Theoretical Computer Science 607 (2015) 16–34

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

The complexity of degree anonymization by vertex addition [☆]

Robert Bredereck [1], Vincent Froese [2], Sepp Hartung, André Nichterlein, Rolf Niedermeier, Nimrod Talmon [*,3]

Institut für Softwaretechnik und Theoretische Informatik, TU Berlin, Germany

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form 30 June 2015
Accepted 4 July 2015
Available online 10 July 2015

Keywords:
Graph modification
Degree-constrained editing
NP-hardness
Parameterized complexity
Fixed-parameter tractability
Kernelization

ABSTRACT

Motivated by applications in privacy-preserving data publishing, we study the problem of making an undirected graph k-anonymous by adding few vertices (together with some incident edges). That is, after adding these "dummy vertices", for every vertex degree d appearing in the resulting graph, there shall be at least k vertices with degree d. We explore three variants of vertex addition (justified by real-world considerations) and study their (parameterized) computational complexity. We derive mostly intractability results, even for very restricted cases (including trees and bounded-degree graphs) but also obtain some encouraging fixed-parameter tractability results.

1. Introduction

This work is concerned with making an undirected graph *k-anonymous*, that is, transforming it (at "low cost") into a graph where every vertex degree occurs either zero or at least k times. This graph modification scenario is motivated by data privacy requests in social networks; it focuses on degree-based attacks on identity disclosure of network nodes. Liu and Terzi [26] (also see [10] for an extended version) pioneered degree-based identity anonymization in graphs, which recently developed into a very active research field [2,4,7–9,20,22,21,27,31] with theoretical as well as practical work. So far, the most frequently studied models have relied on edge modifications (allowing either only edge addition or both edge addition and deletion) [7,10,22,26,27,31]. We are aware of one theoretical work [4] that considers vertex deletion as modification operation; there mostly computational hardness results have been achieved. We are also aware of another theoretical work [20] that considered graph contractions as modification operation. Chester et al. [8] started to investigate vertex addition, and we follow this last line of research.

There is good reason why vertex addition may be preferred to other graph modification operations when aiming at k-anonymity. The central point here is the "utility" of the anonymized graph. For instance, in the edge addition scenario, inserting a new edge always destroys distance properties between vertices and indeed may introduce undesirable and mis-

☆ A preliminary version of this work has been presented at the 2014 International Conference on Algorithmic Aspects of Information and Management (AAIM 2014), Vancouver, Canada, July 2014 [5].

* Corresponding author.
E-mail addresses: robert.bredereck@tu-berlin.de (R. Bredereck), vincent.froese@tu-berlin.de (V. Froese), sepp.hartung@tu-berlin.de (S. Hartung), andre.nichterlein@tu-berlin.de (A. Nichterlein), rolf.niedermeier@tu-berlin.de (R. Niedermeier), nimrodtalmon77@gmail.com (N. Talmon).

1 Supported by the DFG, project PAWS (NI 369/10).
2 Supported by the DFG project DAMM (NI 369/13).
3 Supported by DFG, RTG "Methods for Discrete Structures" (GRK 1408).

http://dx.doi.org/10.1016/j.tcs.2015.07.004

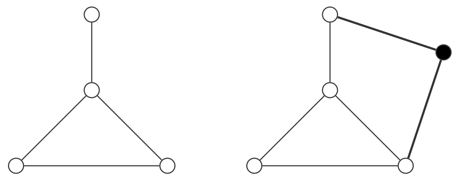

Fig. 1. Degree Anonymization (VA): The input graph on the left is not 2-anonymous. The graph on the right is a possible solution for anonymizing by vertex addition. The new vertex (black) is arbitrarily connected to some other vertices.

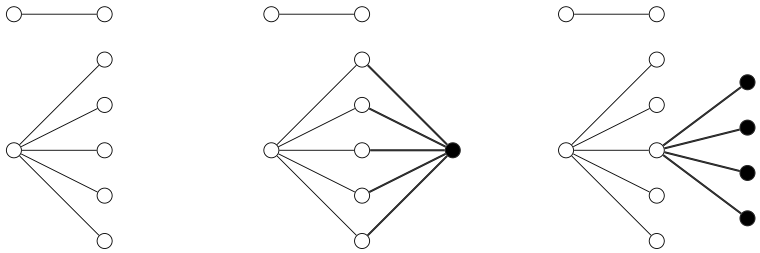

Fig. 2. An example showing the difference between our model and the model of Chester et al. [8]. In this example, $k = 2$, and $X = V$. The input graph on the left is not 2-anonymous. The graph in the middle is a minimum solution for our model, using only one additional vertex, while the graph on the right is a minimum solution for the model of Chester et al. [8], using four addition vertices.

leading "fake relations". Adding new vertices and connecting them to some of the vertices of the original graph could avoid this problem and gives at least a better chance to preserve essential graph properties such as connectivity, shortest paths, or diameter. For example, adding a new vertex and connecting it to only one existing vertex does not change distances between any existing vertices. Chester et al. [8] provided a more thorough discussion of the benefits of vertex addition.

The basic decision version of the vertex addition problem we study is as follows (see Fig. 1 for an example).

Degree Anonymization (VA)

Input: A simple undirected graph $G = (V, E)$ and $k, t \in \mathbb{N}$.
Question: Is there a k-anonymous graph $G' = (V \cup V', E \cup E')$ such that $|V'| \leq t$ and $E' \subseteq \{\{u, v\} \subseteq V \cup V' \mid u \in V' \vee v \in V'\}$?

It is important to note that Chester et al. [8] studied a slightly different model, with decisive consequences for computational complexity: Their model gets as input a simple undirected graph $G = (V, E)$, integers t and k, and also a vertex subset $X \subseteq V$, and the task is to k-anonymize the degree sequence (that is, the vertex degrees sorted in ascending order) of $X \cup V'$ *and* the degree sequence of X. On the contrary, we consider the simpler model where $X = V$, and we require to k-anonymize only the degree sequence of $X \cup V'$ $(= V \cup V')$.

To better understand the difference in the models, consider the example depicted in Fig. 2: In this example, the minimum solution size for the model of Chester et al. [8] is four, while the minimum solution size for our model is two. The crucial difference is that in the solution for our model, the new vertex and the old vertex of degree five *together* will form a 2-anonymized "block". Nevertheless, we conjecture that our results (both positive and negative) extend to the model of Chester et al. [8].

Our contributions Partially answering an open question of Chester et al. [8], we show that Degree Anonymization (VA) is weakly NP-hard for a compact encoding of the input. Based on this encoding, we provide several (fixed-parameter) tractability results, exploiting parameterizations by the maximum vertex degree of the input graph, the number of added vertices, and the maximum number of (implicitly) added new edges. The tractability result regarding the parameter maximum number of (implicitly) added new edges is given by developing a *bikernelization* [1,24] to a closely related number problem. This is one of our most technical results. Moreover, we also study variants of Degree Anonymization (VA) where we only allow "cloned" vertices to be added[4] (that is, identical copies of existing vertices with exactly the same neighborhood; we denote this problem variant by Degree Anonymization (VC); see Fig. 3 for an example) or we explicitly demand the preservation of some desirable features of the input graph (expressed by Π) such as distance properties (this problem variant is denoted by Π-Preserving Degree Anonymization (VA)). For these practically interesting variants, we prove computational hardness already for very restricted cases (for instance even on trees). Table 1 surveys most of our results, and some open questions.

[4] The cloning operation is frequently studied in the context of privacy, see, for example, the work by Bilge et al. [3]. It is also studied in other contexts, for example, in social choice [12].