Discrete Applied Mathematics 184 (2015) 130–138

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

Parameterized clique on inhomogeneous random graphs ☆

Tobias Friedrich, Anton Krohmer *

Institute for Mathematics and Computer Science, Friedrich-Schiller-Universität Jena, D-07743, Germany

ARTICLE INFO

Article history:
Received 19 May 2014
Received in revised form 8 October 2014
Accepted 15 October 2014
Available online 7 November 2014

Keywords:
Clique
Scale-free network
Power law graph
Inhomogeneous random graph

ABSTRACT

Finding cliques in graphs is a classical problem which is in general NP-hard and parameterized intractable. In typical applications like social networks or biological networks, however, the considered graphs are scale-free, i.e., their degree sequence follows a power law. Their specific structure can be algorithmically exploited and makes it possible to solve clique much more efficiently. We prove that on inhomogeneous random graphs with n nodes and power law exponent β, cliques of size k can be found in time $\mathcal{O}(n)$ for $\beta \geq 3$ and in time $\mathcal{O}(ne^{k^4})$ for $2 < \beta < 3$.

1. Introduction

The proliferation of scale-free network models has recently been propelled by experimental findings: Many real-world graphs like social networks, electricity maps, biological networks, co-author graphs, sex graphs, etc. have been found to exhibit similar properties even though they stem from vastly different fields and sources [8,19,21]. They all have a power law degree distribution, meaning that the number of vertices with degree k is proportional to $k^{-\beta}$, where β is a constant intrinsic to the network.

Following this discovery, an abundance of different theoretical models for these networks has been proposed, among which the probably most well known are the Preferential Attachment [3] and the Inhomogeneous Random Graphs [23]. There has been a significant body of research devoted to finding more similarities between these networks (e.g. low diameter, large clustering); there has been little work, however, on how to exploit these properties for algorithmic problems. In fact, many such problems like k-Clique that are believed to be intractable were originally inspired by scale-free networks—even though at the time the term "clique" was coined, the notion of scale-free networks did not yet exist [17].

It is therefore natural to investigate these real-world inspired problems on power law graphs. In this paper, we present different algorithms for finding fully connected subgraphs (cliques) of size k in inhomogeneous random graphs. It has been well studied that on general inputs, k-Clique is NP-complete [15]. An efficient polynomial time algorithm is therefore unlikely to exist. In fact, the combinatorial explosion of this problem is in both parameters n and k: A runtime $\mathcal{O}(f(k)\cdot \text{poly}(n))$ that scales arbitrarily in the problem parameter k, but only polynomially in the input size n is considered to be unachievable, as this problem is also $W[1]$-complete [11].

Recent findings, however, suggest that things look differently on scale-free networks: Although there can be cliques of polynomial size [4], Janson, Łuczak, and Norros [14] proved formally that one can retrieve a $1 - o(1)$ approximation of the largest clique with high probability when given the underlying theoretical model of the graph. Eppstein and Strash [10] furthered this intuition experimentally by enumerating all maximal cliques on several data sets in feasible time. In this paper,

☆ This paper improves its conference version [13].

* Corresponding author. Tel.: +49 3641 946325; fax: +49 3641 946322.
E-mail address: anton.krohmer@uni-jena.de (A. Krohmer).

http://dx.doi.org/10.1016/j.dam.2014.10.018

we show that there exist *exact* algorithms that only need the graph as input and no further metainformation. In particular – in contrast to previous theoretical results – these algorithms do not require access to the generally unknown node weights.

Results. The behavior of scale-free networks depends significantly on the exponent β of the power law degree distribution. In the case $\beta > 3$ (e.g. sexual contacts, citations or electronic circuits [8,20]), the expected maximal size of a clique is constant [5,14]. This implies that large cliques are unlikely, but does not imply a fast algorithm that always answers correctly. The difficulty is certifying a negative answer. We prove the following theorem.

Theorem 1. *The k-*Clique *problem can be solved in expected time $\mathcal{O}(n)$ on inhomogeneous random graphs with power law exponent $\beta \geq 3$.*

Throughout this paper, all asymptotic notation is given in both parameters n and k, but hides constants like average degree δ and power law exponent β. Consequently, the runtime of our algorithm does not depend on k and is therefore asymptotically optimal. Our algorithm is deterministic and always returns the correct answer. Moreover, it does not use any underlying information of the model (e.g. weights). Note that this theorem implies that k-Clique, which is NP-complete in general, in this setting becomes avgP, which is the average-case analog of P [16]. The best result so far was $\mathcal{O}(n^4)$ by Janson et al. [14] and an algorithm with $\mathcal{O}(n^2)$ runtime by the conference version [13] of this paper. Note that an application of a Markov bound yields the following high probability bound on the runtime.

Corollary 2. *Let $f(n)$ be any function such that grows asymptotically slower than n, i.e. $f(n) \in \omega(n)$. Then, the k-*Clique *problem can be solved in time $f(n)$ on inhomogeneous random graphs with power law $\beta \geq 3$ with high probability.*

On the other hand, many scale-free networks (e.g. co-actors, protein interactions, internet, peer-to-peer [20]) have a power law exponent β with $2 < \beta < 3$. In this case, the expected maximal size of a clique diverges [5,14] and there exists a *core*. The core is a subgraph that has a diameter of $\mathcal{O}(\log\log n)$ and contains a dense Erdős–Rényi graph [7,23]. As this is a known hard problem, we cannot expect similarly good results as for $\beta \geq 3$. We prove the following theorem.

Theorem 3. *The k-*Clique *problem can be solved in time $\mathcal{O}(n \exp(k^4))$ with overwhelming[1] probability on inhomogeneous random graphs with power law exponent $2 < \beta < 3$.*

While in general k-Clique is not believed to be parameterized tractable, i.e. in FPT, the above theorem shows that in this setting k-Clique is typically parameterized tractable, i.e. in typFPT, which is an average-case analog of FPT as defined in [12]. We are confident that this result extends to exponents $\beta \leq 2$, but as those networks exhibit significantly different properties and are rare in reality compared to the above mentioned cases, we did not investigate them formally.

2. Preliminaries

In order to achieve high general validity, we use the inhomogeneous random graph model of van der Hofstad [23], which generalizes the models of Chung–Lu [7,1,2] and Norros–Reittu [22] as well as the generalized random graphs. The model has two adjustable parameters: the exponent of the scale-free network β and the average degree δ. Depending on these two parameters, each node i has a weight w_i. This determines the edge probability $p_{ij} := \Pr[\{i,j\} \in E]$, which intuitively should be set proportional to $w_i w_j$.

Weights w_i. A simple way to fix the weights would be for example $w_i = \delta(n/i)^{\frac{1}{\beta-1}}$. However, we aim for a more general setting and proceed differently. Given the weights w_i, we can use the empirical complementary cumulative distribution function (CCDF) $F_n(w) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[w_i \geq w]$. This gives us $F_n(w) = \Pr[W \geq w]$, where W is a random variable chosen uniformly from the weights $w_1, \ldots, w_n$. Instead of fixing w_i, it is now easier to start from $F_n(w)$ and assume the following.

Definition 4 (*Power-Law Weights*)**.** We say that an empirical CCDF $F_n(w)$ follows the *power law with exponent β*, if there exist two positive constants α_1, α_2 such that

$$\alpha_1 w^{-\beta+1} \leq F_n(w) \leq \alpha_2 w^{-\beta+1}.$$

Then, we require the weights $w_1, \ldots, w_n$ to have the empirical CCDF $F_n(w)$. Following van der Hofstad [23], we moreover require that the empirical CCDF F_n satisfies the following properties.

Definition 5 (*Regularity Conditions for Node Weights*)**.**

(1) *Weak convergence of node weights.* There exists a function F such that $\lim_{n\to\infty} F_n(x) = F(x)$.
(2) *Convergence of average node weight.* Let W_n and W have distribution functions F_n and F, respectively. Then, it holds that $\lim_{n\to\infty} \mathbb{E}[W_n] = \mathbb{E}[W]$. Furthermore, $\mathbb{E}[W] > 0$.

[1] We use the terms *high probability* for probability $1 - o(1)$, *negligible probability* for probability $1/f(n)$, and *overwhelming probability* for probability $1 - 1/f(n)$, where $f(n)$ is any superpolynomially increasing function.