A Framework for Evaluating Network Measures for Functional Importance

Tieyun Qian
State Key Lab of Software Eng., Wuhan Univ.
Hubei, China
qty@whu.edu.cn

Qing Li
Dept. of Comp. Sci.
City Univ. of Hong Kong
HongKong, China
itqli@cityu.edu.hk

Jaideep Srivastava
Dept. of Comp. Sci. and Eng.
Univ. of Minnesota, Twin Cities
Minnesota, USA
srivasta@cs.umn.edu

ABSTRACT

Many metrics such as degree, closeness, and PageRank have been introduced to determine the relative importance of a node within a network. The desired function of a network, however, is domain-specific. For example, the robustness can be crucial for a communication network, while efficiency is more preferred for fast spreading of advertisements in viral marketing. The information provided by some widely used measures are often conflicting under such varying demands. In this paper, we present a novel framework for evaluating network metrics regarding typical functional requirements. We also propose an analysis of five well established measures to compare their performance of ranking nodes on functional importance in a real-life network.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications

General Terms

Measurement, Performance

Keywords

Network Metrics, Functional Importance

1. INTRODUCTION

Identifying important nodes is a ubiquitous problem in graph theory and network analysis. There are a wide variety of metrics introduced in the literature for users to choose. For example, the closeness centrality reflects the mean geodesic distance between a vertex and all other vertices reachable from it, and PageRank [1] considers the importance of each page by recursively counting the number and PageRank metric of all pages that link to it. These measures have their own biases and may provide inconsistent information. Figure 1 shows a sample network, and the three top nodes of this graph under betweenness, eigenvector and pagerank metrics are listed in Table 1. It is clear from Table 1 that different measures generate substantially different orderings. Moreover, the desired function of a network varies case by case. For instance, the vulnerability under attack or random breakdown could be crucial for a communication network. In contrast, viral marketing requires influence maximization through the world-of-mouth effect among recipients. Hence the question is: which is the most appropriate metric for finding the most important nodes to a given application?

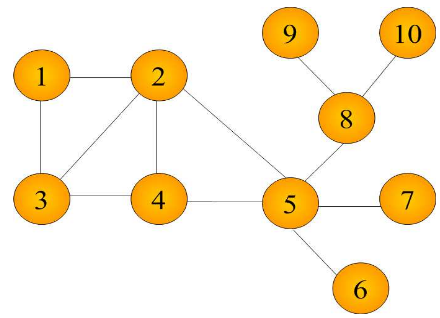

Figure 1: A sample network with 10 nodes.

Table 1: Rankings of top 3 nodes for sample network by different metrics

	Betweenness	EigenVector	PageRank
top 1	5	2	5
top 2	8	5	2
top 3	2	4	8

In this paper, we relate the raised problem with the permutation of nodes according to their functional importance of an application domain, and we present an evaluation methodology for this problem based on ranking correlation, which will allow users to compare the metrics from a domain-specific view. Given a network, firstly we characterize the function of a network as a structural feature in the graph. Secondly, we define the functional importance of a node in terms of the loss in structural property incurred by the deletion of the node. Thirdly, we evaluate the ranking correlation between the permutation generated by each metric and by the functional importance. By comparing the correlation coefficients between different metrics and functional importance, we finally get a complete vision of all metrics, and the right metric for a given application domain can be easily selected.

2. EVALUATION OF NETWORK MEASURES FOR FUNCTIONAL IMPORTANCE

In the field of network analysis, the networks are often characterized from a topological point of view, and recent advances in network theory reveal that there is a connection between network structure and function [3]. As a preliminary attempt, in this paper we focus the application of our framework on the requirement of connectivity of a graph, and we report our results on five well established network measures, i.e. degree centrality(dc), closeness centrality(cc), betweenness centrality(bc), eigenvector centrality(ec), and PageRank(pr).

2.1 Characterizing the functional requirement

First, we present our preliminary study of the functional need about connectivity. Connectivity is a basic property determining the robustness of a network. For example, research in complex network has demonstrated that intentional attacks on nodes critical in connectivity will quickly disrupt a network.

Let $G(V, E)$ be a graph, directed or undirected, with $n = |V|$ nodes and $m = |E|$ edges. Let SCC be the the largest strongly connected component in G. Following the notion described in [2], we quantify the connectivity of a network G as:

$$\Omega_c(G) =_{argmax} |SCC(G)| \tag{1}$$

2.2 Computing the functional importance

The importance of a node to the network varies with its topology. For instance, the breakdown of a leaf node in a star network will not affect communications among others. In contrast, failure of any middle nodes in a chain network renders the network inoperable. For a structural property p, which relates to a specific demand in network function, we define the importance of a vertex v as:

$$\omega_p(v) = \Omega_p(G) - \Omega_p(G_v^-), \tag{2}$$

where graph G_v^- is the remaining graph of G after the deletion of a node v.

With this definition, we can calculate the functional importance for each node in G. For example, if we investigate the connectivity property of the sample network shown in Figure 1, we can find that vertex #5 is of the largest importance, and then follows vertex #8. The removal of these two nodes will bring *6* and *3* loss in Ω_c, respectively, whereas the connectivity score only decreases *1* upon the deletion of any other nodes in the graph.

2.3 Measuring the ranking distance

Given a network measure m, there exists a permutation corresponding to the values of m for each node in V. The n raw scores are converted to ranks, thus we get a ranking list $\aleph_m$ according to the descending order of measure m. Similarly, based on Equation 2, we can also generate a standard permutation $\aleph_p$ for each desired property according to the descending order of the functional importance. Then the association between the network metric and the functional importance can be evaluated by the correlation between the two ordering lists $\aleph_m$ and $\aleph_p$.

In statistics, Kendall's tau and Spearman's rho coefficient are two widely accepted distance metrics for rankings. Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of joint observations on nodes in V from two ranking lists $\aleph_m$ and $\aleph_p$ respectively. Note that all the values of (x_i) and (y_i) should be unique. To ensure this, if the values of node i and node j are the same, then their order is decided by their node id. Kendall's tau coefficient is defined as:

$$\tau(\aleph_m, \aleph_p) = \frac{(\sum_{x_i<x_j} |y_i < y_j| + \sum_{x_i>x_j} |y_i > y_j|)}{\frac{1}{2}n(n-1)} - \frac{(\sum_{x_i<x_j} |y_i > y_j| + \sum_{x_i>x_j} |y_i < y_j|)}{\frac{1}{2}n(n-1)} \tag{3}$$

Spearman's rho coefficient is defined as:

$$\rho(\aleph_m, \aleph_p) = 1 - \frac{6\sum_i (x_i - y_i)^2}{n(n^2-1)} \tag{4}$$

3. EXPERIMENTAL EVALUATION

We conduct our experiment on a social network - the Enron corpus. After preprocessing, this dataset contains *33696* nodes and *180811* edges. Each node represents an Enron email address and each edge represents one email communication.

We first construct *5* permutations for network measures and *1* permutation for connectivity property using the proposed method. Then we compute the pairwise ranking correlation coefficient between each network metric and connectivity importance. The experimental results are listed in Table 2.

Table 2: Correlation coefficient between metrics and connectivity importance

	Kendall's τ	Spearman's ρ
($\aleph_{dc}$, $\aleph_c$)	0.4167	0.5351
($\aleph_{cc}$, $\aleph_c$)	**0.7637**	**0.8577**
($\aleph_{bc}$, $\aleph_c$)	0.6897	0.7779
($\aleph_{ec}$, $\aleph_c$)	0.5918	0.7505
($\aleph_{pr}$, $\aleph_c$)	0.1490	0.2301

From Table 2, we observe that, compared to other metrics, closeness centrality is more correlated with connectivity. A potential application of this result could be that, the targeted removal of nodes with the highest scores of closeness may prevent the diffusion of rumor in the network.

4. CONCLUSIONS

We have investigated the problem of metric selection for a specific requirement on network functions. The connectivity is used as an example of the desired function, and the problem is formalized as a problem of computing ranking coefficient between permutations according to network metric and connectivity. And, the decrease in connectivity upon removal of a node is used to characterize the functional importance of a node. Experimental results show that the metric of closeness performs well for capturing connectivity importance.

Please note that the proposed framework is quite general. It can be applied to the evaluation of many other metrics in terms of their performance on various functional importance.

5. ACKNOWLEDGEMENTS

This research was supported in part by the 111 Project (B07037), the NSFC Project (60873007, 61070011), Specialized Research Fund for the Doctoral Program of Higher Education, China (20090141120050), the NSFC Overseas, HongKong & Macao Scholars Collaborated Researching Fund (61028003).

6. REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proc. of the 7th WWW*, pages 107–117, 1998.

[2] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proc. of IMC*, pages 29–42, 2007.

[3] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.