Journal of Discrete Algorithms 34 (2015) 128–136

Contents lists available at ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

On the kernel size of clique cover reductions for random intersection graphs

Tobias Friedrich [a,*], Christian Hercher [b]

[a] Hasso Plattner Institute, Potsdam, Germany
[b] Friedrich-Schiller-Universität Jena, Germany

ARTICLE INFO

Article history:
Received 30 November 2014
Received in revised form 21 April 2015
Accepted 27 May 2015
Available online 9 June 2015

Keywords:
Parameterized algorithms
Clique cover
Kernelization
Average case
Random intersection graphs
Erdős–Rényi graphs

ABSTRACT

Covering all edges of a graph by a minimum number of cliques is a well known NP-hard problem. For the parameter k being the maximal number of cliques to be used, the problem becomes fixed parameter tractable. However, assuming the Exponential Time Hypothesis, there is no kernel of subexponential size in the worst-case.

We study the average kernel size for random intersection graphs with n vertices, edge probability p, and clique covers of size k. We consider the well-known set of reduction rules of Gramm, Guo, Hüffner, and Niedermeier (2009) [17] and show that with high probability they reduce the graph completely if p is bounded away from 1 and $k < c \log n$ for some constant $c > 0$. This shows that for large probabilistic graph classes like random intersection graphs the expected kernel size can be substantially smaller than the known exponential worst-case bounds.

1. Introduction

In the past few years, many results on upper and lower bounds on the kernel sizes of parameterized problems have been shown. Nearly all of them only consider the worst-case. We take a different route and consider the average kernel size for a probabilistic graph model and present a tight characterization depending on the graph density.

Our study problem is the NP-hard problem Clique Cover, which aims to covering the edges of a graph with a minimum number of cliques. The problem arises in studies of the interaction of entities in real-world networks [18] and in protein–protein interaction networks [3]. It also has applications in compiler optimization [29], computational geometry [1], and computational statistics [16,27]. In different domains the problem has been described by varying names, other variants are Keyword Conflict [22], Covering by Cliques [15], and Intersection Graph Basis [14].

Formal definition of the problem. A clique in an undirected graph $G = (V, E)$ is a subgraph, where any two vertices are connected by an edge. For an undirected graph G and an integer $k \geq 0$, the decision problem (Edge) Clique Cover answers yes iff there is a set of at most k cliques $\{G_1, G_2, \dots, G_k\}$ in G such that each edge in G has both its endpoints in at least one of the cliques G_i.

* Corresponding author.
E-mail address: surname@hpi.de (T. Friedrich).

http://dx.doi.org/10.1016/j.jda.2015.05.014
1570-8667/

Previous results for special graph classes. Clique Cover has been studied from many perspectives. It remains NP-hard even when restricting to planar input graphs [6] or graphs with maximum degree 6 [21]. However there are polynomial-time algorithms when restricting to graphs with maximum degree 5 [21], chordal graphs [24] or line graphs [26]. Back in the general setting Lund and Yannakakis have shown that the optimization variant of Clique Cover is not approximable within a factor of $|V|^{\varepsilon}$ for some $\varepsilon > 0$ (unless $\mathsf{P} = \mathsf{NP}$) [23]. In fact, it remains APX-hard even when restricting to biconnected graphs of maximum degree 7 [20]. Hence there is no hope of good and fast approximation algorithms.

Previous parametrized results. Another popular approach on handling the problem is studying its parametrized version. Fixing the output size as a parameter k, Gyárfás [19] presents a set of simple reduction rules. Successively applying such reduction rules leads to a smaller instance and/or parameter. This process is called a *kernelization* as it results in a so-called kernel, which is a yes-instance iff the original input was a yes-instance, too. The reduction rules considered in [19] leave a kernel of size (number of vertices) at most 2^k. Computing this kernelization can be done in $\mathrm{poly}(n)$ time. As this kernel size does not depend on $|V| = n$, Clique Cover with this parametrization is therefore in FPT, see Gramm et al. [17]. Performing these reduction rules, we get an algorithm with runtime $\mathcal{O}\left(n^4 + f\left(2^k\right)\right)$. Recent results of Cygan et al. [7,8] show that there is no guaranteed polynomial sized kernel if not $\mathsf{NP} \subseteq \mathsf{coNP/poly}$ and there is no subexponential sized kernel, unless the Exponential Time Hypothesis fails. In fact, unless $\mathsf{P} = \mathsf{NP}$, the kernel has to be of exponential size in worst case [28]. Thus we do not expect a $2^{2^{o(k)}} \cdot \mathrm{poly}(n)$ time algorithm – a double exponential runtime is required. These results are not depending on the set of reduction rules used for kernelization. Moreover, kernel sizes (whether polynomial or not) are of great interest because they lead to a finer structural analysis of FPT problems, see Bodlaender et al. [4].

Previous average case results. Clique Cover has also been studied on random graphs. Bollobás et al. [5] give lower and upper bounds on the number of cliques required to cover the entire graph, which hold with high probability (meaning with probability $1 - o(1)$ as n tends to infinity, w.h.p.) for random graphs (Erdős–Rényi) with edge probability $p = \frac{1}{2}$. This is equivalent to a uniform distribution over all graphs with set of vertices $|V| = \{1, 2, \ldots, n\}$. This result was improved to $\Theta(n^2/\log^2 n)$ for all constant p with $0 < p < 1$ by Frieze and Reed [13]. Also on random intersection graphs the problem was studied. Behrisch and Taraz [2] give algorithms, which find w.h.p. a clique cover of minimal size in polynomial time for certain probability functions, if the underlying feature set (and thus expected size of the clique cover) is n^{α} for a constant $0 < \alpha < 1$.

Our results. We study the kernel size in a probabilistic graph model called random intersection graphs similar to Erdős–Rényi that ensures that the drawn graphs are coverable by at most k cliques (for details see Section 3). First we study a set of reduction rules (defined in Section 4), which intuitively arises from the definition of the graph model. In Section 5 we show that for sparse graphs (edge probabilities p decreasing at least polynomially to 0) we get w.h.p. polynomial sized kernels with respect to the parameter k. Therefore the worst case instances there are rare. If the graphs are dense (i.e., edge probabilities p not decreasing at least polynomially to 0, e.g., constant p with $0 < p < 1$) the set of reduction rules w.h.p. reduces the instances only to kernels with exponential size. We give a full characterization of the (w.h.p.) kernel sizes with respect to the edge probability in Theorem 5.1. The situation of stepwise decreasing kernel size is visualized in Fig. 1 on page 132. At probabilities $p(n)$ which are asymptotically $n^{-2/i}$, with $1 < i \leq k$ (more precisely, $\lim_{n\to\infty} \frac{\log p}{\log n} = -\frac{2}{i} =: -a$), the behavior of the kernel size cannot be described with full certainty by the parameter a alone and only an interval can be stated.

In a second step we generalize the previous set of reduction rules and study the original set of reduction rules given by Gramm et al. [17]. For this we can use the same techniques to show in Section 6 that for an edge probability of $p < 1 - r^k$ for a constant $0 < r < 1$ and $k < c(r) \log n$ for a positive constant $c(r)$ depending on r this set of rules reduces the graph w.h.p. completely. For random intersection graphs the reduction rules of Gramm et al. [17] can therefore be seen as an 'optimal set' of reduction rules for kernelization.

2. Erdős–Rényi random graphs

The well known observation that searching for a minimal clique cover is NP-complete is a pure worst case statement. For a better understanding, we want to study probabilistic graph models and perform an average analysis. One such popular model was constructed by Erdős and Rényi [9] and is defined as follows.

Definition 2.1 *(Erdős–Rényi graphs).* Let n be a positive integer and $p(n)$ a function $\mathbb{N} \to [0, 1]$. Then $G(n, p)$ is the probability space on the set of graphs with n vertices, where each edge is in the graph with probability $p = p(n)$ and every two edges are drawn independently.

This model was used in the analyses of Clique Cover by Bollobás et al. [5] and Frieze and Reed [13]. For a constant edge probability $0 < p < 1$ w.h.p., the number of cliques needed to cover the graph tends to infinity. Consequently, if we fix the number k of such cliques, most instances will be no-instances. This can be stated more rigorously as follows.