Information and Computation 256 (2017) 212–225

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

The complexity of degree anonymization by graph contractions ☆

Nimrod Talmon [*,1], Sepp Hartung

Institut für Softwaretechnik und Theoretische Informatik, Technische Universität Berlin, D-10587 Berlin, Germany

A R T I C L E   I N F O

Article history:
Received 18 June 2015
Available online 10 July 2017

Keywords:
Computational complexity
Parameterized complexity
Fixed-parameter tractability
Privacy
Data publishing
Graph modifications
Degree-constrained editing

A B S T R A C T

We study the computational complexity of k-anonymizing a given graph by as few graph contractions as possible. A graph is said to be k-anonymous if for every vertex in it, there are at least $k-1$ other vertices with exactly the same degree. The general degree anonymization problem is motivated by applications in privacy-preserving data publishing, and was studied to some extent for various graph operations (most notable operations being edge addition, edge deletion, vertex addition, and vertex deletion). We complement this line of research by studying the computational complexity of degree anonymization by graph contractions. We consider several variants of graph contractions, which are operations of interest, for example, in the contexts of social networks and clustering algorithms. We show that the problem of degree anonymization by graph contractions is NP-hard even for some very restricted inputs, and identify some fixed-parameter tractable cases.

1. Introduction

Motivated by concerns of data privacy in social networks, Clarkson et al. [9] introduced the general degree anonymization problem, defined as follows. Given an input graph G and a set of allowed operations O, the task is to transform G into a k-anonymous graph by performing as few operations from O as possible; a graph is said to be k-anonymous if for every vertex in it, there are at least $k-1$ other vertices with exactly the same degree.

The general idea is that an adversary might know the degrees of the vertices, thus, even if we hide the identifying information from each vertex (for example, the name of the person corresponding to the vertex), the adversary might be able to de-anonymize some vertices. The task is to preserve the privacy of the entities comprising the underlying social network, while not modifying the graph too much. Here, we allow performing as few graph contractions as possible, in order to k-anonymize the input graph. That is, the set O of our allowed operations is (several variants of) graph contractions.

Studying graph contractions in the context of degree anonymization is interesting for several reasons. First, some variants of contractions can preserve original properties of the input graph (for example, connectivity). Second, vertex contraction (where also non-adjacent vertices can be contracted), is the inverse operation of vertex cleaving (as defined by Oxley [24, Chapter 3]), which was studied in the context of degree anonymization by Bredereck et al. [5] (there, called *vertex cloning*).

☆ A preliminary short version of this work has been presented at the 12th Annual Conference on Theory and Applications of Models of Computation (TAMC 2015) [17].

* Corresponding author.
E-mail addresses: nimrodtalmon77@gmail.com (N. Talmon), sepp.hartung@tu-berlin.de (S. Hartung).

1 Was supported by DFG, Research Training Group “Methods for Discrete Structures” (GRK 1408).

http://dx.doi.org/10.1016/j.ic.2017.07.007

We mention also the relation of graph contractions to communities detection in social networks and to clustering (see, for example, Delling et al. [10]).

1.1. Related work

The general problem of degree anonymization has been studied, both theoretically and practically, for several graph modification operations. Clarkson et al. [9], which introduced the problem, gave a heuristic, based on dynamic programming, for k-anonymizing an input graph by adding as few edges as possible, while Lu et al. [21] proposed another heuristic for the same problem, based on a greedy approach. Hartung et al. [18] studied the parameterized complexity of degree anonymization by edge additions. The main result there is a fixed-parameter tractable algorithm with respect to the maximum degree in the input graph. The problem of k-anonymizing an input graph by performing as few edge modification as possible, that is, edge switchings, edge deletions, and edge additions, was studied by Casas-Roma et al. [7], which gave a heuristic solution for it. Anonymization by adding new vertices was studied by Chester et al. [8], which gave a non-optimal algorithm with some approximation guarantees for the problem. The computational complexity as well as the parameterized complexity of this problem was studied by Bredereck et al. [5], which developed fixed-parameter tractable algorithms for some special cases. Bredereck et al. [4] considered degree anonymization by vertex deletion, again from a parameterized complexity point-of-view.

This paper can be seen as complementing the above line of research by considering graph contractions, as a natural graph modification operation. Specifically, we study the parameterized complexity of degree anonymization by graph contractions.

This paper also complements research done on the following problem. Given an input graph $G = (V, E)$ and a family $\mathcal{F}$ of graphs, the task is to find a subset of edges $E' \subseteq E$ of minimum size, such that after contracting the edges in E', G would be in the family $\mathcal{F}$. Indeed, in our case, $\mathcal{F}$ is the family of all k-anonymous graphs. Asano and Hirata [1] defined a set of conditions on $\mathcal{F}$, which is sufficient for the NP-hardness of this problem, while others studied the complexity of this problem for different some specific graph classes (acting as $\mathcal{F}$). Planar graphs were considered by Golovach et al. [15], bipartite graphs were considered by Heggernes et al. [19], paths were considered by Heggernes et al. [19], and trees were considered by Guillemot and Marx [16]. Most relevant to our work is the work done by Belmonte et al. [3], which considered the parameterized complexity of making a graph to respect some degree constraints, such as being d-regular, by performing as few as possible graph contractions. This work is of particular interest, as the concept of k-anonymity is a generalization of the notion of regularity (in particular, it is clear that a graph is n-anonymous if and only if it is regular).

This paper is a full version of a paper presented at the 12th Annual Conference on Theory and Applications of Models of Computation [17].

2. Preliminaries

We assume familiarity with standard notions regarding algorithms, computational complexity, and graph theory. For a positive integer z, we denote $\{1, \ldots, z\}$ by $[z]$.

2.1. Parameterized complexity

Parameterized complexity and parameterized algorithms provide one way to cope with NP-hardness. The general idea is that, while some problems are presumably intractable, that is, provably NP-hard, they might still be tractable as long as some parameters are kept small. Thus, in parameterized complexity theory, we measure the complexity of a given problem with respect to both the input size and a particular *parameter* of the problem. Typical parameters for degree anonymization problems include the solution size, the anonymity level, and the maximum degree. We say that a parameterized problem is *fixed-parameter tractable* (is in FPT) if there is an algorithm that, given an input instance I with parameter k, solves the problem in time $g(k)|I|^{O(1)}$, where g is some computable function and $|I|$ is the length of the encoding of I. There is also a hierarchy of hardness classes for parameterized problems, of which the two most important levels are formed by the classes W[1] and W[2]. The most convenient way of defining these classes is through an appropriate reduction notion and their complete problems. Specifically, we say that a parameterized problem A reduces to a parameterized problem B if there are two computable functions, g and g', with the following properties: given an instance I of A with parameter k, $g(I)$ outputs in FPT time an instance I' of B with parameter $k' \leq g'(k)$, such that I is a "yes"-instance of A if and only if I' is a "yes"-instance of B. In other words, g is a standard many-one reduction from A to B, which is allowed to run in FPT time, but such that the parameter of the output instance must be upper-bounded by a function of the input instance's parameter.

The class W[1] is defined as the class of problems that the Clique problem [12] can be parameterically reduced to, and the class W[2] is defined as the class of problems that the Set Cover problem [12] can be parameterically reduce to, where both problems are parameterized by the solution size.

Clique
Input: An undirected graph $G = (V(G), E(G))$ and an integer h.
Question: Does there exist a set H of h vertices such that there is an edge between each pair of vertices from H?