Computers & Operations Research 43 (2014) 261–270

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

A derandomized approximation algorithm for the critical node detection problem

Mario Ventresca*, Dionne Aleman

Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada M5S 3G8

ARTICLE INFO

Available online 27 September 2013

Keywords:
Critical node detection problem
Randomized rounding
Complex network

ABSTRACT

In this paper we propose an efficient approximation algorithm for determining solutions to the critical node detection problem (CNDP) on unweighted and undirected graphs. Given a user-defined number of vertices $k>0$, the problem is to determine which k nodes to remove such as to minimize pairwise connectivity in the induced subgraph. We present a simple, yet powerful, algorithm that is derived from a randomized rounding of the relaxed linear programming solution to the CNDP. We prove that the expected solution quality obtained by the linear-time algorithm is bounded by a constant. To highlight the algorithm quality four common complex network models are utilized, in addition to four real-world networks.

1. Introduction

In general, critical node detection involves the identification of a typically small subset of vertices $R \subseteq V$ from graph $G=(V,E)$. These vertices are somehow important, as specified by the objective function being optimized. Commonly, the critical nodes are intended to be removed from G with the goal of minimizing or maximizing some properties of the induced subgraph. Some examples of critical node problems include important junctions in cell-signalling or protein–protein networks [31,9], highly influential individuals in anti-terrorism networks [20], targeted vaccination for pandemic prevention [10,23,7] or keys to decipher brain functionality [18]. In some contexts, an accurate mathematical definition for a critical node does not yet exist, particularly for highly complex systems such as the brain [30].

This paper focuses on the particular problem formulation of [5], termed the Critical Node Detection Problem. It is assumed that the given network is connected, and the user has defined a desired upper limit on the number of critical nodes k to be detected and removed from the graph. The goal is to discover the subset of vertices whose removal leaves the residual graph with minimum pairwise connectivity. As a consequence, the optimal network will be the one that is fragmented into many components of approximately equal size. This problem was shown to be $\mathcal{NP}-$complete [5].

Our main contributions revolve around an efficient randomized algorithm for estimating solutions to the CNDP. The approach follows the algorithmic and analytical framework for rounding a fractional solution to an integral one through a process known as randomized rounding [27]. Many problems have been solved using this general framework but CNDP is not among them. We show a constant approximation factor for the number of vertices and objective value, dependent on the minimum rounding threshold. Through the use of four common random graph models and real-world networks, we examine the practical performance of the approach in the context of different connectivity patterns in network structure.

Recently, the CNDP has been attracting attention. The restricted case where G is a tree structure with non-unit edge costs has been shown to be $\mathcal{NP}-$complete [11], and the authors also provide an efficient dynamic programming approach. Slight variants to the CNDP formulation are also solved using this dynamic programming solution [2]. An integer linear programming model with a non-polynomial number of constraints is given and branch and cut algorithms are proposed in [12].

Heuristics without guaranteed approximation bounds have also been studied. The original CNDP work devises a heuristic based on the maximum-independent set problem as a starting point for local search, repeating the process until a desired termination criteria is reached [5]. The algorithm is tested on a limited number of network structures with promising results. Two stochastic search algorithms are proposed in [33] that allow for significantly larger networks to be solved within reasonable time and without significant resources. These simulated annealing and population-based incremental learning algorithms are shown to yield very promising results for common network models of up to 5000 vertices. However, all of these works lack any approximation bound on solution quality or the number of vertices selected.

The critical node detection problem resembles other common graph partitioning problems in the literature. The minimum

* Corresponding author.
E-mail addresses: mario.ventresca@utoronto.ca, mventresca@purdue.edu (M. Ventresca), aleman@mie.utoronto.ca (D. Aleman).

0305-0548/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cor.2013.09.012

multi-cut problem aims to separate a set of source–sink pairs by removing a subset of minimum weighted edges from the given graph. An $\mathcal{O}(\log n)$ approximation for general graphs has been proven for this problem [17]. Another well studied graph partitioning problem is the k-cut problem. Given an undirected weighted graph, the goal is to find a minimum cost set of edges that separates the graph into at least k connected components. In [29], an $\mathcal{O}(2-2/k)$–approximation algorithm has been proposed. Classical multi-way cut, multi-cut and k-cut problems have been extended to include budget constraints that act to limit the number of edges or vertices that can be cut from the graph have been studied in [14]. The authors also propose the problem of maximizing the number of connected components within a desired budget.

An approximation algorithm with $\mathcal{O}(\sqrt{n})$ bound was presented for the sparsest cut, edge expansion, balanced separator, and graph conductance problems in [3,4]. Each of these problems are based on graph partitioning, while also minimizing the size of the interface between the resulting components. Their method is based on semi-definite relaxation to these problems in concert with expander flows, and has influenced much subsequent research.

Numerous other optimization problems similar to CNDP have also been defined recently. The minimum contamination problem has the goal of minimizing the expected size of a network contamination by removing a subset of edges of at most a given cardinality [19]. The authors identify a variant of the problem in which the goal is to minimize the proportion of vertices in the largest resulting induced subgraph. Their bi-criteria algorithm achieves an $\mathcal{O}(1+\varepsilon,(1+\varepsilon)/\varepsilon(\log n))$ approximation. A game-theoretic analysis is conducted in [10] that requires a solution to a generalization of the sum-of-squares partitioning problem [7]. Their solution is based on the region growing framework of [16,21] and an $\mathcal{O}(\log n)$ approximation is proven.

The remainder of this paper is organized as follows. Section 2 formally describes the critical node detection problem. The proposed randomized rounding-based approach is described in Section 3, where its approximation abilities are also proven. A set of random graph and complex network models are described in Section 4, and the results of experimentation are given in Section 5. Conclusion and suggested directions for future work are given in Section 6.

2. The critical node detection problem

In an undirected graph $G=(V,E)$ where V is the set of vertices and E is the set of edges, find $R\subseteq V$ such that $|R|\leq k$ and the residual graph $G\backslash R$ has minimum pairwise connectivity. Formally, the problem can be summarized as

$$\operatorname*{argmin}_{R\subseteq V} \sum_{u,v\in(V\backslash R)} x_{uv} \tag{1}$$

where,

$$x_{uv}=\begin{cases}1 & \text{if } u,v \text{ are in the same component of } G(V\backslash R)\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

Defining

$$y_u=\begin{cases}1 & \text{if } u \text{ is deleted in the solution}\\ 0 & \text{otherwise}\end{cases} \tag{3}$$

then, the problem can be expressed as the following integer program [6]:

$$\text{minimize} \quad \sum_{u,v\in V} x_{uv} \tag{4}$$

$$\text{subject to} \quad x_{uv}+y_u+y_v\geq 1 \quad \forall (u,v)\in E \tag{5}$$

$$x_{uv}+x_{vw}-x_{wu}\leq 1 \quad \forall u,v,w\in V \tag{6}$$

$$x_{uv}-x_{vw}+x_{wu}\leq 1 \quad \forall u,v,w\in V \tag{7}$$

$$-x_{uv}+x_{vw}+x_{wu}\leq 1 \quad \forall u,v,w\in V \tag{8}$$

$$\sum_{v\in V} y_v\leq k \tag{9}$$

$$x_{uv}\in\{0,1\} \quad \forall u,v\in V \tag{10}$$

$$y_u\in\{0,1\} \quad \forall u\in V \tag{11}$$

The first constraint (Eq. (5)) enforces the separation of vertices into different components by permitting edge x_{uv} to be deleted if either or both vertices u and v are deleted. Eqs. (6)–(8) are essentially the enforcement of a triangle inequality whereby if vertices u, v and v, w are connected in the residual graph, then u and w must also be connected. The final constraint (Eq. (9)) ensures that no more than k vertices are removed from G.

3. Proposed approach

Before presenting our proposed algorithm and analysis we outline some important preliminary definitions. We will use the following notation. Let $(\mathbf{x}^*,\mathbf{y}^*)$ with objective function value z^* be the optimal solution to CNDP, and $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ with objective function value $\hat{z}$ be the optimal solution to the LP relaxation of CNDP. The rounded solution obtained by the proposed algorithm is $(\overline{\mathbf{x}},\overline{\mathbf{y}})$, and it has objective function value $\overline{z}$.

3.1. Randomized rounding approach

We show that the randomized rounding-based method is capable of attaining approximation ratios within a tunable constant factor with respect to the lower rounding threshold of $\hat{\mathbf{y}}$.

3.1.1. Preliminaries

Our analysis requires one simple concept that establishes the relationship between the expected value of the sum of random variables (we assume binary, but the definition can be generalized) and the expected values of each variable.

When utilizing randomness in approximation algorithms it is important to understand the probability of violating a constraint, which can be accomplished using the Chernoff Bound.

Definition 1 (*Chernoff Bound*). Let $X_1,\ldots,X_n$ be n independent random variables from the set $\{0,1\}$, not necessarily identically distributed. Then for $X=\sum_{i=1}^{n}X_i$ and $\mu=\mathbb{E}[X]\leq U$ and $\delta>0$,

$$\Pr[X\geq(1+\delta)U]<\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{U} \tag{12}$$

Fig. 1 depicts the Chernoff Bound considering different U for $\delta=[0,1]$. As would be expected, larger U values yield a sharper decrease in probability. That is, the probability approaches zero faster as U increases.

3.1.2. The algorithm

Randomized rounding views the values of $\hat{x}_{uv},\hat{y}_v\in[0,1]$ as independent probabilities, and determines which $\hat{y}_v$ to retain in the solution set $\mathcal{S}$ probabilistically. The naive approach is to utilize