A Refined Complexity Analysis of Degree Anonymization on Graphs

Sepp Hartung[1], André Nichterlein[1], Rolf Niedermeier[1], and Ondřej Suchý[2]

[1] Institut für Softwaretechnik und Theoretische Informatik, TU Berlin
{sepp.hartung, andre.nichterlein, rolf.niedermeier}@tu-berlin.de
[2] Faculty of Information Technology, Czech Technical University in Prague
ondrej.suchy@fit.cvut.cz

Abstract. Motivated by a strongly growing interest in graph anonymization in the data mining and databases communities over the last five years, we study the NP-hard problem of making a graph k-anonymous by adding as few edges as possible. Herein, a graph is k-anonymous if for every vertex in the graph there are at least $k - 1$ other vertices of the same degree. Our algorithmic results shed light on the performance quality of a popular heuristic due to Liu and Terzi [ACM SIGMOD 2008]; in particular, we show that the heuristic provides optimal solutions in case that many edges need to be added. Based on this, we develop a polynomial-time data reduction, yielding a polynomial-size problem kernel for the problem parameterized by the maximum vertex degree. This result is in a sense tight since we also show that the problem is already NP-hard for H-index three, implying NP-hardness for smaller parameters such as average degree and degeneracy.

1 Introduction

For many scientific disciplines, including the understanding of the spread of diseases in a globalized world or power consumption habits with impact on fighting global warming, the availability of (anonymized) social network data becomes more and more important. In a landmark paper Liu and Terzi [16] introduced the following simple graph-theoretic model for identity anonymization on (social) networks. Herein, they transferred the k-anonymity concept known for tabular data in databases [9] to graphs.

Degree Anonymity [16]
Input: An undirected graph $G = (V, E)$ and two positive integers k and s.
Question: Is there an edge set E' over V with $|E'| \leq s$ such that $G' = (V, E \cup E')$ is k-anonymous, that is, for every vertex $v \in V$ there are at least $k - 1$ other vertices in G' having the same degree?

Liu and Terzi [16] assume in this model that an adversary (who wants to de-anonymize the network) knows only the degree of the vertex of a target individual;

this is a modest adversarial model. Clearly, there are stronger adversarial models which (in many cases very realistically) assume that the adversary has more knowledge, making it possible to breach privacy provided by a "k-anonymized graph" [20]. Moreover, it has been argued that graph anonymization has fundamental theoretical barriers which prevent a fully effective solution [1]. DEGREE ANONYMITY, however, provides the perhaps most basic and still practically relevant model for graph anonymization; it is the subject of active research [4, 5, 18].

Graph anonymization problems are typically NP-hard. Thus, almost all algorithms proposed in this field are heuristic in nature, this also being true for algorithms for DEGREE ANONYMITY [16, 18]. Indeed, as the field of graph anonymization is young and under strong development, there is very little research on its theoretical foundations, particularly concerning computational complexity and algorithms with provable performance guarantees [6].

Our contributions. Our central result is to show that DEGREE ANONYMITY has a polynomial-size problem kernel when parameterized by the maximum vertex degree Δ of the input graph. In other words, we prove that there is a polynomial-time algorithm that transforms any input instance of DEGREE ANONYMITY into an equivalent instance with at most $O(\Delta^7)$ vertices. Indeed, we encounter a "win-win" situation when proving this result: We show that Liu and Terzi's heuristic strategy [16] finds an optimal solution when the size s of a minimum solution is larger than $2\Delta^4$. As a consequence, we can bound s in $O(\Delta^4)$ and, hence, a polynomial kernel we provide for the combined parameter (Δ, s) is also a polynomial kernel only for Δ. Furthermore, our kernelization has the useful property (e. g. for approximations) that each solution derived for the kernel instance has a one-to-one correspondence to a solution of the original instance. While this kernelization directly implies fixed-parameter tractability for DEGREE ANONYMITY parameterized by Δ, we also develop a further improved fixed-parameter algorithm.

In addition, we prove that the polynomial kernel for Δ is tight in the sense that even for constant values of the "stronger" parameter (that is, provably smaller) H-index[1], DEGREE ANONYMITY becomes NP-hard. The same proof also yields NP-hardness in 3-colorable graphs. Further, from a parameterized perspective, we show that DEGREE ANONYMITY is W[1]-hard when parameterized by the solution size s (the number of added edges), even when $k = 2$. In other words, there is no hope for tractability even when the level k of anonymity is low and the graph needs only few edge additions (meaning little perturbation) to achieve k-anonymity.

Why is the parameter "maximum vertex degree Δ" of specific interest? First, note that from a parameterized complexity perspective it seems to be a "tight" parameterization in the sense that for the only little "stronger" parameter H-index our results already show NP-hardness for H-index three (also implying hardness e.g. for the parameters degeneracy and average degree). Social networks typically have few vertices with high degree and many vertices of small degree.

---

[1] The H-index of a graph G is the maximum integer h such that G has at least h vertices with degree at least h. Thus G has at most h vertices of degree larger than h.

Leskovec and Horvitz [15] studied a huge instant-messaging network (180 million nodes) with maximum degree bounded by 600. For the DBLP co-author graph generated in February 2012 with more than 715,000 vertices we measured a maximum degree of 804 and an H-index of 208, so there are not more than 208 vertices with degree larger than 208. Thus, a plausible strategy might be to only anonymize vertices of "small" degree and to remove high-degree vertices for the anonymization process because it might be overly expensive to anonymize these high-degree vertices and since they might be well-known (that is, not anonymous) anyway. Indeed, high-degree vertices can be interpreted as outliers [2], potentially making their removal plausible.

Related work. The most important reference is Liu and Terzi's work [16] where the basic model was introduced, sophisticated (heuristic) algorithms (also using algorithms to determine the realizability of degree sequences) have been developed and validated on experimental data. Somewhat more general models have been considered by Zhou and Pei [25] (studying the neighborhood of vertices instead of only the degree) and by Chester et al. [5] (anonymizing a *subset* of the vertices of the input). Chester et al. [4] investigate the variant of adding vertices instead of edges. Building on Liu and Terzi's work, Lu et al. [18] propose a "more efficient and more effective" algorithm for DEGREE ANONYMITY. Again, this algorithm is heuristic in nature. Today, the field of graph anonymization has grown tremendously with numerous surveys and research directions. We only mention some directly related work.

There are many and more complicated models for graph anonymization. Weaknesses (mainly depending on the assumed adversary model where for many practical situations the adversary may e.g. have an auxiliary network that helps in de-anonymizing) of DEGREE ANONYMITY and other models have been pointed out [1, 20, 24]. In conclusion, given the generality of background knowledge an adversary may or may not have, graph anonymization remains a chimerical target [18] and, thus, a universally best model is not available.

Finally, from a (parameterized) computational complexity perspective, the closest work we are aware of is due to Mathieson and Szeider [19] who provide a study on editing graphs to satisfy degree constraints. In their basic model, each vertex is equipped with a degree list and the task is to edit the graph such that each vertex achieves a degree contained in its degree list. They study the editing operations edge addition, edge deletion, and vertex deletion and provide numerous parameterized tractability and intractability results. Interestingly, on the technical side they also rely on the computation of general factors in graphs (as we do) and they also study kernelization, where they leave as most challenging open problem to extend their kernelization results to cases that include vertex deletion and edge addition, emphasizing that the presence of edge additions makes their approach inapplicable.

Due to the lack of space, many technical details are deferred to a full version of the paper.

2 Preliminaries

Parameterized complexity. The concept of parameterized complexity was pioneered by Downey and Fellows [7] (see also [8, 21] for more recent textbooks). A parameterized problem is called *fixed-parameter tractable* if there is an algorithm that decides any instance (I, k), consisting of the "classical" instance I and a parameter $k \in \mathbb{N}_0$, in $f(k) \cdot |I|^{O(1)}$ time, for some computable function f solely depending on k. A core tool in the development of fixed-parameter algorithms is polynomial-time kernelization [3, 12]. Here, the goal is to transform a given problem instance (I, k) in polynomial time into an equivalent instance (I', k'), the so-called *kernel*, such that $k' \leq g(k)$ and $|I'| \leq g(k)$ for some function g. If g is a polynomial, then it is called a *polynomial kernel*. A parameterized problem that is classified as W[1]-hard (using so-called parameterized reductions) is unlikely to admit a fixed-parameter algorithm. There is good complexity-theoretic reason to believe that W[1]-hard problems are not fixed-parameter tractable.

Graphs and anonymization. We use standard graph-theoretic notation. All graphs studied in this paper are simple, i.e., there are no self-loops and no multi-edges. For a given graph $G = (V, E)$ with vertex set V and edge set E we set $n = |V|$ and $m = |E|$. Furthermore, by $\deg_G(v)$ we denote the degree of a vertex $v \in V$ in G and Δ_G denotes the maximum degree of any vertex of G. For $0 \leq a \leq \Delta_G$ let $D_G(a) = \{v \in V \mid \deg_G(v) = a\}$ be the *block* of degree a, that is, the set of all vertices with degree a in G. Thus, being k-anonymous is equivalent to each block being of size either zero or at least k. The complement graph of G is denoted by $\overline{G} = (V, \overline{E})$, $\overline{E} = \{\{u, v\} \mid u, v \in V, \{u, v\} \notin E\}$. The subgraph of G induced by a vertex subset $V' \subseteq V$ is denoted by $G[V']$. For an edge subset $E' \subseteq E$, $V(E')$ denotes the set of all endpoints of edges in E' and $G[E'] = (V(E'), E')$. For a set of edges S with endpoints in a graph G, we denote by $G + S$ the graph that results by inserting all edges in S into G and we call S an edge insertion set for G. Thus, DEGREE ANONYMITY is the question whether there is an edge insertion set S of size at most s such that $G + S$ is k-anonymous. In this case S is called *k-insertion set* for G. We omit subscripts if the graph is clear from the context.

3 Hardness Results

In this section we provide two polynomial-time many-to-one reductions yielding three (parameterized) hardness results.

Theorem 1. DEGREE ANONYMITY *is NP-hard on 3-colorable graphs and on graphs with H-index three.*

Proof (Sketch). We give a reduction from the NP-hard INDEPENDENT SET problem, where given a graph $G = (V, E)$ and a positive integer h, the question is whether there is a size-h independent set, that is, a vertex subset of pairwise nonadjacent vertices. We assume without loss of generality that in the given INDEPENDENT SET instance (G, h) it holds that $|V| \geq 2h + 1$. We construct an

equivalent instance $(G' = (V', E'), k, s)$ for DEGREE ANONYMITY as follows. We start with a copy G' of G, denoting with $v' \in V'$ the copy of the vertex $v \in V$. Then, for each vertex $v \in V$ we add to G' degree-one vertices adjacent to v' such that v' has degree Δ_G in G'. Next we add a star with $\Delta_G + h - 1$ leaves and denote its central vertex c. We conclude the construction by setting $k = h + 1$ and $s = \binom{h}{2}$.

INDEPENDENT SET is NP-hard on 3-colorable graphs [23, Lemma 6] and on graphs with maximum degree three [11, GT20]. Observe that if G is 3-colorable, then G' is also 3-colorable. Furthermore, if G has maximum degree three, then only the central vertex c might have degree larger than three, implying that the H-index of G' is three. □

DEGREE ANONYMITY is W[1]-hard with respect to the standard parameterization, that is, by the size of edges s that are allowed to add:

Theorem 2. DEGREE ANONYMITY *is W[1]-hard parameterized by the number of inserted edges s, even if $k = 2$.*

4   Polynomial Kernel for the Maximum Degree Δ

In this main section we provide a polynomial kernel with respect to the parameter maximum degree Δ (Theorem 4). Our proof has two main ingredients: first we show in Section 4.2 a polynomial kernel with respect to the combined parameter (Δ, s); second we show in Section 4.3 that a slightly modified variant of Liu and Terzi's heuristic [16] exactly solves any instance having a minimum-size k-insertion set of size at least $(\Delta^2 + 4\Delta + 3)^2$. Hence, either we can solve a given instance in polynomial time or we can upper-bound s by $(\Delta^2 + 4\Delta + 3)^2$, implying that the kernel polynomial in (Δ, s) is indeed polynomial only in Δ. We begin by presenting the main technical tool used in our work, the so-called f-FACTOR problem.

4.1 f-Factor problem

DEGREE ANONYMITY has a close connection to the polynomial-time solvable f-FACTOR problem [17, Chapter 10]: Given a graph $G = (V, E)$ and a function $f : V \to \mathbb{N}_0$, does there exist an f-*factor*, that is, a subgraph $G' = (V, E')$ of G such that $\deg_{G'}(v) = f(v)$ for all vertices? One can reformulate DEGREE ANONYMITY using f-FACTOR as follows: Given an instance (G, k, s), the question is whether there is a function $f : V \to \mathbb{N}_0$ such that the complement graph $\overline{G}$ contains an f-factor, $\sum_{v \in V} f(v) \le 2s$ (every edge is counted twice in the sum of degrees), and for all $v \in V$ it holds that $|\{u \in V \mid \deg_G(u) + f(u) = \deg_G(v) + f(v)\}| \ge k$ (the k-anonymity requirement). The following lemma guarantees the existence of an f-factor in graphs fulfilling certain requirements on the maximum degree and the number of vertices.

Lemma 1 ([14]). *Let $G = (V, E)$ be a graph with minimum vertex degree δ and let $a \le b$ be two positive integers. Suppose further that*

$$\delta \ge \frac{b}{a+b}|V| \ \ and \ \ |V| > \frac{a+b}{a}(b + a - 3).$$

5

Then, for any function $f : V \to \{a, a + 1, ..., b\}$ where $\sum_{v \in V} f(v)$ is even, G has an f-factor.

As we are interested in an f-factor in the complement graph of our input graph G, we use Lemma 1 with $a = 1$, $b = \Delta + 2$, and minimum degree $\delta \geq n - \Delta - 1$. This directly leads to the following.

Corollary 1. *Let $G = (V, E)$ be a graph with n vertices, minimum degree $n - \Delta - 1$, $\Delta \geq 1$, and let $f : V \to \{1, \ldots, \Delta + 2\}$ be a function such that $\sum_{v \in V} f(v)$ is even. If $n \geq \Delta^2 + 4\Delta + 3$, then G has an f-factor.*

4.2 Polynomial kernel for (Δ, s)

Our kernelization algorithm is based on the following observation. For a given graph G, consider for some $1 \leq i \leq \Delta_G$ the block $D_G(i)$, that is, the set of all vertices of degree i. If $D_G(i)$ contains many vertices, then the vertices are "interchangeable":

Observation 1. *Let $G = (V, E)$ be a graph, let S be a k-insertion set for G, and let $v \in V(S) \cap D_G(i)$ be a vertex such that $|D_G(i)| > (\Delta + 2)s$. Then there exists a vertex $u \in D_G(i) \setminus V(S)$ such that replacing in every edge of S the vertex v by u results in a k-insertion set for G.*

Proof. Since $|S| \leq s$, the vertex v can be incident to at most s edges in S. Denoting the set of these edges by S^v, obviously one can replace v by $u \in D_G(i)$ if u is non-adjacent to all vertices in $V(S^v) \setminus \{v\}$ (this allows to insert all edges) and $u \notin V(S)$ (the size of all blocks in $G + S$ does not change). However, as $V(S)$ contains at most $2s$ vertices from $D_G(i)$ and each of the at most s vertices in $V(S^v) \setminus \{v\}$ has at most Δ_G neighbors in G, it follows that such a vertex $u \in D_G(i)$ exists if $|D_G(i)| > (\Delta + 2)s$. $\square$

By Observation 1, in our kernel we only need to keep at most $(\Delta + 2)s$ vertices in each block: If in an optimal k-insertion set S there is a vertex $v \in V(S)$ that we did not keep, then by Observation 1 we can replace v by some vertex we kept. There are two major problems that need to be fixed to obtain a kernel: First, when removing vertices from the graph, the degrees of the remaining vertices change. Second, k might be "large" and, thus, removing vertices (during kernelization) in one block may breach the k-anonymity constraint. To overcome the first problem we insert some "dummy-vertices" which are guaranteed not to be contained in any k-insertion set. However, to solve the second problem we need to adjust the parameter k as well as the number of vertices that we keep from each block.

We now explain the kernelization algorithm in detail (see Algorithm 1 for the pseudocode). Let (G, s, k) be an instance of DEGREE ANONYMITY. For brevity we set $\beta = (\Delta + 4)s + 1$. We compute in polynomial time an equivalent instance (G', s, k') with at most $O(\Delta^3 s)$ vertices: First set $k' = \min\{k, \beta\}$ (Line 4). We arbitrarily select from each block $D_G(i)$ a certain number x of vertices and collect all these vertices into the set A (Line 14). To cope with the above mentioned

Algorithm 1 The pseudocode or the algorithm producing a polynomial kernel with respect to (Δ, s).

1: **procedure** PRODUCEPOLYKERNEL($G = (V, E)$, k, s)
2:     **if** $|V| \leq \Delta(\beta + 4s)$ **then**   // β is defined as $\beta = (\Delta + 4)s + 1$
3:         **return** (G, k, s)
4:     $k' \leftarrow \min\{k, \beta\}$; $A \leftarrow \emptyset$
5:     **for** $i \leftarrow 1$ **to** Δ **do**
6:         **if** $2s < |D_G(i)| < k - 2s$ **then**
7:             **return** trivial no-instance   // insufficient budget for $D_G(i)$
8:         **if** $k \leq \beta$ **then**   // determine retained vertices
9:             $x \leftarrow \min\{|D_G(i)|, \beta + 4s\}$
10:        **else if** $|D_G(i)| \leq 2s$ **then**
11:            $x \leftarrow |D_G(i)|$
12:        **else**
13:            $x \leftarrow k' + \min\{4s, (|D_G(i)| - k)\}$ // observe that $k' = \beta$.
14:        add x vertices from $D_G(i)$ to A
15:     $G' = G[A]$
16:     **for each** $v \in A$ **do**   // add vertices to preserve degree of retained vertices
17:         add to G' $\deg_G(v) - \deg_{G'}(v)$ many degree-one vertices adjacent to v
18:     denote with P the set of vertices added in Line 17
19:     by adding matched pairs of vertices, ensure that $|P| \geq \max\{4\Delta + 4s + 4, k'\}$
20:     **if** $\Delta + s + 1$ is even **then**
21:         $G^F = (P, E^F) \leftarrow (\Delta + s + 1)$-factor in $\overline{G'[P]}$
22:     **else**
23:         $G^F = (P, E^F) \leftarrow (\Delta + s + 2)$-factor in $\overline{G'[P]}$
24:     $G' \leftarrow G' + E^F$
25:     **return** (G', k', s)

---

second problem, the "certain number" is defined in a case distinction on the value of k (see Lines 5 to 14). Intuitively, if k is large then we distinguish between "small" blocks of size at most $2s$ and "large" blocks of size at least $k - 2s$. Obviously, if there is a block which is neither small nor large, then the instance is a no-instance (see Line 7). Thus, in the problem kernel we keep for small blocks the "distance to size zero" and for large blocks the "distance to size k". Furthermore, in order to separate between small and large blocks it is sufficient that $k' > 4s$. However, to guarantee that Observation 1 is applicable, the case distinction is a little bit more complicated, see Lines 5 to 14.

We start building G' by first copying $G[A]$ into it (Line 15). Next, adding a pendant vertex to v means that we add a new vertex to G' and make it adjacent to v. For each $v \in A$ we add pendant vertices to v to ensure that $\deg_{G'}(v) = \deg_G(v)$ (Line 17). The vertices of A stay untouched in the following. Denote the set of all pendant vertices by P. Next, we add enough pairwise adjacent vertices to P to ensure that $|P| \geq \max\{k', 4\Delta + 4s + 4\}$ (Line 19). Hence, $|P| \leq \max\{|A| \cdot \Delta, k', 4\Delta + 4s + 4\} + 1$. To avoid that vertices in P help to anonymize the vertices in A we "shift" the degree of the vertices in P (see

Lines 20 to 24): We add edges between the vertices in P to ensure that the degree of all vertices in P is $\Delta + s + 2$ (when $\Delta + s + 1$ is even) or $\Delta + s + 3$ (when $\Delta + s + 2$ is even). For the ease of notation let χ denote the new degree of the vertices in P. Observe that before adding edges all vertices in P have degree one in G'. Thus, the minimum degree in $\overline{G'[P]}$ is $|P| - 2$. Furthermore, for each $v \in P$ we denote by $f(v)$ the number of incident edges v requires to have the described degree. It follows that $f(v)$ is even and hence $\sum_{v \in P} f(v)$ is even. Hence setting $a = b = \chi$ fulfills all conditions of Lemma 1. Thus, the required f-factor exists and can be found in $O(|P|\sqrt{|P|(\Delta + s)})$ time [10]. This completes the description of the kernelization algorithm.

The key point of the correctness of the kernelization is to show that without loss of generality, no k-insertion set S for G' of size $|S| \leq s$ affects any vertex in P. This is ensured by "shifting" the degree of all vertices in P by $s + 1$ (or $s + 2$), implying that none of the vertices in A can "reach" the degree of any vertex in P by adding at most s edges. Hence each block either is a subset of A or of P. We now prove that we may assume that an edge insertion set does not affect any vertex in P. All what we need to prove this is the fact that A contains at least $\beta + 4s$ vertices from at least one block in G. Observe that this is ensured by the check in Line 2.

Lemma 2. *If there is a k-insertion set S for G' with $|S| \leq s$, then there is also a k-insertion set S' for G' with $|S'| = |S|$ such that $V(S') \cap P = \emptyset$.*

Based on Lemma 2 we now prove the correctness of our kernelization algorithm.

Theorem 3. DEGREE ANONYMITY *admits a kernel with $O(\Delta^3 s)$-vertices.*

Proof. The polynomial kernel is computed by Algorithm 1. As f-FACTOR can be solved in polynomial time [10], Algorithm 1 runs in polynomial time. The correctness of the kernelization algorithm is deferred to a full version. It remains to show the size of the kernel. To this end, observe that each block in A has size at most $\beta + 4s$ (see Lines 9, 11, and 13). Thus, $|A| = O(\Delta\beta) = O(\Delta^2 s)$. Furthermore, the set P contains at most $\max\{\Delta|A|, k', 4s + 4\Delta + 1\}$ vertices (see Lines 17 to 19). Thus, $|P| = O(\Delta^3 s)$ and, hence, the reduced instance contains $O(\Delta^3 s)$ vertices. $\qquad\square$

4.3 A polynomial-time algorithm for "large" solution instances

In this section we show that if a minimum size k-insertion set S is "large" compared to Δ, then one can solve the instance in polynomial time (Lemma 5). Towards this, we first show that a large solution influences the degree of many vertices. Then the main idea is that if it influences the degree of "many" vertices from the same block, say $D_G(i)$, then by Observation 1 the corresponding vertices can be arbitrarily "interchanged". Thus it is not important to know which vertex from $D_G(i)$ has to be "moved" up to a certain degree by adding edges, because Observation 1 ensures that we can greedily find one. This, however, implies that the actual structure of the input graph (which forbids to insert certain edges

since they are already present) no longer matters. Hence, we solve DEGREE ANONYMITY without taking the graph structure into account. Thereby, if we can k-anonymize the degree sequence corresponding to G (the sequence of degrees of G) such that "many" degrees have to be adjusted, then by Corollary 1 we can conclude that $\overline{G}$ contains an f-factor where $f(v)$ captures the difference between the degree of v in G and the anonymized degree sequence. The f-factor can be found in polynomial time [10] and, hence, a k-insertion set can be found in polynomial time. We now formalize this idea.

We first show that a "large" minimum-size k-insertion set increases the maximum degree by at most two.

Lemma 3. *Let $G = (V, E)$ be a graph and let S be a minimum-size k-insertion set. If $|V(S)| \geq \Delta^2 + 4\Delta + 3$, then the maximum degree in $G + S$ is at most $\Delta + 2$.*

Proof. Let G be a graph with maximum degree Δ and k an integer. Let S be a minimum-size edge set such that $G + S$ is k-anonymous and $|V(S)| \geq \Delta^2 + 4\Delta + 3$. Now assume towards a contradiction that the maximum degree in $G + S$ is at least $\Delta + 3$. We show that there exists an edge set S' such that $G + S'$ is k-anonymous, $|S'| < |S|$, and $G + S'$ has maximum degree at most $\Delta + 2$.

First we introduce some notation. Let f be a function $f : V \to \mathbb{N}_0$ defined as $f(v) = \deg_{G+S}(v) - \deg_G(v)$ for all $v \in V$. Furthermore, denote with X the set of all vertices having degree more than $\Delta + 2$ in $G + S$, that is, $X = \{v \in V \mid f(v) + \deg_G(v) \geq \Delta + 3\}$. Observe that $G[S]$ is an f-factor of $\overline{G}$ and $2|S| = \sum_{v \in V} f(v)$. We now define a new function $f' : V \to \mathbb{N}_0$ such that $\overline{G}$ contains an f'-factor denoted by $G' = (V, S')$ where the edge set S' has the properties as described in the previous paragraph.

We define f' for all $v \in V$ as follows:

$$
f'(v) = \begin{cases} f(v) & \text{if } v \notin X, \\ \Delta - \deg_G(v) + 1 & \text{if } v \in X \text{ and } f(v) + \deg_G(v) - \Delta - 1 \text{ is even}, \\ \Delta - \deg_G(v) + 2 & \text{else.} \end{cases}
$$

First observe that $\deg_G(v) + f'(v) \leq \Delta + 2$ for all $v \in V$. Furthermore, observe that $f'(v) = f(v)$ for all $v \in V \setminus X$ and for all $v \in X$ it holds that $f'(v) < f(v)$ and $f(v) - f'(v)$ is even. Thus, $\sum_{v \in V} f(v) > \sum_{v \in V} f'(v)$ and $\sum_{v \in V} f'(v)$ is even. It remains to show that (i) $\overline{G}$ contains an f'-factor $G' = (V, S')$ and (ii) $G + S'$ is k-anonymous.

To prove (i) let $\widetilde{V} = \{v \in V \mid f'(v) > 0\}$ and observe that if $f(v) > 0$, then, by definition of f', we have $f'(v) > 0$ and hence $\widetilde{V} = V(S)$. Furthermore, let $\widetilde{G} = \overline{G[\widetilde{V}]}$. Observe that $\widetilde{G}$ has minimum degree $|\widetilde{V}| - \Delta - 1$ and $|\widetilde{V}| = |V(S)| \geq \Delta^2 + 4\Delta + 3$. Thus, the conditions of Corollary 1 are satisfied and hence $\widetilde{G}$ contains an $f'|_{\widetilde{V}}$-factor $\widetilde{G}' = (\widetilde{V}, S')$. By definition of $\widetilde{V}$ it follows that $G' = (V, S')$ is an f'-factor of $\overline{G}$. Thus, it remains to show (ii).

Assume towards a contradiction that $G + S'$ is not k-anonymous, that is, there exists some vertex $v \in V$ such that $1 \leq |D_{G+S'}(\deg_{G+S'}(v))| < k$. Let $d = \deg_{G+S}(v)$ and $d' = \deg_{G+S'}(v)$. Observe that $d' = \deg_G(v) + f'(v)$. Thus, if $v \notin$

9

X, then by definition of f' it holds that $d' = \deg_G(v) + f(v) = d \leq \Delta + 2$. Hence, for all vertices $u \in D_{G+S}(d')$ it follows that $u \notin X$. Thus, $D_{G+S}(d') \subseteq D_{G+S'}(d')$ and since $G + S$ is k-anonymous we have $|D_{G+S'}(d')| \geq k$, a contradiction. On the other hand, if $v \in X$, that is, $d > \Delta + 2$, then $|D_{G+S}(d)| \geq k$ since $G + S$ is k-anonymous. Furthermore, by the definitions of $D_{G+S}(d)$, f, and X we have for all $u \in D_{G+S}(d)$ that $\deg_G(u) + f(u) = d$, $u \in X$, and, thus, $f'(u) + \deg_G(u) = d'$. Therefore, $D_{G+S}(d) \subseteq D_{G+S'}(d')$ and $|D_{G+S'}(d')| \geq k$, a contradiction. □

We now formalize the anonymization of degree sequences. A multiset of positive integers $\mathcal{D} = d_1, \ldots, d_n$, d_i that corresponds to the degrees of all vertices in a graph is called *degree sequence*. A degree sequence $\mathcal{D}$ is k-anonymous if each number in $\mathcal{D}$ occurs at least k times in $\mathcal{D}$. Clearly, the degree sequence of a k-anonymous graph G is k-anonymous. Moreover, if a graph G can be transformed by at most s edge insertions into a k-anonymous graph, then the degree sequence of G can be transformed into a k-anonymous degree sequence by increasing the integers by no more than $2s$ in total (clearly, in the other direction this fails in general because of the graph structure). As we are only interested in a degree sequence corresponding to a graph of a DEGREE ANONYMITY instance where s is large, by Lemma 3 we can require the integers in a k-anonymous degree sequence to be at upper-bounded by $\Delta + 2$.

k-DEGREE SEQUENCE ANONYMITY (k-DSA)
Input: Two positive integers k and s and a degree sequence $\mathcal{D} = d_1, \ldots, d_n$ with $d_1 \leq d_2 \leq \ldots \leq d_n$ and $\Delta = d_n$.
Question: Is there a k-anonymous degree sequence $\mathcal{D}' = d'_1, \ldots, d'_n$ with $d_i \leq d'_i$ and $\max_{1 \leq i \leq n} d'_i \leq \Delta + 2$ such that $\sum_{i=1}^n d'_i - d_i = 2s$?

By slightly modifying a dynamic programming-based heuristic introduced by Liu and Terzi [16], we next prove that k-DEGREE SEQUENCE ANONYMITY is polynomial-time solvable. Note that Liu and Terzi [16] use their heuristic to solve DEGREE ANONYMITY by first solving the problem on the degree sequence of the input graph G and then trying to "realize" (adding the corresponding edges to G) the produced k-anonymous degree sequence.

Lemma 4. k-DEGREE SEQUENCE ANONYMITY *can be solved in polynomial time.*

We now have all ingredients to solve DEGREE ANONYMITY in polynomial time in case it has a "large" minimum-size k-insertion set. More formally, let (G, k, s) be an instance and let $\mathcal{D}$ be the degree sequence of G, then first find the largest $i \leq s$ such that $(\mathcal{D}, k, i)$ is a yes-instance for k-DEGREE SEQUENCE ANONYMITY. If i is "large", then we prove that we can transfer the solution to G. In all other cases, since any k-insertion set for G of size $j \leq s$ directly implies that $(\mathcal{D}, k, j)$ is a yes-instance for k-DEGREE SEQUENCE ANONYMITY, it follows that we can bound the parameter s by a function in Δ.

Lemma 5. *Let (G, k, s) be an instance of* DEGREE ANONYMITY. *Either one can decide the instance in polynomial time or (G, k, s) is a yes-instance if and only if $(G, k, \min\{(\Delta^2 + 4\Delta + 3)^2, s\})$ is a yes-instance.*

By [Lemma 5](#) it follows that in polynomial time we can either find a solution or we have $s < (\Delta^2 + 4\Delta + 3)^2$. By [Theorem 3](#) this implies our main result.

Theorem 4. DEGREE ANONYMITY *admits an $O(\Delta^7)$-vertex kernel.*

5 Fixed-Parameter Algorithm

We provide a direct combinatorial algorithm for the combined parameter (Δ, s). Roughly speaking, for fixed k-insertion set S the algorithm branches into all suitable structures of $G[S]$, that is, graphs of at most $2s$ vertices with vertex labels from $\{1, \ldots, \Delta\}$. Then the algorithm checks whether this structure occurs as a subgraph in $\overline{G}$ such that the labels on the vertices match the degree of the corresponding vertex in G.

Theorem 5. DEGREE ANONYMITY *can be solved in $(6s^2\Delta^3)^{2s} \cdot s^2 \cdot n^{O(1)}$ time.*

Note that due to the upper bound $s < (\Delta^2 + 4\Delta + 3)^2$ (see [Lemma 5](#)) and the polynomial kernel for the parameter Δ (see [Theorem 4](#)), [Theorem 5](#) also provides an algorithm running in $\Delta^{O(\Delta^4)} + n^{O(1)}$ time.

6 Conclusion

One of the grand challenges of theoretical research on computationally hard problems is to gain a better understanding of when and why heuristic algorithms work [13]. In this theoretical study, we contributed to a better theoretical understanding of a basic problem in graph anonymization, on the one side partially explaining the quality of a successful heuristic approach [16] and on the other side providing a first step towards a provably efficient algorithm for relevant special cases (bounded-degree graphs). Our work just being one of the first steps in the so far underdeveloped field of studying the computational complexity of graph anonymization [6], there are numerous challenges for future research. For instance, our focus was on classification results rather than engineering the upper bounds, a natural next step to do. Second, it would be interesting to perform a data-driven analysis of parameter values on real-world networks in order to gain parameterizations that can be exploited in a broad-band multivariate complexity analysis [22] of DEGREE ANONYMITY. Finally, with DEGREE ANONYMITY we focused on a very basic problem of graph anonymization; there are numerous other models (partially mentioned in the introductory section) that ask for similar studies.

Bibliography

[1] C. C. Aggarwal, Y. Li, and P. S. Yu. On the hardness of graph anonymization. In *Proc. 11th IEEE ICDM*, pages 1002–1007. IEEE, 2011.

[2] G. Aggarwal, T. Feder, K. Kenthapadi, S. Khuller, R. Panigrahy, D. Thomas, and A. Zhu. Achieving anonymity via clustering. *ACM Transactions on Algorithms*, 6 (3):1–19, 2010.

[3] H. L. Bodlaender. Kernelization: New upper and lower bound techniques. In *Proc. 4th IWPEC*, volume 5917 of *LNCS*, pages 17–37. Springer, 2009.

[4] S. Chester, B. M. Kapron, G. Ramesh, G. Srivastava, A. Thomo, and S. Venkatesh. *k*-Anonymization of social networks by vertex addition. In *Proc. 15th ADBIS (2)*, volume 789 of *CEUR Workshop Proceedings*, pages 107–116. CEUS-WS.org, 2011.

[5] S. Chester, J. Gaertner, U. Stege, and S. Venkatesh. Anonymizing subsets of social networks with degree constrained subgraphs. In *Proc. ASONAM*, pages 418–422. IEEE Computer Society, 2012.

[6] S. Chester, B. Kapron, G. Srivastava, and S. Venkatesh. Complexity of social network anonymization. *Social Network Analysis and Mining*, 2012. online available.

[7] R. G. Downey and M. R. Fellows. *Parameterized Complexity*. Springer, 1999.

[8] J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer, 2006.

[9] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4):14:1–14:53, 2010.

[10] H. N. Gabow. An efficient reduction technique for degree-constrained subgraph and bidirected network flow problems. In *Proc. 15th STOC*, pages 448–456. ACM, 1983.

[11] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1979.

[12] J. Guo and R. Niedermeier. Invitation to data reduction and problem kernelization. *SIGACT News*, 38(1):31–45, 2007.

[13] R. M. Karp. Heuristic algorithms in computational molecular biology. *J. Comput. Syst. Sci.*, 77(1):122–128, 2011.

[14] P. Katerinis and N. Tsikopoulos. Minimum degree and f-factors in graphs. *New Zealand J. Math*, 29(1):33–40, 2000.

[15] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *Proc. 17th WWW*, pages 915–924. ACM, 2008.

[16] K. Liu and E. Terzi. Towards identity anonymization on graphs. In *Proc. ACM SIGMOD '08*, pages 93–106. ACM, 2008.

[17] L. Lovász and M. D. Plummer. *Matching Theory*, volume 29 of *Annals of Discrete Mathematics*. North-Holland, 1986.

[18] X. Lu, Y. Song, and S. Bressan. Fast identity anonymization on graphs. In *Proc. 23rd DEXA Part I*, volume 7446 of *LNCS*, pages 281–295. Springer, 2012.

[19] L. Mathieson and S. Szeider. Editing graphs to satisfy degree constraints: A parameterized approach. *J. Comput. Syst. Sci.*, 78(1):179–191, 2012.

[20] A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *Proc. 30th IEEE SP*, pages 173–187. IEEE, 2009.

[21] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford University Press, 2006.

[22] R. Niedermeier. Reflections on multivariate algorithmics and problem parameterization. In *Proc. 27th STACS*, volume 5 of *LIPIcs*, pages 17–32. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2010.

[23] C. Phillips and T. J. Warnow. The asymmetric median tree—a new model for building consensus trees. *Discrete Appl. Math.*, 71(1–3):311–335, 1996.

[24] A. Sala, X. Zhao, C. Wilson, H. Zheng, and B. Y. Zhao. Sharing graphs using differentially private graph models. In *Proc. 11th ACM SIGCOMM*, pages 81–98. ACM, 2011.

[25] B. Zhou and J. Pei. The *k*-anonymity and *l*-diversity approaches for privacy preservation in social networks against neighborhood attacks. *Knowl. Inf. Syst.*, 28(1):47–77, 2011.