Local search and heterogeneities in weighted complex networks

Hari P Thadakamalla
Department of Industrial Engineering
The Pennsylvania State University,
University Park, PA 16802, USA.
(001) 814 865 7859
hpt102@psu.edu

Réka Albert
Department of Physics
The Pennsylvania State University,
University Park, PA 16802, USA.
(001) 814 865 6123
ralbert@phys.psu.edu

Soundar R. T. Kumara
Department of Industrial Engineering
The Pennsylvania State University,
University Park, PA 16802, USA.
(001) 814 863 2359
skumara@psu.edu

ABSTRACT
We discuss a new algorithm based on a network measure, local betweenness centrality (LBC), for search in complex networks that are heterogeneous in node degree and edge weights. LBC search utilizes both the heterogeneities and performs the best in scale-free weighted networks. We further argue that the amount of information used by LBC search could be optimal.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Search process.* G.2.2. [**Discrete Mathematics**]: Graph Theory – *Graph algorithms, Network problems.*

General Terms
Algorithms, Design, Theory.

Keywords
Local search, scale-free networks, weighted complex networks, local betweenness centrality

1. INTRODUCTION
The macroscopic properties of many large-scale real-world networks found in communications, biology or sociology have been studied intensively in the last few years [2, 4, 7]. One of the significant findings is the presence of heterogeneity in various properties of the elements in the network. For instance, the degree distributions of many real-world networks, as well as the distribution of clustering coefficients, follow power laws which indicates a high local (node) heterogeneity [2, 4, 7] [8]. Similarly, many studies have demonstrated the heterogeneity of interaction strengths or edge-weights [3] and this heterogeneity's huge impact on network resilience, network navigation, local search, and epidemiological processes [2, 4, 7].

Here we study the effects of these heterogeneities on local search. Local search is a process in the network, where a node tries to find a path to a target node using only local information. We specifically focus on two kinds of heterogeneities, one with respect to node degree and other with respect to the edge-weights. We discuss the results obtained for a new search algorithm based on a network measure local betweenness centrality (LBC). More detailed analysis of the results could be found in [9].

NetSci2006, May 22–25, 2006, Bloomington, IN, USA.

2. LOCAL SEARCH
The problem of local search goes back to the popular experiment by Milgram [6] illustrating the existence of short paths in social networks. An even more striking observation of this study, as pointed out by Kleinberg [5], was the ability of the nodes in the network to find these paths using only local information. Later many models were proposed to explain this phenomenon [5, 10].

Local search can be considered in two kinds of networks, namely spatial and non-spatial networks. In spatial networks, each node knows the global position of the target node and at each step of the search process the node knows whether it is going away from the target or towards the target. The models in [5, 10] are examples of such kind of networks. Where as, in some networks such as peer-to-peer networks, at each step of the search process we do not know whether we are going towards the target node or away from the target node. In this paper we focus on such non-spatial networks.

2.1 Non-spatial networks
Traditional way to search in non-spatial networks is by random walk where the node that has the message passes it to a randomly chosen neighbor. This process continues until the message reaches the target node. Adamic *et al.* [1] have proposed a new algorithm (high degree search) and demonstrated that in networks with a power-law degree distribution (scale-free networks), it is more efficient than random walk search. In high degree search, the node passes the message to the neighbor that has the highest degree among the neighbors.

However, they assumed all the edges to be equivalent which does not hold in real-world networks. The edge weights may represent the cost, bandwidth, distance or power consumption associated with the process described by the edge. Then the total path length (p) obtained by a search path 1 -2 – 3 – l, is given by $p = \sum_{i=1}^{n} w_{i,i+1}$, where $w_{i,i+1}$ is the weight on the edge from node i to node $i + 1$. Therefore, if the edge weights are heterogeneous then the search path obtained by high-degree search may not be optimal. In that case, the search algorithm has to consider the information about the edge weights also. In the next section, we define a search algorithm based on the measure which depends both on the degree of the neighbor and the weight of the edge connecting to the neighbor.

3. LOCAL BETWEENNESS CENTRALITY
We assume that each node in the network knows information about the first neighbors, second neighbors and connections between them. Using this information each node (we call this

node as a root node) forms a local network and calculates the LBC of its neighbor. The LBC of a neighbor *i* (*L(i)*) is the betweenness centrality of this neighbor in the local network. Hence, it is given by

$$L(i) = \sum_{\substack{s \neq i \neq t \\ s,t \in local\ network}} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node *s* to *t*. $\sigma_{st}(i)$ is the number of these shortest paths passing through *i*. Here, a shortest path means the path over which the sum of weights is minimal. If the LBC of a neighbor is high, it implies that this node is critical in the local network. Also, LBC of a neighbor depends both on its degree and the weight of the edge connecting it to the root node [9]. In LBC search, a node passes the message to the neighbor which has the highest LBC in the corresponding local network.

In [9], we have shown that for scale-free networks with power-law exponent between 2.0 to 2.9, LBC search always performs consistently better than various other algorithms. This is mainly due to the fact that LBC search utilizes the heterogeneity in both node degree and edge weights. Further we showed that irrespective of the edge weight distributions, LBC search performed the best (see table 1). When we compare the algorithm specifically with high-degree search, we observed that as the heterogeneity in the edge weights increase, the difference between the high-degree search and LBC search increase. This implies that it is critical to consider the edge weights in the local search algorithms. Moreover, given that many real-world networks are heterogeneous in edge weights, it becomes important to consider an LBC based search rather than high degree search.

4. EXTENSIONS ON LBC SEARCH

One intriguing question about the LBC search is regarding the assumption that each node knows the information until second neighbors. Is it possible to improve the LBC search algorithm if we include information about the third neighbors, to calculate the most critical neighbor? In [9], we have shown that minimum average node weight search, which uses more information than other search algorithms, did not perform better. This implies that using more information may not lead to a better search algorithm. Initial simulations indicate that using information up to second neighbors could be optimal for LBC search. The performances of LBC search was close to the BC search, where the root node passes the message to the most critical neighbor with respect to the whole network. This implies that using more information than up to two neighbors may not help in improving the LBC search process.

5. CONCLUSIONS AND FUTURE WORK

We have discussed a new algorithm based on LBC for local search in complex networks which are heterogeneous in node degree and edge weights. LBC search uses these heterogeneities to perform better in weighted scale-free networks. As the heterogeneity in edge weights increase, the LBC search performs increasingly better than high degree search.

We further hypothesized that for LBC search using information up to second neighbors is optimal. Initial simulations on power-law networks with exponent 2.1 have shown that it could be true. In future, we are planning to test this hypothesis on networks with other exponents and other edge weight distributions. Also, we need to compute the average path lengths obtained by using information up to 3, 4 neighbors to compare with the current LBC search.

6. ACKNOWLEDGMENTS

The authors would like to acknowledge the National Science Foundation (Grant # SST 0427840) and a Sloan Research Fellowship to one of the authors (R. A.) for making this work feasible.

7. REFERENCES

[1] Adamic, L. A., Lukose, R. M., Puniyani, A. R., and Huberman, B. A., Phys. Rev. E, 64, 046135 (2001).

[2] Albert, R., Barabasi, A. L., Reviews of Modern Physics 74, p. 1 (2002).

[3] Barrat, A., Barthelemy, M., and Vespignani, A., Phys. Rev. E 70, 066149 (2004).

[4] Dorogovtsev, S. N., and Mendes, J. F. F., Adv. Phys. 51, p. 1079–1187 (2002).

[5] Kleinberg, J., Nature 406 p. 845 (2000).

[6] Milgram, S., Psychology Today 1, (1967).

[7] Newman, M. E. J., SIAM Review 45, p. 167-256, (2003).

[8] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L., Science 297, p. 1551-1555 (2002).

[9] Thadakamalla, H. P., Albert, R., and Kumara, S. R. T., Phys. Rev. E, 72, 066128 (2005).

[10] Watts, D. J., Dodds, P. S., and Newman, M. E. J., Science 296, (2002).

Table 1. Comparison of search strategies in power-law networks with exponent 2.1 and 2000 nodes with different edge weight distributions. The mean for all the edge weight distributions is 5 and the variance is shown in the brackets. The values in the table are the average distances obtained for each search strategy in these networks. The values in the brackets show the relative difference between average distances for each strategy with respect to the average distance obtained by the LBC strategy.

Search strategy	Beta (2.3)	Uniform (8.3)	Exponential (25)	Power-law (4653)
Random walk	1107.7 (202%)	1097.7 (241%)	1108.7(272%)	1011.2(344%)
Minimum edge weight	704.5 (92%)	414.7 (29%)	319.0 (7%)	358.5 (44%)
Highest degree	380.0 (4%)	368.4 (14%)	375.8 (26%)	395.0 (59%)
Minimum average node weight	1228.7 (235%)	788.2 (145%)	605.4 (103%)	466.2 (88%)
Highest LBC	**366.3**	**322.3**	**298.1**	**247.8**