Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science

ELSEVIER Electronic Notes in Theoretical Computer Science 322 (2016) 153–168

www.elsevier.com/locate/entcs

On the Maximum Betweenness Improvement Problem

Gianlorenzo D'Angelo[1] Lorenzo Severini[1] Yllka Velaj[1]

Gran Sasso Science Institute (GSSI)
Viale F. Crispi 7, L'Aquila, Italy.

Abstract

The betweenness is a well-known measure of centrality of a node in a network. We consider the problem of determining how much a node can increase its betweenness centrality by creating a limited amount of new edges incident to it. If the graph is directed, this problem does not admit a polynomial-time approximation scheme (unless $P = NP$) and a simple greedy approximation algorithm guarantees an almost tight approximation ratio [10].
In this paper we focus on the undirected graph case: we show that also in this case the problem does not admit a polynomial-time approximation scheme (unless $P = NP$). Moreover, we show that, differently from the directed case, the greedy algorithm can have an unbounded approximation ratio. In order to test the practical performance of the greedy algorithm, we experimentally measured its efficiency in term of ranking improvement, comparing it with another algorithm that simply adds edges to the nodes that have highest betweenness. Our experiments show that the greedy algorithm adds only few edges in order to increase the betweenness of a node and to reach the top positions in the ranking. Moreover, the greedy algorithm outperforms the second approach.

Keywords: Betweenness centrality, approximation algorithms, graph augmentation

1 Introduction

One of the main goals of network analysis is that of determining the most important nodes in a given complex network. Several measures of importance have been explicitly formalized in the literature to try to quantify how much a node is important (or "*central*"). The way of defining such so called *centrality measures* depends on the particular feature of the network that the measure wants to capture.

One of the most popular measures of importance is the *betweenness centrality* (see, for example, [8]). It intuitively quantifies how much a node controls the information flow between all pairs of nodes in a graph. More formally, the betweenness centrality of a given node v is the portion of shortest paths that pass through v

[1] Email: gianlorenzo.dangelo@gssi.infn.it, lorenzo.severini@gssi.infn.it, yllka.velaj@gssi.infn.it

http://dx.doi.org/10.1016/j.entcs.2016.03.011
1571-0661/

over all the possible shortest paths between all pairs of nodes different from v. Having high betweenness centrality can have positive impact on the node itself. Let us consider a network where there are messages passing along the edges by using shortest paths (e.g. information is a social network or packets in the Internet). Then, the number of messages passing through a given node v is, to some extent, proportional to the number of shortest paths passing through v. Hence, a node with high betweenness has an high probability of receiving an high number of messages and hence it is central.

Computing betweenness centrality of a node or its ranking in the network can be done in polynomial time but requires $O(nm)$ time [9] on unweighted graphs which is clearly infeasible for huge networks. Therefore, several randomized or approximation algorithms have been proposed [7,14,22]

Besides computing the betweenness centrality, another interesting problem is that of *increasing* the betweenness centrality of a given node. Increasing the centrality of a node can clearly have positive consequences on the node itself. For example, in the field of transportation network analysis, the betweenness centrality seems to be positively related to the efficiency of an airport (see [17] where a network of 57 European airports has been analyzed). On the other hand, in many complex networks, a node can decide to connect itself to some other nodes in the network. Therefore, in this paper, we focus on the problem of maximizing the betweenness centrality of a node by adding a limited number of edges incident to it. More specifically, we consider the problem of efficiently determining, for a given node v, the set of k edges incident to v that, when added to the original graph, allows v to increase as much as possible its betweenness centrality (and as a consequence its ranking according to this measure).

1.1 Related work

The problem of increasing the centrality of a node in a network has been studied for several centrality measures different from betweenness, i.e. page-rank [4,20], eccentricity [11], stress and some measures related to the number of paths passing through a given node [15], closeness [10], and average distance [18].

Regarding betweenness centrality, if the network is directed and the arcs to be added are all directed towards node v, it has been shown that the problem does not admit a polynomial-time approximation scheme (unless $P = NP$) and that a simple greedy approximation algorithm exhibits an almost tight approximation ratio [10]. In detail, the problem cannot be approximated within a factor of $1 - \frac{1}{2e}$ and the greedy algorithm guarantees an approximation factor of $(1 - \frac{1}{e})$. The main part of the proof of the approximation ratio of the greedy algorithm consists in proving that the objective function is monotone and submodular which is not true for the undirected case.