Journal of King Saud University – Computer and Information Sciences xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

δ-Space for real-world networks: A correlation analysis of decay centrality vs. degree centrality and closeness centrality

Natarajan Meghanathan

Jackson State University, Jackson, MS 39217, USA

ARTICLE INFO

Article history:
Received 3 January 2017
Revised 9 April 2017
Accepted 21 April 2017
Available online xxxx

Keywords:
Decay centrality
Decay parameter
Closeness centrality
Degree centrality
Correlation
Real-world network graphs

ABSTRACT

We analyze a suite of 48 real-world networks and compute the decay centrality (DEC) of the vertices for the complete range of values for the decay parameter $\delta \in (0, 1)$ as well as determine the Pearson's correlation coefficient (PCC) between the DEC_δ values and degree centrality (DEG) and closeness centrality (CLC). We observe PCC(DEC_δ, DEG) to decrease with increase in δ and PCC(DEC_δ, CLC) to decrease with decrease in δ. We define the δ-$space_r$ for a real-world network with respect to the DEG, DEC, CLC correlation as the difference between the maximum and minimum δ values under which we observe a particular level of correlation (r) between the DEC, DEG and DEC, CLC metrics respectively. We show that the PCC(DEG, CLC) values for the real-world networks exhibit a very strongly positive correlation with the δ-$space_r$ values and demonstrate that one could predict the δ-$space_r$ value for a real-world network using the PCC(DEG, CLC) value for that network. We also analyze the impact of various topological measures on the δ-$space_r$ values for the real-world networks.

1. Introduction

The Decay Centrality (DEC) metric is a parameter-driven centrality metric that has not been explored much in the literature for complex network analysis. Decay centrality is a measure of the closeness of a node to the rest of the nodes in the network (Jackson, 2010). However, unlike closeness centrality (CLC) Freeman, 1979, the importance given to the geodesic distance (typically, in terms of the number of hops if the edges do not have weights) is weighted in terms of a parameter called the decay parameter δ ($0 < \delta < 1$). The formulation for computing the decay centrality of a vertex v_i for a particular value of the decay parameter δ is (Jackson, 2010): $DEC(v_i) = \sum_{v_i \neq v_j} \delta^{d(v_i, v_j)}$ where $d(v_i, v_j)$ is the distance from node v_i to node v_j. The decay parameter δ essentially controls how important is a node v_j to a node v_i ($v_i \neq v_j$) that are at a distance $d(v_i, v_j)$ from each other. If δ is smaller, the distance to the nearby nodes is weighted relatively larger than the distance to the nodes farther away. If δ is larger, the distance to every node is given almost the same importance. As a result, if δ is closer to 0, the decay centrality of the vertices is more likely to exhibit a very strong positive correlation with the degree centrality of the vertices; if δ is closer to 1, the decay centrality of the vertices is more likely to exhibit a very strong positive correlation with the closeness centrality of the vertices. We adopt the ordinal range of values proposed by Evans (1995) and consider two centrality metrics to exhibit strongly positive (s+) and very strongly positive (vs+) correlation if the Pearson's correlation coefficients (PCC) (Lay et al., 2015) computed on the basis of the values incurred for the two metrics are respectively 0.6 or above and 0.8 or above. As part of our correlation analysis, we analyze a suite of 48 real-world networks whose spectral radius ratio for node degree (Meghanathan, 2014) (a measure of variation in node degree) ranges from 1.01 to 5.51.

The motivation for our research came from the initial results of our correlation study which indicated that the Pearson's correlation coefficient PCC(DEC_δ, DEG) decreases with increase in δ from 0.01 to 0.99 and PCC(DEC_δ, CLC) decreases with decrease in δ. Because of such a trend, we came up with a hypothesis that there could exist a range of δ values (called the δ-*space*) for which we could observe the DEC_δ values to exhibit a particular level of correlation (we focus on strongly and very strongly positive levels of correlation) simultaneously with both the DEG and CLC metrics. In this pursuit, for each real-world network, we identified the maximum δ value (indicated as $\delta^{DEC,DEG}_{\max,r\equiv s+}$ or equivalently as $\delta^{DEC,DEG}_{\max,r\geqslant 0.6}$ and

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

E-mail address: natarajan.meghanathan@jsums.edu

http://dx.doi.org/10.1016/j.jksuci.2017.04.006
1319-1578/

$\delta^{DEC,DEG}_{\max,r\equiv vs+}$ or equivalently as $\delta^{DEC,DEG}_{\max,r\geqslant 0.8}$) until which PCC(DEC$_\delta$, DEG) is 0.6 or above (for strongly positive correlation) or 0.8 or above (for very strongly positive correlation), as well as identified the minimum δ value (indicated as $\delta^{DEC,CLC}_{\min,r\equiv s+}$ or equivalently as $\delta^{DEC,CLC}_{\min,r\geqslant 0.6}$ and $\delta^{DEC,CLC}_{\min,r\equiv vs+}$ or equivalently as $\delta^{DEC,CLC}_{\min,r\geqslant 0.8}$) starting from which PCC(DEC$_\delta$, CLC) continues to be 0.6 or above or 0.8 or above respectively. For real-world networks with $\delta^{DEC,CLC}_{\min,r} \leqslant \delta^{DEC,DEG}_{\max,r}$, there exists a range of δ values $\delta^{DEC,CLC}_{\min,r} \ldots \delta^{DEC,DEG}_{\max,r}$ (quantified and called the δ-*space*$_r$: $\delta^{DEC,DEG}_{\max,r} - \delta^{DEC,CLC}_{\min,r} + \in$; where $\in$ is the level of precision used for δ; for more details see Section 5) under which DEC$_\delta$ would exhibit a particular level of correlation ($r \equiv s+$ or $vs+$) with respect to both DEG and CLC. Statistically, the δ-*space*$_r$ for a real-world network with respect to a particular level of correlation (r) would correspond to the probability with which the decay centrality metric (computed for a randomly chosen value of δ) could exhibit the particular level of correlation with both the degree centrality and closeness centrality metrics. We hypothesize that the Pearson's correlation coefficient between DEG and CLC for a real-world network is likely to be very strongly correlated to the δ-*space*$_r$ values and that one could predict the δ-*space*$_r$ value for a network using the PCC(DEC, CLC) observed for that network.

The rest of the paper is organized as follows: Section 2 discusses related work. In Section 3, we review the centrality metrics (DEG, CLC and DEC) and the Pearson's correlation measure as well as explain their computation with an example graph. Section 4 first introduces the notion of δ-*space* for DEC, DEG and DEC, CLC correlation (hereafter referred to as DEG-DEC-CLC correlation) and its computation on the example graph of Section 3. Section 5 introduces the real-world networks that are analyzed in this paper. Section 6 first presents the results of correlation study involving DEC, DEG and CLC and the notion of δ-*space*$_r$. Section 6 then presents the simulation results to corroborate our hypothesis about the relationship between δ-*space*$_r$ and the Pearson's correlation coefficient for DEG and CLC. Section 6 also analyzes the impact of various topological measures on the δ-*space*$_r$ values (negative values and the highest positive value of 0.99) incurred for the real-world networks. Section 7 concludes the paper.

2. Related work

Decay centrality has not been explored much in the literature for complex network analysis. To the best of our knowledge, ours is the first work to conduct a correlation study focusing on decay centrality. Most of the work (e.g., Li et al., 2015; Meghanathan, 2015) on correlation studies (involving centrality metrics) were focused on the commonly studied centrality metrics such as the neighborhood-based degree centrality and eigenvector centrality (Bonacich, 1987) and shortest path-based betweenness centrality (Freeman, 1977) and closeness centrality. The objective of such correlation studies has been typically to identify computationally-light alternatives (like DEG and its derivatives (Meghanathan, 2017) for computationally-heavy metrics (such as EVC and BWC) for both real-world networks and simulated networks of theoretical models (Renyi, 1959; Barabasi and Albert, 1999). The focus of our paper is different from such typical correlation studies in the literature. We seek to explore the trend of change in the correlation coefficients between a parameter-driven centrality metric (whose values for a node change for different values of the decay parameter) and the degree and closeness centrality metrics whose values are not parameter-driven and remain the same for a particular network.

The most related work to our work is a recent study (Tsakas, 2016) on random networks (Renyi, 1959) for which a single threshold value of the decay parameter (referred here as δ_{thresh}) was observed to exist (for a particular operating condition) such that nodes with high degree centrality also had a high decay centrality computed for δ values less than δ_{thresh} and nodes with high closeness centrality also had a high decay centrality computed for δ values above δ_{thresh}. It was observed in Tsakas (2016) that for random networks: nodes with the largest values for degree centrality and closeness centrality are more likely to be nodes that also incur the largest values for decay centrality for almost all values of δ. In addition, nodes that had the largest decay centrality for a certain value of δ are more likely to be part of the set of nodes that had the largest degree centrality or the largest closeness centrality. The likelihood of all of the above was studied using multinomial logistic regression (Greene, 2011).

In (Dangalchev, 2006), Dangalchev proposed a variant of closeness centrality metric (to quantify the vulnerability of networks to get disconnected) that is essentially the decay centrality of the vertices computed for δ = 0.5. However, there was no correlation analysis reported between Dangalchev's closeness centrality metric and the decay centrality of the vertices for different values of δ. Most of the other works (e.g., Chatterjee and Dutta, 2016; Kang et al., 2012) on decay centrality metric have focused on exploring its suitability for diffusion in socio-economic networks with regards to selecting the seed nodes that could effectively propagate information about a product to putative customers. Nodes that are themselves central and connected to other central nodes (via direct links or shorter paths) in the network are typically preferred for such "agent" roles (Tsakas, 2016; Chatterjee and Dutta, 2016). The use of decay centrality vis-a-vis diffusion centrality (Kang et al., 2012) and eigenvector centrality (Ide et al., 2014; Banerjee et al., 2013) to identify such "agent" nodes for diffusion has been explored in the literature.

3. Review of centrality metrics and Pearson's correlation measure

The centrality metrics that are of interest in this research are degree centrality (DEG), closeness centrality (CLC) and decay centrality (DEC). In this section, we briefly review these three metrics and their computation using a running example graph as well as review the Pearson's correlation measure and its computation with respect to the DEG and CLC metrics for the running example graph.

3.1. Degree centrality

The degree centrality (DEG) of a vertex is the number of neighbors incident on the vertex. Fig. 1 illustrates the degree centrality of the vertices (listed above the vertices) in the example graph used in Sections 3 and 4. A key weakness of the degree centrality metric is that the metric can take only integer values (though, weighted degree centrality can take on any real value) and ties among vertices (with same degree) is quite common and unavoidable in network graphs of any size (in the graph of Fig. 1, we

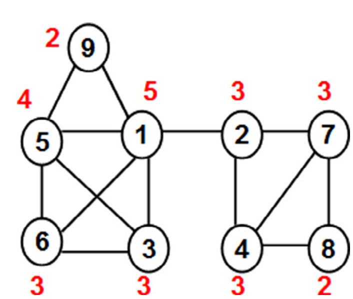

Fig. 1. Degree Centrality of the Vertices in an Example Graph.