay centrality and two widely used and computationally cheaper measures of centrality, namely degree and closeness centrality. We show that for low values of the decay parameter the nodes with maximum decay centrality also have maximum degree, whereas for high values of the decay parameter they also maximize closeness. For intermediate values of the decay parameter, we perform an extensive set of simulations on random networks and find that maximum degree or closeness are good proxies for maximum decay centrality. In particular, in the vast majority of simulated networks, the nodes with maximum decay centrality are characterized by a threshold on the decay parameter below which they belong to the set of nodes with maximum degree and above which they belong to the set of nodes with maximum closeness. The threshold values vary with the characteristics of the network. Moreover, nodes with maximum degree or closeness are highly ranked in terms of decay centrality even when they are not maximizing it. The latter analysis allows us to propose a simple rule of thumb that ensures a nearly optimal choice with very high probability.

JEL Classification: C63, D85

Keywords: Decay Centrality, Centrality Measures, Networks

---

*Part of this project was carried out while the author was at SUTD–MIT International Design Center at Singapore University of Technology and Design supported by grant IDG31300110.

†Correspondence Address: University of Cyprus, Department of Economics, PO Box 20537, 1678 Nicosia, Cyprus. E–mail: `tsakas.nikolaos@ucy.ac.cy`.

1. Introduction

Decay centrality is a measure of centrality in which a node is rewarded for how close it is to other nodes, but in a way that very distant nodes are weighted less than closer ones (see Jackson, 2008). It is defined as $\sum_{j \neq i} \delta^{d(i,j)}$, where $0 < \delta < 1$ is a decay parameter and $d(i,j)$ is the geodesic distance between nodes i and j. For low values of δ decay centrality puts much more weight on closer nodes, thus becoming proportional to degree centrality, whereas for high values of δ it measures the size of the component a node lies in.

It is considered to be richer than other distance related measures, because it captures the idea that the importance of a node for another is proportional to their distance (see for instance Jackson and Wolinsky, 1996). More recently, it has been considered important in problems of optimal targeting selection in networks (see Banerjee et.al, 2013; Chatterjee and Dutta, 2015; Tsakas, 2014). In particular, in two different environments, Chatterjee and Dutta (2015) and Tsakas (2014) find decay centrality to be the measure that helps selecting the node that can lead to the maximum diffusion of a given action in a social network.

Nevertheless, its use is cumbersome for two main reasons. First, except in very simple structures, which are the nodes with maximum decay centrality cannot be easily identified, since the measure depends vastly on the exact network topology and the value of the decay parameter. Second, calculating the decay centrality of all nodes and subsequently choosing the one that maximizes it might be computationally costly, since it requires calculating the geodesic distance between each pair of nodes and subsequently summing a function of them. This is particularly important for large networks.[1]

The aim of this paper is to show the close connection between decay centrality and two well–studied and computationally cheaper measures, namely degree and closeness centrality. The relationships are established both analytically and numerically and provide evidence that the nodes with maximum decay centrality usually belong either to the set of nodes with maximum degree or to the set of nodes with maximum closeness.

In particular, focusing on connected networks, we show that for sufficiently low values of the decay parameter the nodes that maximize decay centrality belong to the set of nodes with maximum degree, whereas for sufficiently high values of the decay parameter the nodes that maximize decay

[1]For a network with n nodes, the time complexity for the calculating degree and closeness centrality are in $O(n^2)$ and $O(n^3)$ respectively (see Brandes and Erlebach, 2005), where for the calculation of shortest paths that is necessary for closeness centrality is used the simple Dijkstra algorithm (see Dijkstra, 1959). Once the shortest paths have been calculated, decay centrality requires the calculation of $\delta^{d(i,j)}$ for each pair of nodes (i,j). Hence, the time complexity of calculating decay centrality is in $O(n^5)$.

centrality maximize closeness as well. The first proposition is not surprising as it is already known that for low values of δ decay centrality is proportional to degree. However, the second proposition establishes a novel relationship between decay and closeness centrality for high values of δ, for which so far decay centrality was associated only with the size of the component a node lied in.

Nevertheless, despite establishing these results what happens for intermediate values of the decay parameter still remains unanswered. We tackle this problem numerically and using an extended set of simulations we find that in the vast majority of cases the nodes with maximum decay centrality belong either to the set of nodes with maximum degree or to the set of nodes with maximum closeness centrality. When the two sets do not intersect, we observe that for low values of δ the decay centrality is maximized by nodes with maximum degree and as δ increases there is a threshold above which decay centrality is maximized by nodes with maximum closeness. The threshold varies with the network parameters. It occurs very rarely that for some value of δ a node maximizes decay centrality without having either maximum degree or closeness and even in that case the rank in terms of decay centrality of nodes with maximum degree or closeness is very close to the top.

When the two sets intersect a node that belongs to their intersection is almost always the one with maximum centrality for all values of δ. We study how this result is affected by several network parameters and analyze the distribution of ranks in decay centrality of nodes with maximum degree or closeness, so as to understand how suboptimal these choices may be. We find that even in when considering the 95th percentile, the rank of the nodes is relatively high, but it is affected significantly by δ. Nevertheless, we find that considering a rule of thumb with a threshold at $\delta = 0.5$, below which a node with maximum degree is chosen and above which a node with maximum closeness is chosen, is sufficient to ensure that the chosen node is ranked among the top nodes in terms of decay centrality, with probability at least equal to 95%. This provides a very useful rule of thumb, as it is clear–cut and computationally cheap. We also present a simple econometric analysis that shows the effect of other measurable network characteristics on the likelihood that a node maximizing either degree or closeness maximizes decay centrality as well.

In general, the results suggest that in most cases, given a decay parameter, one can use degree or closeness as proxies for decay centrality, without risking to make a particularly suboptimal choice.

2. Notation

Consider a set of nodes N, with cardinality n, which are connected through a network. A network is represented by a family of sets $\mathcal{N} := \{N_i \subseteq N \mid i = 1, \ldots, n\}$, with N_i denoting the set of nodes that are directly connected with i. N_i is called i's neighborhood and its cardinality, $|N_i|$, is called

i's *degree*. We focus on undirected networks, where $j \in N_i$ if and only if $i \in N_j$. It is also useful to define the set of nodes with maximum degree, i.e. $I_{deg} = \arg\max_{i \in N} |N_i|$.

A path in a network between nodes i and j is a sequence $i_1, ..., i_K$ such that $i_1 = i$, $i_K = j$ and $i_{k+1} \in N_{i_k}$ for $k = 1, ..., K-1$. The geodesic distance, $d(i,j)$, between two nodes in the network is the length of the shortest path between them. We say that two nodes are connected if there exists a path between them. The network is connected if every pair of nodes is connected. We focus on connected networks, nevertheless for disconnected networks the analysis would be identical for each of their connected components.[2]

The *closeness centrality* (or simply *closeness*) of a node $i \in N$ is defined as the inverse of the sum of the geodesic distances from each other agent in the network, i.e. $C_i = \frac{1}{\sum_{j \neq i} d(i,j)}$. Notice that closeness centrality measures how easily a node can reach all other nodes in the network. According to this definition, we define the set of nodes with maximum closeness centrality, i.e. $I_{clos} = \arg\max_{i \in N} \sum_{j \neq i} C_i$.

Finally, given a decay parameter $\delta \in (0,1)$, the decay centrality of node $i \in N$ is $DC_i^\delta = \sum_{j \neq i} \delta^{d(i,j)}$. The decay centrality is a function of distances from each node in the network, adjusted by a decay parameter that makes distant nodes count less than closer ones. As in the previous two cases, for each value of δ, we define the set of nodes with maximum decay centrality, i.e. $I_{dc}^\delta = \arg\max_{i \in N} \sum_{j \neq i} DC_i^\delta$.

3. Analytical Results

Proposition 1. *Exists $\underline{\delta}$ such that for all $\delta \in (0, \underline{\delta})$ holds that $I_{dc}^\delta \subseteq I_{deg}$.*

Proof of Proposition 1. To prove the argument it is enough to ensure that for these values of δ a node with maximum decay centrality, $i \in I_{dc}^\delta$, should necessarily be among the nodes with maximum degree, i.e. $i \in I_{deg}$. Notice that for all $\delta \in (0,1)$, $\arg\max_{i \in N} \sum_{j \neq i} \delta^{d(i,j)} = \arg\max_{i \in N} \sum_{j \neq i} \delta^{d(i,j)-1}$, since in the right hand side all arguments are divided by δ. This modification allows to use the argument that $\lim_{\delta \to 0} \sum_{j \neq i} \delta^{d(i,j)-1} = |N_i|$, which holds because $\lim_{\delta \to 0} \delta^{d(i,j)-1} = 1$ if $d(i,j) = 1$ and is equal to 0 otherwise. Therefore, in the limit, maximizing decay centrality coincides with maximizing degree. Moreover, the objective function is continuous in δ, which means that if $\lim_{\delta \to 0} \sum_{j \neq i} \delta^{d(i,j)-1} = |N_i| > |N_k| = \lim_{\delta \to 0} \sum_{j \neq i} \delta^{d(k,j)-1}$ for $i, k \in N$, hence there exists $\underline{\delta} > 0$ such that $\sum_{j \neq i} \delta^{d(i,j)-1} \geq \sum_{j \neq i} \delta^{d(k,j)-1}$, or equivalently $\sum_{j \neq i} \delta^{d(i,j)} \geq \sum_{j \neq i} \delta^{d(k,j)}$, for all $\delta \in (0, \underline{\delta})$. $\square$

[2] A connected component is a non-empty sub-network N' such that (i) $N' \subset N$, (ii) N' is connected and (iii) if $i, j \in N'$ and $j \in N_i$, then $j \in N'_i$.

Proposition 2. *Exists $\overline{\delta}$ such that for all $\delta \in (\overline{\delta}, 1)$ holds that $I_{dc}^{\delta} \subseteq I_{clos}$.*

Proof of Proposition 2. To prove the argument is enough to ensure that for these values of δ a node with maximum decay centrality, $i \in I_{dc}^{\delta}$, should necessarily be among the nodes with maximum closeness, $i \in I_{clos}$. Consider decay centrality as a function of δ, i.e. $f_i(\delta) = \sum_{j \neq i} \delta^{d(i,j)}$ and notice that the function is continuously differentiable in δ, as well as that $\lim_{\delta \to 1} f_i(\delta) = |N|$ for all $i \in N$. Differentiating f_i with respect to δ we get that $f_i'(\delta) = \sum_{j \neq i} d(i,j)\delta^{d(i,j)-1} > 0$ and $\lim_{\delta \to 1} f_i'(\delta) = d(i,j)$. The result becomes straightforward noticing that for f_i, f_k being increasing functions for which it holds that $f_i(1) = f_k(1)$ and $f_i'(1) < f_k'(1)$ there must exist $\overline{\delta}$ such that $f_i(\delta) > f_k(\delta)$ for all $\delta \in (\overline{\delta}, 1)$. Therefore, in order a node to maximize decay centrality for $\delta \in (\overline{\delta}, 1)$ it should minimize $\sum_{j \neq i} d(i,j)$, or equivalently maximize $\frac{1}{\sum_{j \neq i} d(i,j)}$ (i.e. closeness). $\square$

The two propositions establish the fact that in the two limits decay centrality coincides with degree and closeness respectively. However, the characteristics that make nodes have high decay centrality for intermediate values of the decay parameter still remain unexplored.

4. Numerical Results

We simulate random undirected Èrdos-Renyi networks (Erdős and Rényi, 1959), $G(n, p)$, where n is the network size and p is the probability of two nodes being linked. The networks are required to be connected so that geodesic distances are well defined. We consider five distinct network sizes spanning from 10 to 200 nodes and ten link probabilities spanning from 0.05 to 0.5 and perform 10000 trials for each configuration. Our aim is to understand to what extent there is a connection between nodes with high decay centrality and nodes with either high degree or high closeness for intermediate values of δ.

The first question we pose is how often I_{dc}^{δ}, i.e. the set of nodes with maximum decay centrality, intersects with either I_{deg}, i.e. the set of nodes with maximum degree, or with I_{clos}, i.e. the set of nodes with maximum closeness. We find that inn the vast majority of the cases $I_{dc}^{\delta} \subseteq (I_{deg} \cup I_{clos})$ for almost all intermediate values of δ and not only for the limit values (close to 0 or 1), as the theory has predicted. This suggests that focusing on nodes with either maximum degree or maximum closeness will often be sufficient to ensure the maximization of decay centrality.

Before exploring this result further, it is important to mention that I_{deg} intersects with I_{clos} quite often for random networks.[3] The reasons why this occurs are outside the scope of this paper, however

[3]A non–empty intersection between I_{deg} and I_{clos} is observed very often for sufficiently dense networks, namely

it has an apparent effect on our results as it provides a natural connection between the two limit cases explored by theory. In fact, in "almost all" of the cases where there are nodes that belong both to I_{deg} and I_{clos}, those nodes also belong to I^{δ}_{dc}. This result cannot be generalized theoretically as there are cases in which, for some intermediate values of δ, the nodes with maximum decay centrality do not belong to either I_{deg} or I_{clos}. Nevertheless, as it becomes apparent when comparing Tables 1 and 2 the frequency with which such cases arise is practically negligible. Later on, we explore the rank in decay centrality of nodes with maximum degree or closeness when they are not ranked first and we find that they are still very highly ranked. This result provides a first strong argument, which is that if there exists a node $i \in N$ such that $i \in (I_{deg} \cap I_{clos})$ then almost always $i \in I^{\delta}_{dc}$ as well.

n \ p	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	8620	8756	8951	9035	9341	9570	9752	9879	9971	9987
20	6600	7229	8113	8711	9349	9752	9969	9998	10000	10000
50	5921	7620	8220	9330	9964	10000	10000	10000	10000	10000
100	6783	7401	8794	9989	10000	10000	10000	10000	10000	10000
200	7209	7593	9976	10000	10000	10000	10000	10000	10000	10000

Table 1: Frequency of occassions where $I_{deg} \cap I_{clos} \neq \emptyset$.

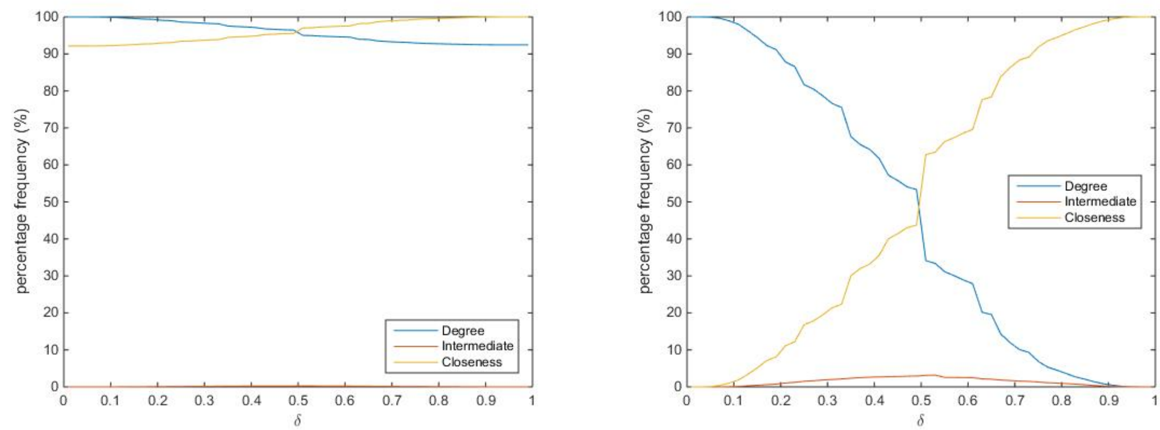
n \ p	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0	0	0	0	0	0	0	0	0	0
20	11	2	0	0	0	0	0	0	0	0
50	30	0	0	0	0	0	0	0	0	0
100	14	0	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0	0	0

Table 2: Frequency of occassions where $I_{deg} \cap I_{clos} \neq \emptyset$ and $I^{\delta}_{dc} \nsubseteq (I_{deg} \cap I_{clos})$ for some value of δ.

We turn our attention to the case where I_{deg} and I_{clos} do not intersect. In this case, we expect from theory a transition in the nodes that belong to I^{δ}_{dc} as δ increases. It turns out that even in this cases, most of the times the transition is immediate, meaning that a node $i \in I^{\delta}_{dc}$ belongs either to I_{deg} or to I_{clos}. This can become apparent in Figure 1, which contains the percentage frequencies

$p > 0.15$, as it can be seen in Table 1 and Figure 9 in the Appendix.

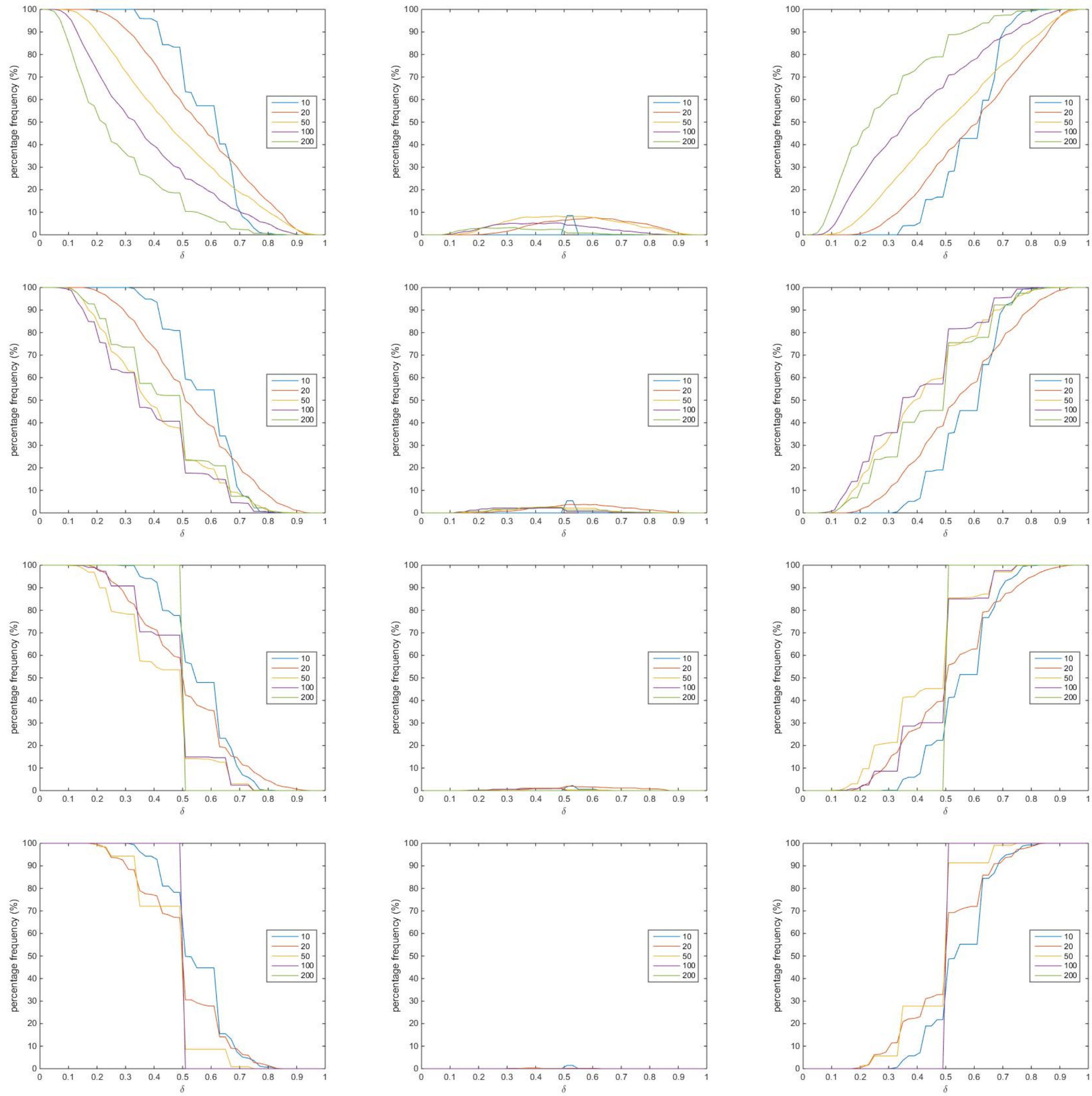
with which $I^{\delta}_{dec} \subseteq I_{deg}$ (blue), $I^{\delta}_{dec} \subseteq I_{clos}$ (yellow) and $I^{\delta}_{dc} \cap (I_{deg} \cup I_{clos}) = \emptyset$ (red). For each value of δ the frequencies correspond to the fraction of the simulated networks in which each of the three conditions held true. The fact that in the left subfigure I_{deg} and I_{clos} can intersect means that the sum of the frequencies may exceed 100%. In fact, this seems to be the case rather often. This in no longer possible in the right subfigure, where we include only the cases where I_{deg} and I_{clos} do not intersect, therefore the three conditions are mutually exclusive. Note that, the latter case is never observed in more than a 2% of the trials, with the percentage becoming much lower as we get further away from $\delta = 0.5$. This suggests that for most networks there is a threshold value of δ below which $I^{\delta}_{dec} \subseteq I_{deg}$ and above which $I^{\delta}_{dec} \subseteq I_{clos}$.



(a) Including also cases where $I_{deg} \cap I_{clos} \neq \emptyset$ (b) Including only cases where $I_{deg} \cap I_{clos} = \emptyset$

Figure 1: The blue (yellow) line shows the percentage frequency with which $I^{\delta}_{dec} \subseteq I_{deg}$ ($I^{\delta}_{dec} \subseteq I_{clos}$), whereas the red line shows the frequency with which it does not belong to any of the two sets.

Figure 1 contains the percentage frequencies after pooling all different (n, p)–configurations, which yields a reasonable question on how results might differ for different values of parameters p and n. Figure 2 contains the three percentage frequencies of interest for all network sizes and values of $p \in \{0.05, 0.1, 0.15, 0.2\}$, focusing again on networks where I_{deg} and I_{clos} do not intersect. The results are qualitatively similar in all configurations, presenting an inverted S-shaped curve for the frequency of $I^{\delta}_{dec} \subseteq I_{deg}$, an S-shaped curve for the frequency of $I^{\delta}_{dec} \subseteq I_{clos}$ and an inverted bell curve for the frequency of $I^{\delta}_{dc} \cap (I_{deg} \cup I_{clos}) = \emptyset$. Regarding the latter one, we observe that its frequency never exceeds 10%, with this being the case only for $p = 0.05$ and values of δ close to 0.5. As far as it concerns the transition from nodes with maximum degree to those with maximum centrality, we observe this to occur for lower values of δ as the networks become larger. This result is more prevalent



(a) $I_{dc}^{\delta} \subseteq I_{deg}$ (b) $I_{dc} \nsubseteq (I_{deg} \cup I_{clos})$ (c) $I_{dc}^{\delta} \subseteq I_{clos}$

Figure 2: Percentage frequency of decay parameters for the three transitions presented separately for each network size. The four rows correspond to $p = 0.05, 0.1, 0.15$ and 0.2 respectively.

for low values of p; as p increases we observe a sharp transition occurring for δ very close to 0.5.[4]

The previous observations raise two important questions. First, what is the rank in terms of decay centrality of nodes with maximum degree and closeness, when they are not ranked first? The answer to this question would determine how suboptimal could be the choice of a node based on its degree or closeness when one cares about maximizing decay centrality. Figures 3 and 4 show the average rank, as well as the 5th and 95th percentiles of rank distribution in decay centrality of nodes with maximum degree and closeness respectively.[5] It turns out that nodes belonging to I_{deg} or I_{clos} are highly ranked in terms of decay centrality for all values of δ, even when they are not ranked first. The result is similar if we exclude the networks in which I_{deg} and I_{clos} intersect. This reinforces the argument that choosing among nodes with maximum degree or closeness can be an adequate proxy for nodes that maximize decay centrality.

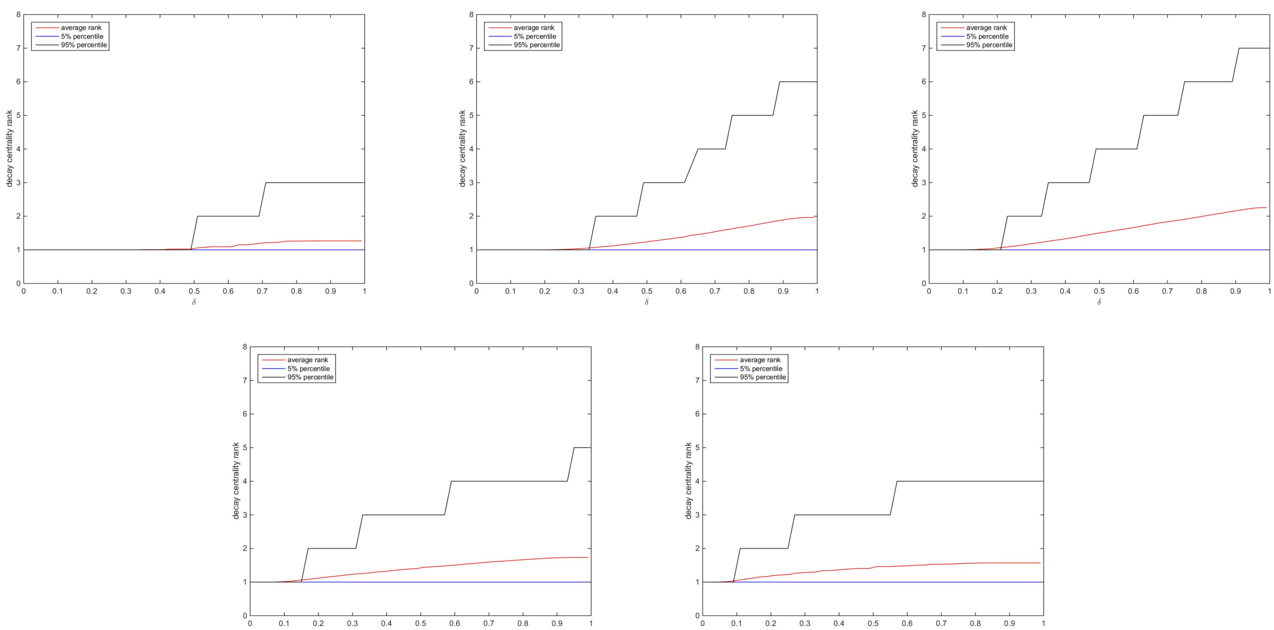


Figure 3: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum degree, including all networks. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.05$.

Nevertheless, focusing either only on degree or only on closeness leads to increasingly suboptimal choices as δ moves towards one of the extremes. Hence, it remains to be clarified which of the two

[4]It should be mentioned that for $p \geq 0.2$ there are very few observations where I_{deg} and I_{clos} do not intersect, which may affect the weight of each individual trial in the result.

[5]The two figures corresponds to a particular pair (n, p), nevertheless the results are qualitatively similar for other configurations, some of which can be found in the Appendix (see Figures 10, 11, 12 and 13).

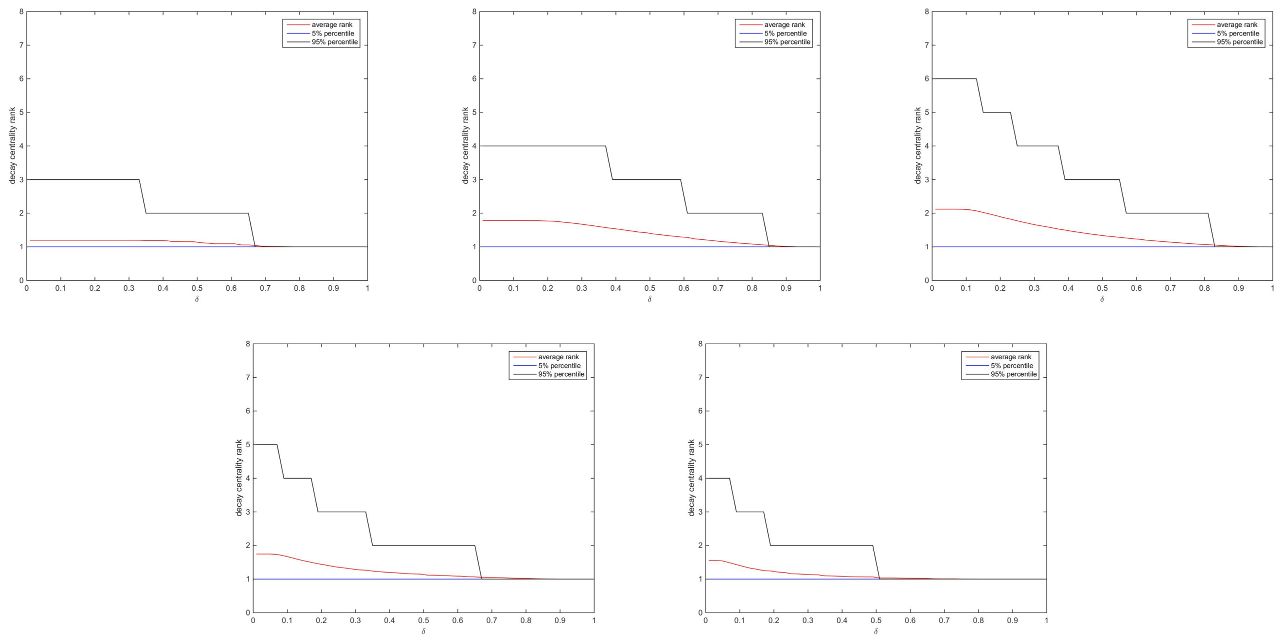


Figure 4: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum closeness, including all networks. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.05$.

sets would provide better candidates depending on the value of the decay parameter one is interested in. Intuitively, when interested in high values of δ, nodes that maximize closeness centrality would be more natural candidates and vice versa when interested in low values of δ nodes that maximize degree centrality would be more natural candidates.

Ideally, we would like to have a simple rule of thumb that would facilitate this choice. A quite natural rule would be to choose among the nodes from I_{deg} for $\delta < 0.5$ and a node from I_{clos} for $\delta > 0.5$. Given that the two sets usually contain few nodes, it should not be too costly to calculate the decay centrality of each of these nodes and pick the one that maximizes it.

It turns out that this rule of thumb is sufficient to ensure that the chosen node will be ranked among the top in terms of decay centrality. Figure 5 shows the same three statistics as before for $p = 0.05$ and all network sizes, in which it can be seen that in all networks a node chosen according to this rule will be ranked in terms of decay centrality among the top three with probability 95%.[6]

In addition to this, we try to identify the threshold values at which the transitions actually occur. Figure 6 shows the values of δ that correspond to the three potential transitions, namely: (a) the lowest δ for which I_{deg} stops containing I_{dc}^{δ}, (b) the lowest δ for which I_{dc}^{δ} does not coincide with either

[6]Figure 14 in the Appendix shows a similar picture when excluding networks where I_{deg} and I_{clos} intersect. Even in this case, a node chosen according to the rule is ranked among the top five with probability at least equal to 95%.

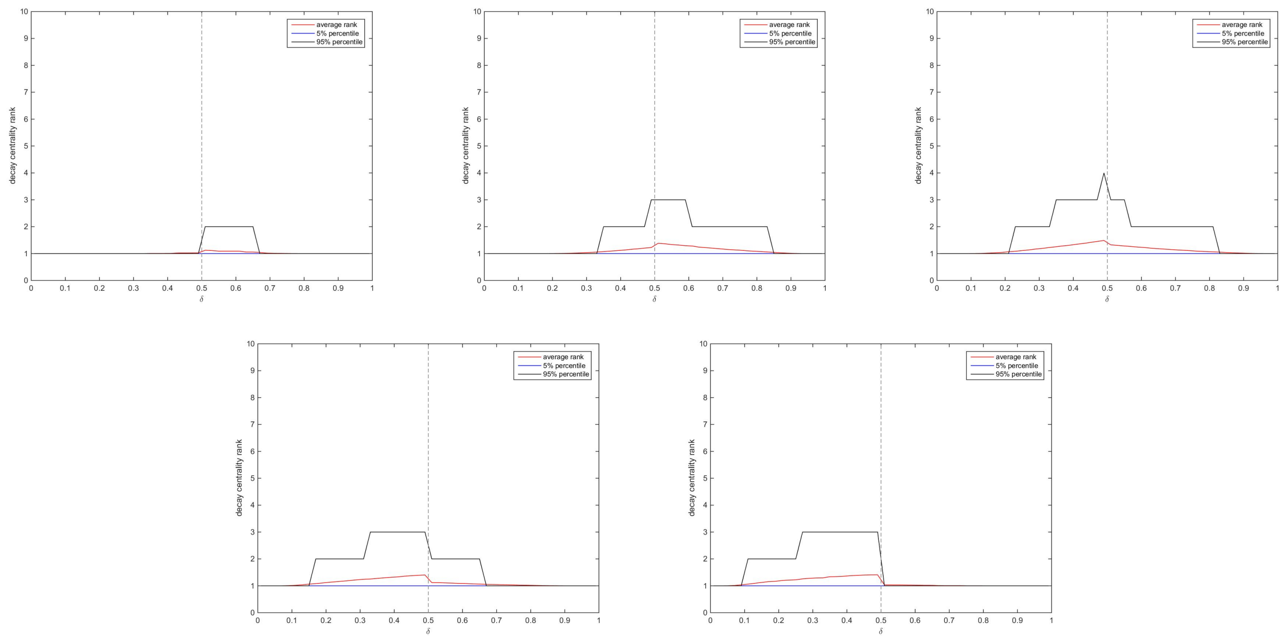
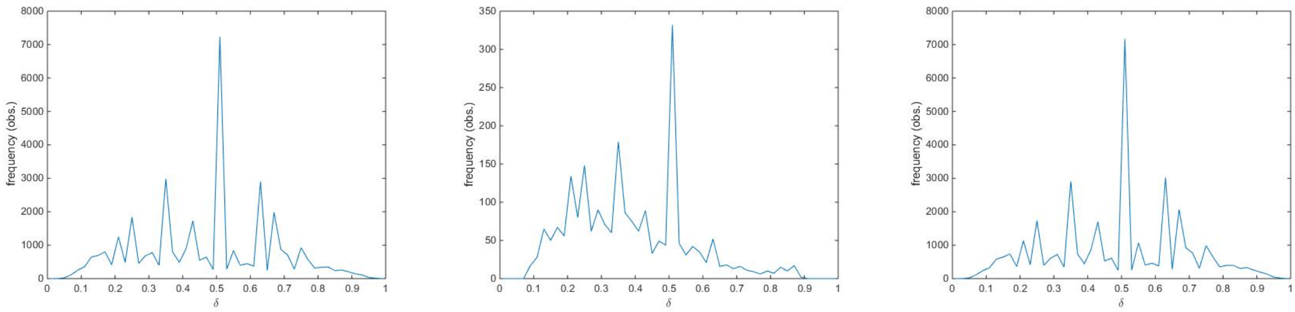


Figure 5: Average rank (red), as well as 5% (blue) and 95% (black) percentiles, of decay centrality of nodes with maximum degree for $\delta < 0.5$ and with maximum closeness for $\delta > 0.5$, including all networks. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.05$.

I_{deg} or I_{clos} and (c) the lowest δ for which I_{clos} starts containing I_{dc}^{δ}.[7] Figure 7 shows the respective cumulative distributions of the values. This visualization is more useful for the comparison between different networks. Once again, we focus on graphs where I_{deg} and I_{clos} do not intersect. A first observation is that the graphs corresponding to transitions (a) and (c) seem very similar, which is expected given that the transition occurs most of the times directly from one set to the other. Moreover, when reporting the number of observations, it becomes apparent that the intermediate region, where I_{dc}^{δ} is not contained in $I_{deg} \cup I_{clos}$, occurs very rarely compared to the other two; a result that is much more prevalent as the values of p increase. However, there is little additional information one can gain from observing the three plots. As an attempt to obtain a better understanding, we also present the plots corresponding to different network sizes and link probabilities.

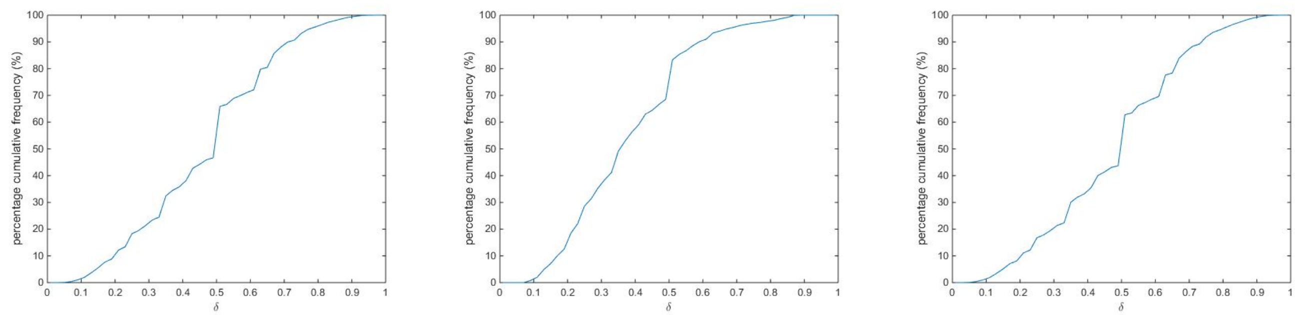
Figure 8 shows the cumulative distributions of the values of δ per network size, for p equal to $0.05, 0.1, 0.15$ and 0.2 respectively. A comparison across networks reveals that the transition tends to occur earlier in larger networks, mainly for low values of p. As p becomes larger, we observe an increasing concentration of observations around the cutoff point of $\delta = 0.5$. The graphs seem quite

[7]The three plots are sufficient to account for almost all observations, as transitions in the opposite direction occur almost never.



(a) I_{deg} stops containing I_{dc}^{δ} (b) $I_{deg} \cup I_{clos}$ stops containing I_{dc}^{δ} (c) I_{clos} starts containing I_{dc}^{δ}

Figure 6: Frequency of decay parameters for the three transitions presented pooled across all networks sizes and link probabilities.



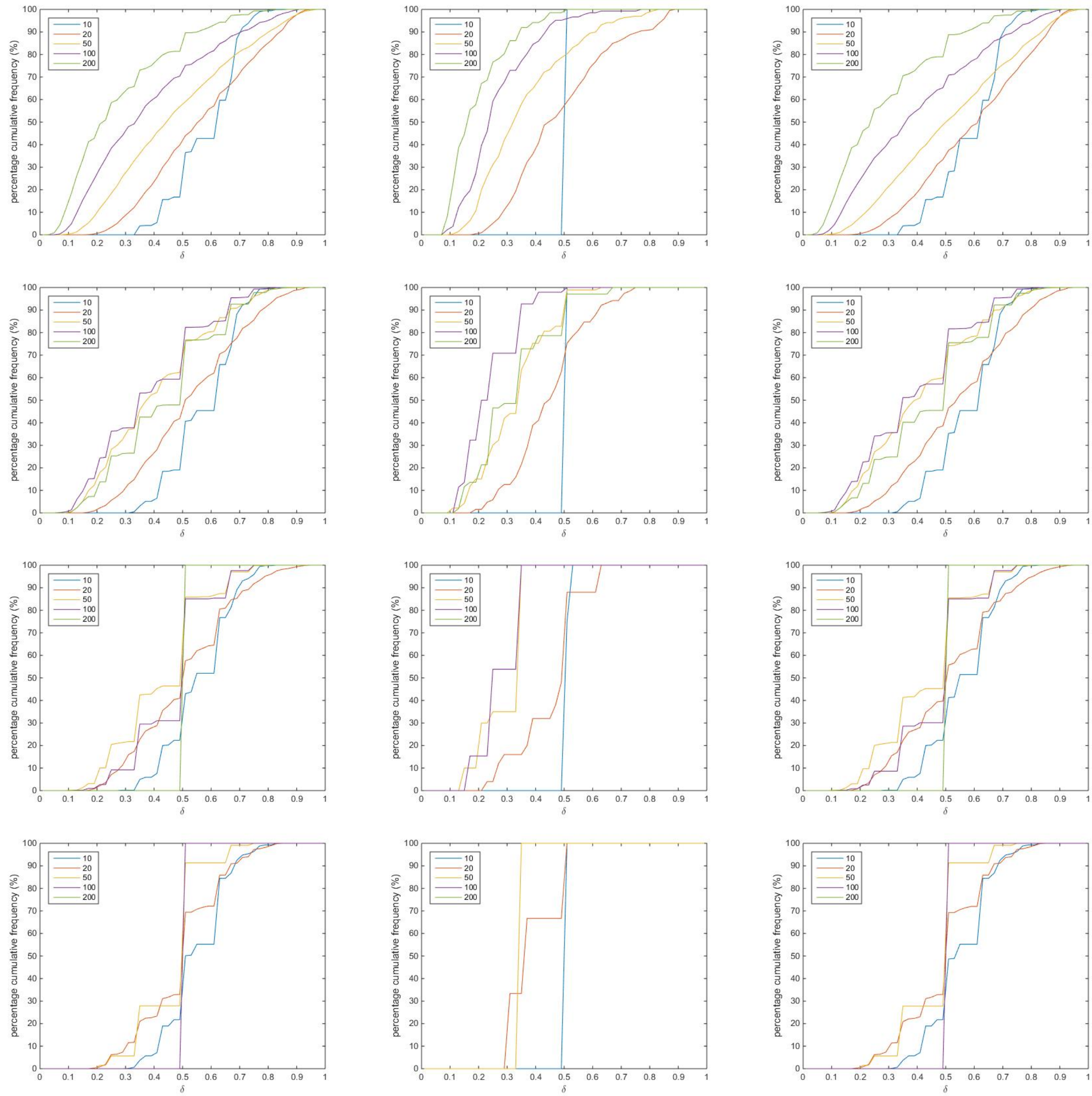
(a) I_{deg} stops containing I_{dc}^{δ} (b) $I_{deg} \cup I_{clos}$ stops containing I_{dc}^{δ} (c) I_{clos} starts containing I_{dc}^{δ}

Figure 7: Percentage cumulative frequency of decay parameters for the three transitions presented pooled across all networks sizes and link probabilities.

similar with those obtained by Figure 2, despite referring to different measures. This similarity is mainly due to the fact that in the vast majority of cases $I_{dc}^{\delta} \subseteq (I_{deg} \cup I_{clos})$. A piece of information that is missing from those graphs, is that for $p > 0.1$ the intermediate region, where I_{dc}^{δ} is not contained in $I_{deg} \cup I_{clos}$, occurs extremely rarely, as the two sets either intersect or the transition from one towards the other occurs directly.

Finally, we try and identify some other observable characteristics of the network that could provide information on which of the two notions of centrality would be adequate in each case. Doing so graphically would be cumbersome, as the results would need to be reported for specific values of n, p and δ, thus limiting their expositional clarity. For this reason, we perform a simple econometric analysis that allows us to get an idea on the effect of each characteristic at once.

We focus on networks where I_{deg} and I_{clos} do not intersect and we construct a discrete choice model with three alternatives: (i) $I_{dc}^{\delta} \subseteq I_{deg}$, (ii) $I_{dc}^{\delta} \subseteq I_{clos}$ and (iii) $I_{dc}^{\delta} \nsubseteq (I_{deg} \cup I_{clos})$. Given that



(a) I_{deg} stops containing I_{dc}^{δ} (b) $I_{deg} \cup I_{clos}$ stops containing I_{dc}^{δ} (c) I_{clos} starts containing I_{dc}^{δ}

Figure 8: Percentage cumulative frequency of decay parameters for the three transitions presented separately for each network size. The four rows correspond to $p = 0.05, 0.1, 0.15$ and 0.2 respectively.

I_{deg} and I_{clos} do not intersect these three alternatives are mutually exclusive. We run a standard multinomial logit regression attempting to estimate the relative probability of each alternative arising, controlling for several network characteristics. Table 3 shows the results for all networks for $p = 0.05$.[8] Alternative (i) is considered as the base and only the coefficients related to (ii) are reported. The main information one gets from these is the sign of the coefficient, which shows the effect of a characteristic on the relative likelihood of (ii) compared to (i), i.e. a positive coefficient signifies that an increase in the given variable makes more likely that a node with maximum closeness will maximize decay centrality compared to a node with maximum degree.[9] Apart from the already established effect of δ and the not surprising effect of maximum and average degree and maximum closeness, we find that a larger diameter increases the likelihood of (ii) compared to (i), except in very small networks. This result is quite intuitive, as in networks with larger diameter a node with high degree may still be quite far from certain nodes located in the periphery of the network, which affects negatively the decay centrality of the node.[10]

5. Discussion

We have established a clear relationship between decay centrality and two widely used measures of centrality, namely degree and closeness, showing that nodes that maximize one of the two measures are natural candidates for maximizing decay centrality. In fact, the majority of networks has a threshold value of δ below which maximum decay centrality coincides with maximum degree and above which it coincides with maximum closeness. We show that a simple rule of thumb that considers a common threshold at $\delta = 0.5$ seems to perform particularly well. The variety of threshold values observed for different networks raises the question on whether there are some particular characteristics of the network that can allow the characterization of this threshold value with some accuracy; a question that has been only partially tackled here. Finally, simulations are limited to networks of small to medium size (up to 200 nodes), due to computational limitations. There is no observation suggesting that passing to larger networks should alter the results qualitatively, however an extension of the analysis to large networks would ensure their applicability to problems where decay centrality has been shown to play an important role.

[8] Table 4 in the Appendix shows the same analysis for $p = 0.10$. For larger values of p and large networks convergence of the method is not always guaranteed, mainly because there are very few observations of alternative (iii).

[9] The coefficients are expected to be strongly significant because of the large sample size, which should make one even more cautious when attempting to explain them.

[10] This result might be partially driven by the fact that random networks tend to have small diameters (for a thorough analysis see Bollobás, 1981; Vega–Redondo, 2007; Jackson, 2008).

For p=0.05	n=10	n=20	n=50	n=100	n=200
Maximum Degree	-1.980***	-0.833***	-0.641***	-0.868***	-0.884***
Maximum Closeness	0.353***	0.175***	0.0897***	0.183***	0.180***
Delta	16.18***	10.91***	8.817***	8.972***	10.16***
Diameter	-0.913***	0.133***	0.124***	0.0691***	0.0931***
Average Degree	-1.578***	-1.372***	-1.534***	-1.824***	-1.139***
St. Dev. Degree	-0.155	-0.0220	-0.103***	-0.235***	-0.0526***
Simulation	-6.20e-09	3.16e-07***	1.03e-07*	-1.85e-07***	-5.70e-07***
Constant	13.59***	8.369***	15.97***	60.53***	106.0***
N	69000	170000	203950	160850	139550

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Coefficients of multinomial logit regression with alternatives (i) $I^{\delta}_{dc} \subseteq I_{deg}$, (ii) $I^{\delta}_{dc} \subseteq I_{clos}$ and (iii) $I^{\delta}_{dc} \nsubseteq (I_{deg} \cup I_{clos})$, for networks where I_{deg} and I_{clos} do not intersect. Alternative (i) is taken as the base and the reported coefficients correspond to (ii). Results are for all networks and $p = 0.05$.

References

Banerjee, A., Chandrasekhar, A.G., Duflo, E. & Jackson, M. (2013). The Diffusion of Microfinance. *Science* 341 (6144).

Bollobás (1981). The Diameter of Random Graphs. *Transactions of the American Mathematical Society* 267(1), 41–52.

Brandes, U. & Erlebach, T. (2005). *Network Analysis: Methodological Foundations*. Springer.

Chatterjee, K. & Dutta, B. (2015). Credibility and Strategic Learning in Networks. *International Economic Review*, forthcoming.

Dijkstra, E.W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* 1, 269–27.

Erdős, P. & Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae Debrecen* 6, 290–291.

Jackson, M.O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. and Wolinsky, A. (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71:44–74.

Tsakas, N. (2014). Optimal Influence under Observational Learning. *SSRN Working Paper* No. 2449420.

Vega–Redondo (2007). *Complex Social Networks*. Cambridge University Press.

A. Graphs

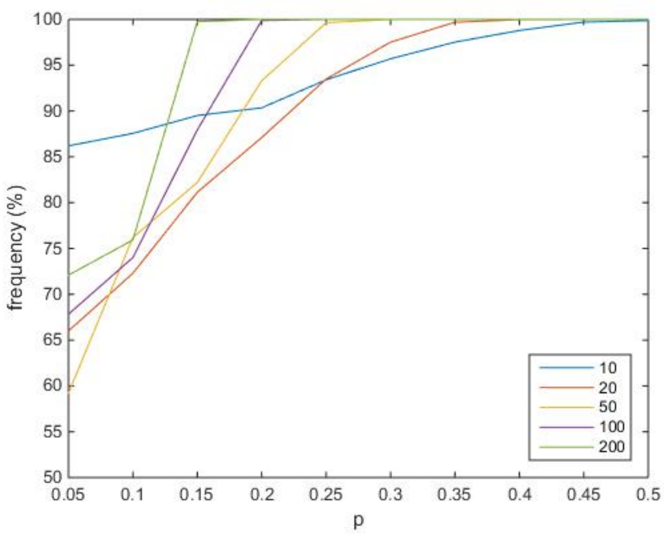


Figure 9: Frequency with which I_{deg} and I_{clos} intersect given network size and connection probability.

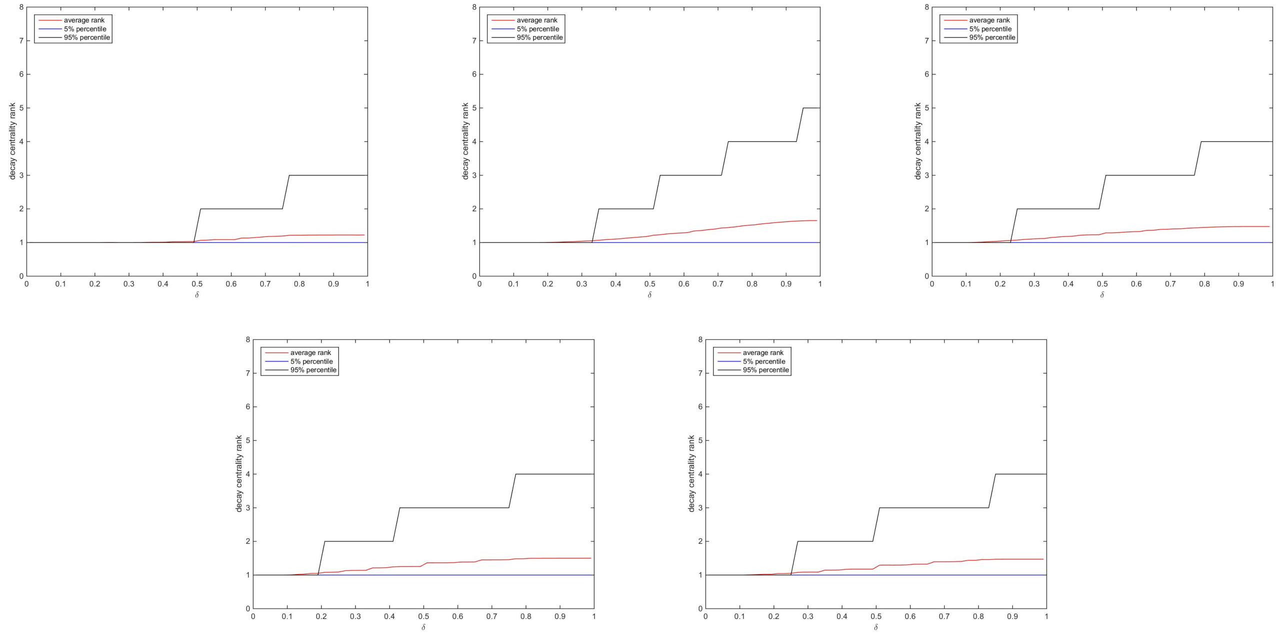


Figure 10: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum degree, including all networks. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.10$.

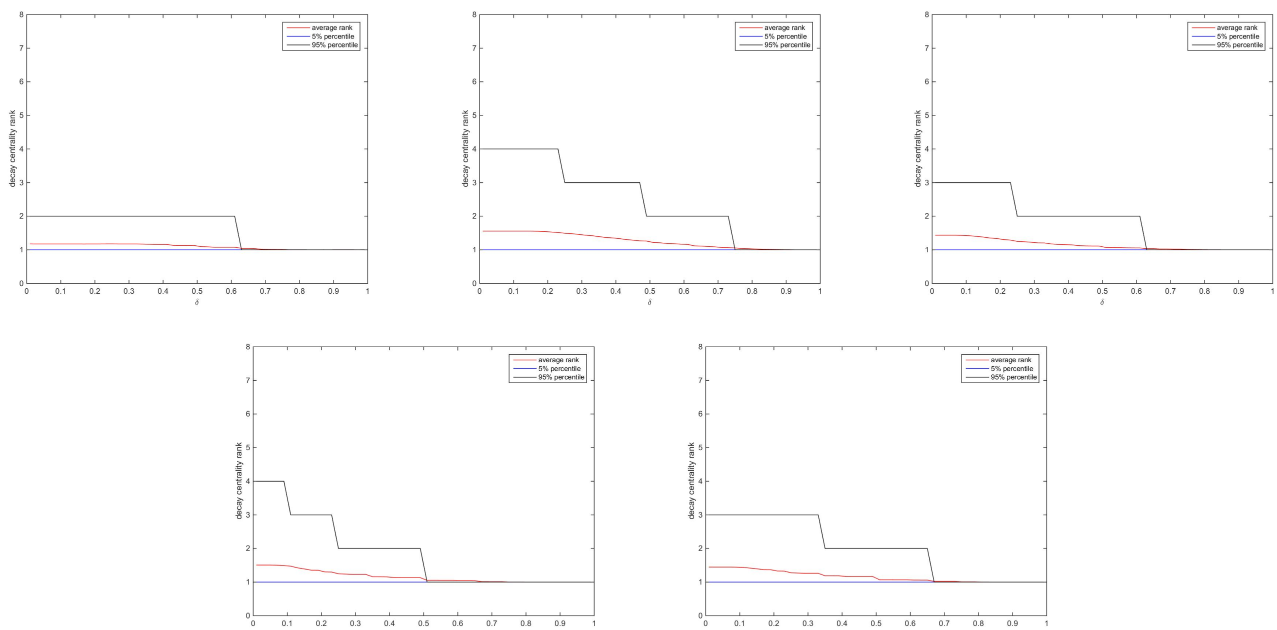


Figure 11: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum closeness, including all networks. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.10$.

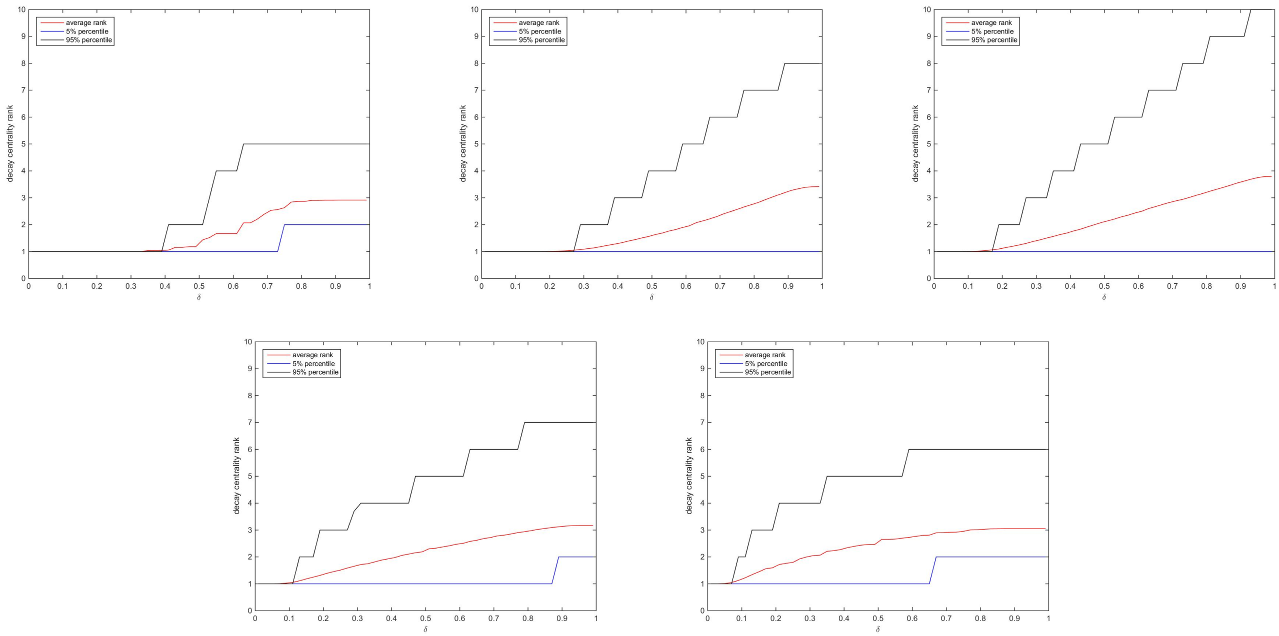


Figure 12: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum degree, excluding networks where I_{deg} and I_{clos} intersect. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.05$.

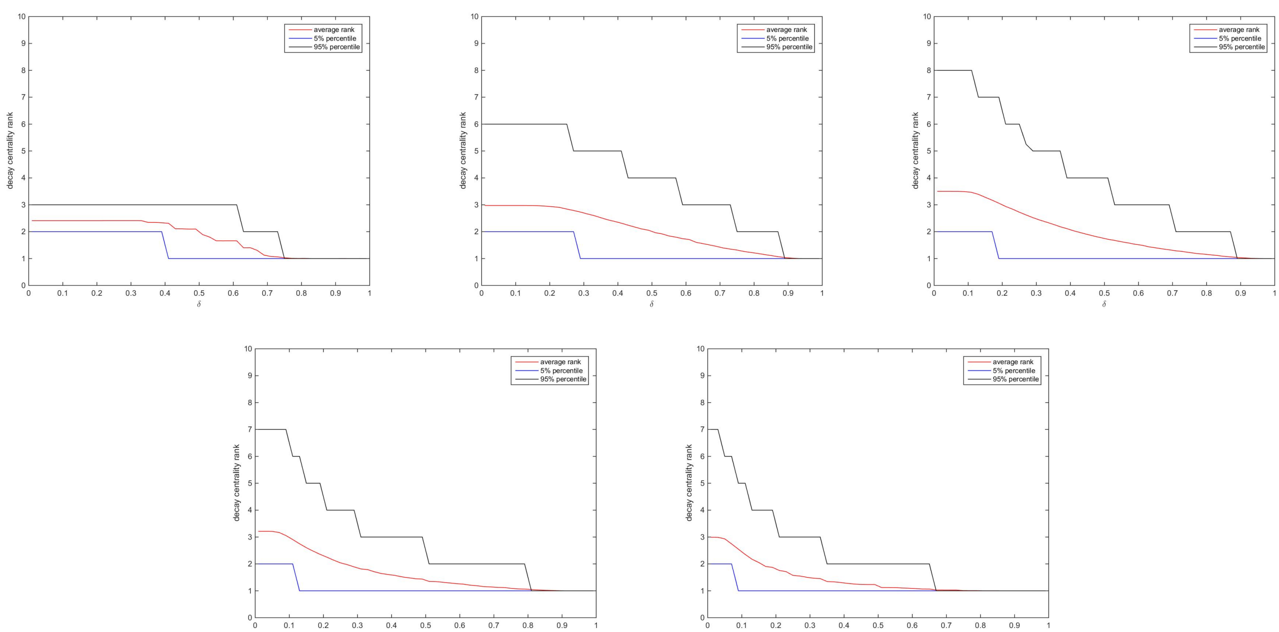


Figure 13: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum closeness, excluding networks where I_{deg} and I_{clos} intersect. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.05$.

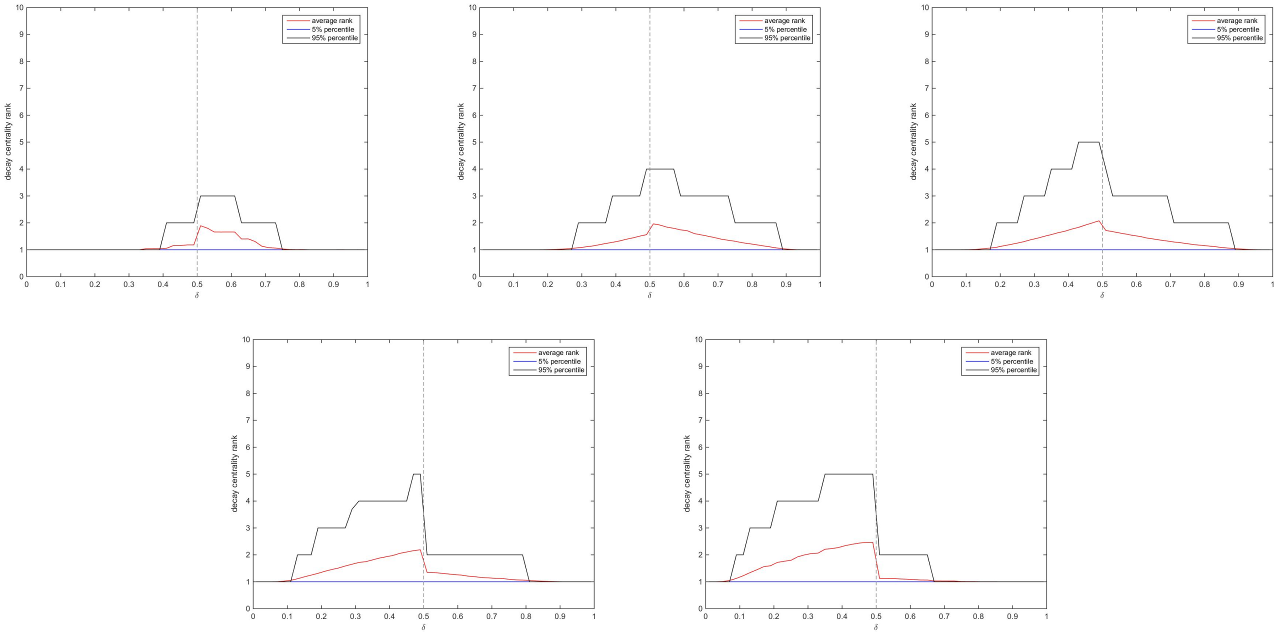


Figure 14: Average rank (red), as well as 5th (blue) and 95th (black) percentiles, of decay centrality of nodes with maximum degree for $\delta < 0.5$ and with maximum closeness for $\delta > 0.5$, excluding networks where I_{deg} and I_{clos} intersect. The five subfigures correspond to $n = 10, 20, 50, 100, 200$ from top–left to bottom–right and $p = 0.05$.

For p=0.10	n=10	n=20	n=50	n=100	n=200
Maximum Degree	-2.502***	-1.129***	-1.059***	-1.025***	-0.973***
Maximum Closeness	0.370***	0.185***	0.380***	0.384***	0.553***
Delta	15.31***	10.90***	10.72***	10.95***	11.46***
Diameter	-0.656***	0.140***	0.399***	0.386***	3.118***
Average Degree	-0.836***	-0.718***	-0.946***	-1.114***	-0.577***
St. Dev. Degree	0.904***	0.606***	-0.250***	0.0690***	0.0304***
Simulation	5.93e-07***	-1.98e-07**	-6.35e-08	-8.92e-09	3.06e-07***
Constant	12.20***	8.208***	49.01***	96.82***	232.9***
N	62200	138550	119000	129950	120350

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Coefficients of multinomial logit regression with alternatives (i) $I^{\delta}_{dc} \subseteq I_{deg}$, (ii) $I^{\delta}_{dc} \subseteq I_{clos}$ and (iii) $I^{\delta}_{dc} \nsubseteq (I_{deg} \cup I_{clos})$, for networks where I_{deg} and I_{clos} do not intersect. Alternative (i) is taken as the base and the reported coefficients correspond to (ii). Results are for all networks and $p = 0.10$.