Theoretical Computer Science 554 (2014) 275–282

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Approximating the minimum independent dominating set in perturbed graphs

Weitian Tong, Randy Goebel, Guohui Lin *

Department of Computing Science, University of Alberta, Edmonton, Alberta T6G 2E8, Canada

ARTICLE INFO

Available online 14 November 2013

Keywords:
Independent set
Independent dominating set
Dominating set
Approximation algorithm
Perturbed graph
Smooth analysis

ABSTRACT

We investigate the minimum independent dominating set in perturbed graphs $\mathfrak{g}(G, p)$ of input graph $G = (V, E)$, obtained by negating the existence of edges independently with a probability $p > 0$. The minimum independent dominating set (MIDS) problem does not admit a polynomial running time approximation algorithm with worst-case performance ratio of $n^{1-\epsilon}$ for any $\epsilon > 0$. We prove that the size of the minimum independent dominating set in $\mathfrak{g}(G, p)$, denoted as $i(\mathfrak{g}(G, p))$, is asymptotically almost surely in $\Theta(\log |V|)$. Furthermore, we show that the probability of $i(\mathfrak{g}(G, p)) \geqslant \sqrt{\frac{4|V|}{p}}$ is no more than $2^{-|V|}$, and present a simple greedy algorithm of proven worst-case performance ratio $\sqrt{\frac{4|V|}{p}}$ and with polynomial expected running time.

1. Introduction

An *independent set* in a graph $G = (V, E)$ is a subset of vertices that are pairwise non-adjacent to each other. The independence number of G, denoted by $\alpha(G)$, is the size of a maximum independent set in G. One close notion to independent set is the *dominating set*, which refers to a subset of vertices such that every vertex of the graph is either in the subset or is adjacent to some vertex in the subset. In fact, an independent set becomes a dominating set if and only if it is maximal. The size of a minimum independent dominating set of G is denoted by $i(G)$, while the domination number of G, or the size of a minimum dominating set of G, is denoted by $\gamma(G)$. It follows that $\gamma(G) \leqslant i(G) \leqslant \alpha(G)$.

Another related notion is the (vertex) *coloring* of G, in which two adjacent vertices must be colored differently. Note that any subset of vertices colored the same in a coloring of G is necessarily an independent set. The *chromatic number* $\chi(G)$ of G is the minimum number of colors in a coloring of G. Clearly, $\alpha(G) \cdot \chi(G) \geqslant |V|$.

The independence number $\alpha(G)$ and the domination number $\gamma(G)$ (and the chromatic number $\chi(G)$) have received numerous studies due to their central roles in graph theory and theoretical computer science. Their exact values are NP-hard to compute [4], and hard to approximate. Raz and Safra showed that the domination number cannot be approximated within $(1 - \epsilon)\log|V|$ for any fixed $\epsilon > 0$, unless NP $\subset$ DTIME$(|V|^{\log\log|V|})$ [9,3]; Zuckerman showed that neither the independence number nor the chromatic number can be approximated within $|V|^{1-\epsilon}$ for any fixed $\epsilon > 0$, unless P = NP [14]; for $i(G)$, Halldórsson proved that it is also hard to approximate within $|V|^{1-\epsilon}$ for any fixed $\epsilon > 0$, unless NP $\subset$ DTIME$(2^{o(|V|)})$ [5].

The above inapproximability results are for the worst case. For analyzing the average case performance of approximation algorithms, a probability distribution of the input graphs must be assumed and the most widely used distribution of graphs on n vertices is the random graph $G(n, p)$, which is a graph on n vertices, and each edge is chosen to be an edge of G independently and with a probability p, where $0 \leqslant p = p(n) \leqslant 1$. A graph property holds *asymptotically almost surely* (a.a.s.)

* Corresponding author.
E-mail addresses: weitian@ualberta.ca (W. Tong), rgoebel@ualberta.ca (R. Goebel), guohui@ualberta.ca (G. Lin).

0304-3975/$ – see front matter
http://dx.doi.org/10.1016/j.tcs.2013.11.010

in $G(n,p)$ if the probability that a graph drawn according to the distribution $G(n,p)$ has the property tends to 1 as n tends to infinity [1].

Let $\mathbb{L}n = \log_{1/(1-p)} n$. Bollobás [2] and Łuczak [7] showed that a.a.s. $\chi(G(n,p)) = (1+o(1))n/\mathbb{L}n$ for a constant p and $\chi(G(n,p)) = (1+o(1))np/(2\ln(np))$ for $c/n \leqslant p(n) \leqslant o(1)$ where c is a constant. It follows from these results that a.a.s. $\alpha(G(n,p)) = (1-o(1))\mathbb{L}n$ for a constant p and $\alpha(G(n,p)) = (1-o(1))2\ln(np)/p$ for $C/n \leqslant p \leqslant o(1)$. The greedy algorithm, which colors vertices of $G(n,p)$ one by one and picks each time the first available color for a current vertex, is known to produce a.a.s. in $G(n,p)$ with $p \geqslant n^{\epsilon-1}$ a coloring whose number of colors is larger than the $\chi(G(n,p))$ by only a constant factor (see Ch. 11 of the monograph of Bollobás [1]). Hence the largest color class produced by the greedy algorithm is a.a.s. smaller than $\alpha(G(n,p))$ only by a constant factor.

For the domination number $\gamma(G(n,p))$, Wieland and Godbole showed that a.a.s. it is equal to either $\lfloor \mathbb{L}n - \mathbb{L}((\mathbb{L}n)(\ln n))\rfloor + 1$ or $\lfloor \mathbb{L}n - \mathbb{L}((\mathbb{L}n)(\ln n))\rfloor + 2$, for a constant p or a suitable function $p = p(n)$ [13]. It follows that a.a.s. $i(G(n,p)) \geqslant \lfloor \mathbb{L}n - \mathbb{L}((\mathbb{L}n)(\ln n))\rfloor + 1$. Recently, Wang proved for $i(G(n,p))$ an a.a.s. upper bound of $\lfloor \mathbb{L}n - \mathbb{L}((\mathbb{L}n)(\ln n))\rfloor + k + 1$, where $k = \max\{1, \mathbb{L}2\}$ [12].

Average case performance analysis of an approximation algorithm over random instances could be inconclusive, because the random instances usually have very special properties that distinguish them from real-world instances. For instance, for a constant p, the random graph $G(n,p)$ is expected to be dense. On the other hand, an approximation algorithm performs very well on most random instances can fail miserably on some "hard" instances. For instance, it has been shown by Kučera [6] that for any fixed $\epsilon > 0$ there exists a graph G on n vertices for which, even after a random permutation of vertices, the greedy algorithm produces a.a.s. a coloring using at least $n/\log_2 n$ colors, while $\chi(G) \leqslant n^{\epsilon}$. To overcome this, Spielman and Teng [10] introduced the smoothed analysis. This new analysis is a hybrid of the worst case and the average-case analyses, and it inherits the advantages of both, by measuring the expected performance of the algorithm under slight random perturbations of the worst-case inputs. If the smoothed complexity of an algorithm is low, then it is unlikely that the algorithm will take long time to solve practical instances whose data are subject to slight noises and imprecision.

Formally, let A be the algorithm we want to analyze and Q be the quality measurement. Without loss of generality, we assume the larger Q the worse the algorithm A performs, such as Q being the running time. Given an instance x, $Q(A,x)$ measures the performance of algorithm A on x. Let U denote the set of all instances. Let r be a random noise, σ be the perturbation parameter measuring the strength of noise, and $\sigma \cdot |x| \cdot r$ be the perturbation added to x (the magnitude of the perturbation is related to the magnitude of input). The expected performance of algorithm A in the small neighborhood of instance x, defined by the above perturbation model, is $E_r[Q(A, x + \sigma \cdot |x| \cdot r)]$. The smoothed performance measure of algorithm A under Q is taken as the worst of the expected performances over all instances of U. We observe that if the smoothed performance measure of algorithm A under Q is good with some relatively small σ and some reasonable random model for r, then it is unlikely algorithm A would perform very bad in real-world applications under quality measure Q, because real-world instances are often subject to a slight amount of noise, especially when they are obtained from measurements of real-world phenomena. A classic example is the Simplex method for linear programming. Simplex method is a very practical algorithm, but it has exponential running time in the worst case. Spielman and Teng [10] had shown that Simplex method has polynomial smoothed running time, which explains the above phenomenon perfectly. Though the smoothed analysis concept was originally introduced for the complexity of algorithms, we extend its idea to depict the essential properties of computational problems.

In this paper, we study the approximability of the minimum independent dominating set (MIDS) problem under the smoothed analysis, and we present a simple deterministic greedy algorithm beating the strong inapproximability bound of $n^{1-\epsilon}$, with polynomial expected running time. The MIDS problem, and the closely related independent set and dominating set problems, have important applications in wireless networks, and have been studied extensively in the literature. Our probabilistic model is the smoothed extension of random graph $G(n,p)$ (also called semi-random graphs in [8]), proposed by Spielman and Teng [11]: given a graph $G = (V, E)$, we define its perturbed graph $\mathfrak{g}(G,p)$ by negating the existence of edges independently with a probability of $p > 0$. That is, $\mathfrak{g}(G,p)$ has the same vertex set V as G but it contains edge e with probability p_e, where $p_e = 1-p$ if $e \in E$ or otherwise $p_e = p$. For sufficiently large p, Manthey and Plociennik presented an algorithm approximating the independence number $\alpha(\mathfrak{g}(G,p))$ with a worst-case performance ratio $O(\sqrt{np})$ and with polynomial expected running time [8].

Re-define $\mathbb{L}n = \log_{1/p} n$. We first prove on $\gamma(\mathfrak{g}(G,p))$, and thus on $i(\mathfrak{g}(G,p))$ as well, an a.a.s. lower bound of $\mathbb{L}n - \mathbb{L}((\mathbb{L}n)(\ln n))$ if $p > \frac{1}{n}$. We then prove on $\alpha(\mathfrak{g}(G,p))$, and thus on $i(\mathfrak{g}(G,p))$ as well, an a.a.s. upper bound of $2\ln n/p$ if $p < \frac{1}{2}$ or $2\ln n/(1-p)$ otherwise. Given that the a.a.s. values of $\alpha(G(n,p))$ and $i(G(n,p))$ in random graph $G(n,p)$, our upper bound comes with no big surprise; nevertheless, our upper bound is derived by a direct counting process which might be interesting by itself. Furthermore, we extend our counting techniques to prove on $i(\mathfrak{g}(G,p))$ a tail bound that, when $4\ln^2 n/n < p \leqslant \frac{1}{2}$, $\Pr[i(\mathfrak{g}(G,p)) \geqslant \sqrt{4n/p}] \leqslant 2^{-n}$. We then present a simple greedy algorithm to approximate $i(\mathfrak{g}(G,p))$, and prove that its worst-case performance ratio is $\sqrt{4n/p}$ and its expected running time is polynomial.

2. A.a.s. bounds on the independent domination number

We need the following several facts.