Computing Top-k Closeness Centrality in Fully-dynamic Graphs

Master's Thesis

Patrick Bisenius

1640015

At the Department of Informatics
Institute of Theoretical Informatics

Reviewer: Jun. Prof. Dr. Henning Meyerhenke
Advisor: Elisabetta Bergamini

31st May 2016

Abstract

Zusammenfassung

Contents

List of Figures

List of Tables

1. Introduction

In network analysis, one fundamental problem is finding the most important and central nodes. Several measures have been proposed to ascertain the importance of nodes within their graphs. These measures can be used to find the most important people in social networks or central places in street networks.

Closeness centrality

Closeness centrality is based on the intuition that a node is important if its distance to other nodes in the graph is small. There are different definitions of closeness centrality that are applicable in different contexts. For strongly connected graphs, one can simply compute the sum of the distances from one node to all the other nodes. However, this approach leads to problems on disconnected graphs. It is not initially clear how to treat nodes which are not connected by a path in the graph. Simply assuming an infinite (or arbitrarily large) distance would completely distort the resulting closeness values of affected nodes. One possible solution for this problem is to only take into account the distances of reachable nodes, then scaling the result with the number of reachable nodes. Another approach is called *harmonic centrality*. Instead of summing all distances and then computing the inverse, the harmonic centrality is obtained by computing the sum of the inverse distances between nodes. Disconnected node pairs do not contribute to the total sum of inverse distances.

Computing the closeness centrality of a node in an unweighted graph requires a complete breadth-first search (BFS), and a complete run of Dijkstra's algorithm [Dij59] with weighted graphs. In order to compute the closeness centrality of all nodes in the graph, one such search is required from all nodes. The computational effort for this is quite high, especially on large real-world networks. Moreover, this effort cannot be avoided if the application requires an exact ranking of all nodes by their closeness centrality.

> Cite survey of different centrality measures

For some applications, however, it is enough to compute a list of the k most central nodes. This problem is called *Top-k closeness*. Limiting the problem to the k most central nodes can decrease the required computational effort significantly. It is only necessary to compute the exact closeness centrality of the k most central nodes to rank them. For all other nodes, it is sufficient to obtain an upper bound for their closeness that is smaller than the exact closeness of the k-th most central node.

Dynamic Top-k closeness centrality

In some cases, it is enough to compute closeness centralities only once because the underlying graph is static. Now consider a social network which constantly adds new users, which is effectively a node insertion in the underlying graph, and existing users befriend other users, which is an edge insertion. Terminating a friendship in the social network corresponds to an edge removal.

Each modification of a graph affects at least the closeness centralities of the directly affected nodes, that is, the nodes incident to a newly inserted or removed edge. It is also possible that there are new shortest paths between pairs of nodes that use the newly inserted edge. Analogously, removing an edge might increase the distance between node pairs because there was only one shortest path between them and it contained the removed edge. A simple strategy to get the new closeness centralities of each node is to re-run the static algorithm on the modified graph, ignoring any information collected by previous runs of the algorithm.

Group closeness

The concept of closeness centralities for single nodes can be extended to groups of nodes. The distance between a node v and a group S is defined as the smallest distance between v and any node of the group. The problem to find a group of size k such that the total distance of all nodes in the graph to the group is minimal is called the *maximum closeness centrality group identification* (MCGI) problem by Chen et al. in [CWW16]. Since the problem is shown to be NP-hard, no efficient exact algorithm exists at this point. However, there are approximative greedy algorithms [CWW16, ZLTG14] for the problem. Bergamini et al. improve on the work in [CWW16] by reducing the memory requirements and total number of operations .

> Cite un-
> published
> work

Contributions

This thesis contributes a dynamic algorithm for Top-k closeness that handles both edge insertions and edge removals. It is based on the static algorithm first proposed by Borassi et al for complex networks in [BCM15], and the additional optimizations for street networks proposed by Bergamini et al. in [BBC+16]. Our algorithm re-uses information obtained by an initial run of the static algorithm and tries to skip the re-computation of closeness centralities for nodes that are unaffected by modifications of the graph. In some cases, even the upper bounds for the closeness centralities of affected nodes can be updated with little computational effort.

We also contribute an algorithm to update the group of nodes with the highest group closeness after an edge insertion. It is based on Bergamini and colleagues' improved version of the algorithm by Chen et al. The basic idea is to verify whether the choices of the greedy algorithm are still valid on the modified graph with as little computational expense as possible.

2. Conclusion

3. Declaration

Ich versichere hiermit wahrheitsgemäß, die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, den 31. Mai 2015   _____

Patrick Bisenius

Bibliography

[BBC+16] Elisabetta Bergamini, Michele Borassi, Pierluigi Crescenzi, Andrea Marino, and Henning Meyerhenke. Computing top-k closeness centrality faster in unweighted graphs. In *2016 Proceedings of the Eighteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 68–80. SIAM, 2016.

[BCM15] Michele Borassi, Pierluigi Crescenzi, and Andrea Marino. Fast and simple computation of top-k closeness centralities. *arXiv preprint arXiv:1507.01490*, 2015.

[CWW16] Chen Chen, Wei Wang, and Xiaoyang Wang. *Efficient Maximum Closeness Centrality Group Identification*, pages 43–55. Springer International Publishing, Cham, 2016.

[Dij59] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[ZLTG14] Junzhou Zhao, John Lui, Don Towsley, and Xiaohong Guan. Measuring and maximizing group closeness centrality over disk-resident graphs. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 689–694. ACM, 2014.

Todo list