Parameterized Inapproximability of Degree Anonymization

Cristina Bazgan[1,3] and André Nichterlein[2]

[1] PSL, Université Paris-Dauphine, LAMSADE UMR CNRS 7243, France
bazgan@lamsade.dauphine.fr
[2] Institut für Softwaretechnik und Theoretische Informatik, TU Berlin, Germany
andre.nichterlein@tu-berlin.de
[3] Institut Universitaire de France

Abstract. The Degree Anonymity problem arises in the context of combinatorial graph anonymization. It asks, given a graph G and two integers k and s, whether G can be made k-*anonymous* with at most s modifications. Here, a graph is k-anonymous if the graph contains for every vertex at least $k-1$ other vertices of the same degree. Complementing recent investigations on its computational complexity, we show that this problem is very hard when studied from the viewpoints of approximation as well as parameterized approximation. In particular, for the optimization variant where one wants to minimize the number of either edge or vertex deletions there is no factor-$n^{1-\varepsilon}$ approximation running in polynomial time unless P = NP, for any constant $0 < \varepsilon \leq 1$. For the variant where one wants to maximize k and the number s of either edge or vertex deletions is given, there is no factor-$n^{1/2-\varepsilon}$ approximation running in time $f(s) \cdot n^{O(1)}$ unless W[1] = FPT, for any constant $0 < \varepsilon \leq 1/2$ and any function f. On the positive side, we classify the general decision version as fixed-parameter tractable with respect to the combined parameter solution size s and maximum degree.

1 Introduction

Releasing social network data without violating the privacy of the users has become an important and active field of research [15]. One model aiming for this goal was introduced by Liu and Terzi [13] who transferred the k-anonymity concept from tabular data in databases [9] to graphs. Herein, Liu and Terzi [13] require that a released graph contains for every vertex at least $k-1$ other vertices with the same degree. The parameter k controls how many individuals are at least linked to one particular degree and thus higher values for k give higher levels of anonymity. We remark that this model has also some weaknesses. Refer to Wu et al. [15] for more details and further anonymization models.

Here, we study the following variant of the model of Liu and Terzi [13].

Degree Anonymity (Anonym)

Input: An undirected graph $G = (V, E)$ and two positive integers k and s.

Question: Can G be transformed with at most s modifications into a k-anonymous graph $G' = (V', E')$, that is, for each vertex in G' there are $k-1$ other vertices of the same degree?

We will use the name scheme Anonym-{E/V}-{Ins/Del/Edt} to distinguish the different graph modification operations edge/vertex insertion/deletion/editing. Liu and Terzi [13] studied edge insertions (Anonym-E-Ins), but also vertex deletions (Anonym-V-Del) [2] and vertex insertions (Anonym-V-Ins) [3, 5] have been considered. While the focus of previous work was on experimentally evaluated heuristics and algorithms [12, 13] or computational complexity and fixed-parameter algorithms [2, 3, 11], we study the polynomial-time and parameterized approximability of these problems. To this end, we mostly concentrate on natural optimization variants of the two problems where either edge deletions (Anonym-E-Del) or vertex deletions (Anonym-V-Del) are allowed. Partially answering an open question of Chester et al. [4], we show strong inapproximable results, even when allowing the running time to be exponential in s. We remark that our results do not transfer to the problem variants allowing to edit up to s edges (Anonym-E-Edt) and the status of the (parameterized) approximability of the corresponding optimization problems remains unsolved.

Related Work. The basic degree anonymization model was introduced by Liu and Terzi [13] (also see Clarkson et al. [6] for an extended version); they also gave an experimentally evaluated heuristic for Anonym-E-Ins. One of the first theoretical works on this model was done by Chester et al. [4]. They provided polynomial-time algorithms for bipartite graphs and showed NP-hardness of generalizations of Anonym-E-Ins with edge labels. In particular, they asked for effective approximation algorithms for Anonym-E-Ins and generalizations. Hartung et al. [11] proved that Anonym-E-Ins is NP-hard and W[1]-hard with respect to (w.r.t.) the solution size s, even if $k = 2$. On the positive side, using the heuristic of Liu and Terzi [13], they showed fixed-parameter tractability of Anonym-E-Ins w.r.t. the maximum degree in the input graph.

Chester et al. [5] considered a variant of Anonym-V-Ins and gave an approximation algorithm with an additive error of at most k. Bredereck et al. [3] investigated the parameterized complexity of several variants of Anonym-V-Ins which differ in the rules how the inserted vertices can be made adjacent to existing vertices. The ANONYM-V-DEL variant studied by Bredereck et al. [2] turned out to be NP-hard even on very restricted graph classes such as trees, split graphs, or trivially perfect graphs.

Our Results. We investigate the approximability of natural optimization variants of ANONYM-V-DEL and ANONYM-E-DEL: Either the budget s is given and one wants to maximize the level k of anonymity, or k is given and the goal is to minimize the number of modifications s. The optimization problems maximizing k with a given budget s are denoted by MAX ANONYM-V-DEL and MAX ANONYM-E-DEL. The variants minimizing s with given k are denoted by MIN ANONYM-E-DEL and MIN ANONYM-V-DEL.

We show that one cannot approximate MAX ANONYM-E-DEL (MAX ANONYM-V-DEL) within a factor of $n^{1-\varepsilon}$ ($n^{1/2-\varepsilon}$) in $f(s)n^{O(1)}$ time unless FPT = W[1], for any function f and any $0 < \varepsilon \leq 1$ ($0 < \varepsilon \leq 1/2$). As the parameter k has size $\Theta(n)$ in all employed gap-reductions, we only manage to exclude *polynomial-time* approximations for the minimization versions. More precisely, both MIN ANONYM-E-DEL and MIN ANONYM-V-DEL cannot be approximated in polynomial time within a factor of $n^{1-\varepsilon}$ unless P = NP.

Complementing the NP-hardness of ANONYM-V-DEL with $k = 2$ on trees [2], we show that ANONYM-E-DEL remains NP-hard on caterpillars (a tree having a dominating path), even if $k = 2$. Extending the fixed-parameter tractability of ANONYM-V-DEL w.r.t. the combined parameter budget and maximum degree (s, Δ), we classify ANONYM (allowing edge and vertex insertion as well as deletion) as fixed-parameter tractable w.r.t. (s, Δ).

Due to the space constraints, some proofs are deferred to a full version.

2 Preliminaries

Graph terminology. We use standard graph-theoretic notation. All graphs studied in this paper are undirected and simple, that is, 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 $0 \leq d \leq \Delta_G$, let $D_G(d) := \{v \in V \mid \deg_G(v) = d\}$ be the *block* of degree d, that is, the set of all vertices with degree d in G. Thus, being k-anonymous is equivalent to each block being of size either zero or at least k.

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 $G[E'] := (V(E'), E')$. Furthermore, for a vertex subset $V' \subseteq V$ we set $G - V' := G[V \setminus V']$ and for an edge set $E' \subseteq \binom{V}{2}$ we set $G - E' := (V, E \setminus E')$ and $G + E' = (V, E \cup E')$. A graph G is *k-anonymous* if for every vertex $v \in V$ there are at least $k - 1$ other vertices in G having the same degree.

A vertex subset $V' \subseteq V$ (an edge subset $E' \subseteq E$) is called *k-deletion set* if $G - V'$ ($G - E'$, respectively) is k-anonymous. Analogously, for a set E'' of edges with endpoints in a graph G such that $V + E''$ is k-anonymous, we call E'' an *k-insertion set* for G. We omit subscripts if the graph is clear from the context.

Approximation. Let Σ be a finite alphabet. Given an optimization problem $Q \subseteq \Sigma^*$ and an instance I of Q, we denote by $|I|$ the size of I, by $opt(I)$ the optimum value of I and by $val(I, S)$ the value of a feasible solution S of I. The *performance ratio* of S (or *approximation factor*) is $r(I, S) = \max\left\{\frac{val(I,S)}{opt(I)}, \frac{opt(I)}{val(I,S)}\right\}$. For a function ρ, an algorithm is a *$\rho(n)$-approximation*, if for every instance I of Q, it returns a solution S such that $r(I, S) \leq \rho(|I|)$. An optimization problem is *$\rho(n)$-approximable in polynomial time* if there exists a $\rho(n)$-approximation algorithm running in time $|I|^{O(1)}$ for any instance I. A parameterized optimization problem $Q \subseteq \Sigma^* \times \mathbb{N}$ is *$\rho(n)$-approximable in fpt-time w.r.t. the parameter k* if there exists a $\rho(n)$-approximation algorithm running in time $f(k) \cdot |I|^{O(1)}$ for any instance (I, k) and f is a computable function [14]. It is worth pointing that in this case, k is not related to the optimization value.

In this paper we use a *gap*-reduction between a decision problem and a minimization or maximization problem. A decision problem A is called *gap-reducible* to a maximization problem Q with gap $\rho \geq 1$ if there exists a polynomial-time computable function that maps any instance I of A to an instance I' of Q, while satisfying the following properties: (i) if I is a yes-instance, then $opt(I') \geq c\rho$, and (ii) if I is a no-instance, then $opt(I') < c$, where c and ρ are functions of $|I'|$. If A is NP-hard, then Q is not ρ-approximable in polynomial time, unless P = NP. In this paper we also use a variant of this notion, called fpt gap-reduction.

Definition 1 (fpt gap-reduction). A parameterized (decision) problem A is called *fpt gap-reducible* to a parameterized maximization problem Q with gap $\rho \geq 1$ if any instance (I, k) of A can be mapped to an instance (I', k') of Q in $f(k) \cdot |I|^{O(1)}$ time while satisfying the following properties: (i)

$k' \leq g(k)$ for some function g, (ii) if I is a yes-instance, then $opt(I') \geq c\rho$, and (iii) if I is a no-instance, then $opt(I') < c$, where c and ρ are functions of $|I'|$ and k.

The interest of the fpt gap-reduction is the following result that immediately follows from the previous definition:

Lemma 1. *If a parameterized problem A is $\mathcal{C}$-hard and fpt gap-reducible to a parameterized optimization problem Q with gap ρ, then Q is not ρ-approximable in fpt-time, unless* $\mathsf{FPT} = \mathcal{C}$*, where $\mathcal{C}$ is any class of the parameterized complexity hierarchy.*

3 Inapproximability of vertex deletion versions

In this section we consider the optimization problems associated to Anonym-V-Del, that is Min Anonym-V-Del and Max Anonym-V-Del. We prove that Min Anonym-V-Del is not $n^{1-\varepsilon}$-approximable in polynomial time, while Max Anonym-V-Del is not $n^{1/2-\varepsilon}$-approximable in fpt-time w.r.t. parameter s, even on trees.

Theorem 1. Min Anonym-V-Del *is not $n^{1-\varepsilon}$-approximable for any* $0 < \varepsilon \leq 1$*, unless $P = NP$.*

Theorem 2. *For every* $0 < \varepsilon \leq 1/2$, Max Anonym-V-Del *is not $n^{1/2-\varepsilon}$-approximable in fpt-time w.r.t. parameter s, even on trees, unless $FPT = W[2]$.*

Proof. Let $0 < \varepsilon \leq 1/2$ be a constant. We provide an fpt gap-reduction from the W[2]-hard Set Cover problem [7] parameterized by the solution size h. Set Cover is defined as follows: given a universe $U = \{e_1, \ldots, e_m\}$, a collection $\mathcal{C} = \{S_1, \ldots, S_n\}$ of sets over U, and $h \in \mathbb{N}$ the task is to decide whether there is a set cover $\mathcal{C}' \subseteq \mathcal{C}$ of size $|\mathcal{C}'| \leq h$, that is $\bigcup_{S \in \mathcal{C}'} S = U$. Let $I = (U, \mathcal{C}, h)$ be an instance of Set Cover. We assume without loss of generality that for each element $e_i \in U$ there exists a set $S_j \in \mathcal{C}$ with $e_i \in S_j$. To reduce the amount of indices in the construction given below we introduce the function $f\colon U \to \mathbb{N}$ that maps an element $e_i \in U$ to $f(e_i) = (h+4)i$. Let t be an integer greater than or equal to $(mn)^{(1-2\varepsilon)/(2\varepsilon)}$. (We will aim for making the constructed graph t-anonymous.)

The instance I' of Max Anonym-V-Del is defined by $s = h$ and on a graph $G = (V, E)$ constructed as follows: For each element $e_i \in U$ add a star $K_{1,f(e_i)}$ with the center vertex v_i^e. Denote with $V_U = \{v_1^e, \dots, v_m^e\}$ the set of all these center vertices. Furthermore, for each element $e_i \in U$ add t stars $K_{1,f(e_i)+1}$.

For each set $S_j \in \mathcal{C}$ add a tree rooted in a vertex v_j^S. The root has $|S_j|t$ child vertices where each element $e_i \in S_j$ corresponds to exactly t of these children, denoted by $v_1^{e_i,S_j}, \dots, v_t^{e_i,S_j}$. Additionally, for each $\ell \in \{1, \dots, t\}$ we add to $v_\ell^{e_i,S_j}$ exactly $f(e_i)$ degree-one neighbors. Hence, the set gadget is a tree of depth two rooted in v_j^S. To ensure that the root v_j^S does not violate the t-anonymous property we add t stars $K_{1,\deg(v_j^S)}$. We denote with $V_\mathcal{C} = \{v_1^S, \dots, v_n^S\}$ the set of all root vertices. Finally, to end up with one tree instead of a forest, repeatedly add edges between any degree-one-vertices of different connected components.

We now show that if I is a yes-instance then $opt(I') \geq t$ and if I is a no-instance then $opt(I') = 1$.

Suppose that I has a set cover of size h. Observe that for each element $e_i \in U$ the only vertex of degree $f(e_i)$ is v_i^e, and there are no other vertices violating the t-anonymous property. The key point in the construction is that, in order to get a t-anonymous graph, one has to delete vertices of $V_\mathcal{C}$. Indeed, let $e_i \in U$ be an element and v_j^S a root vertex such that $e_i \in S_j$. By construction the child vertices $v_\ell^{e_i,S_j}$ of v_j^S correspond to e_i and therefore have $f(e_i)$ child vertices. Thus, deleting v_j^S lowers the degree of all $v_\ell^{e_i,S_j}$ to $f(e_i)$ and, hence, v_i^e no longer violates the t-anonymous property. Hence, given a set cover of size h one can construct a corresponding t-deletion set for G.

Conversely, we show that if there exists a 2-deletion set of size at most h in G, then $(U, \mathcal{C}, h)$ is a yes-instance of Set Cover. Let $S \subseteq V$ be a 2-deletion set of size at most h. First, we show how to construct a 2-deletion set $S' \subseteq V_\mathcal{C}$ such that $|S'| \leq |S|$. To this end, initialize S' as $S' = S \cap V_\mathcal{C}$. If S' is a 2-deletion set, then the construction of S' is finished. Otherwise, there is a vertex v in $G - S'$ such that there is no other vertex with the same degree as v. Observe that since $S' \subseteq V_\mathcal{C}$, it follows that $v \in V_U$, that is $v = v_i^e$ for some $1 \leq i \leq m$. Furthermore, observe that is exactly one vertex in G having a degree d between $f(e_i) - h \leq d \leq f(e_i)$, namely v_i^e. As S is a 2-deletion set, it follows that S either contains v_i^e or a vertex u that is adjacent to a vertex w with $\deg_G(w) > \deg(v_i^e)$. In either case, we add to S' a vertex $v_j^S \in V_\mathcal{C}$ such that $e_i \in S_j$. By exhaustively applying this procedure, we end up with S' being a 2-deletion set. Since the vertices in $V_\mathcal{C}$ are the only ones in G that are adjacent to more than one vertex of degree at least three and all vertices in V_U have degree more than three, it follows that $|S'| \leq |S|$.

It remains to show that the set $\mathcal{C}'$ of sets corresponding to the vertices in S' forms a set cover. To this end, assume by contradiction that $\mathcal{C}'$ is not a set cover, that is, there is an element $e_i \notin \bigcup_{S_j \in \mathcal{C}'} S_j$. However, this implies that in $G - S'$ there is exactly one vertex of degree $f(e_i)$, namely v_i^e, implying that S' is not a 2-deletion set, a contradiction. As $|\mathcal{C}'| = |S'| \leq |S| \leq h$, it follows that if G contains a 2-deletion set of size h, then $(U, \mathcal{C}, h)$ is a yes-instance. Hence, if $(U, \mathcal{C}, h)$ is a no-instance, then there exist no 2-deletion set of size at most h.

Thus, we obtain a fpt gap-reduction with the gap $t = (mn)^{\frac{1-2\varepsilon}{2\varepsilon}} = (t^2m^2n^2)^{1/2-\varepsilon} \geq |V|^{1/2-\varepsilon}$ since $|V| < t^2m^2n^2$. From Lemma 1 and since Set Cover is W[2]-hard [7], we have that Max Anonym-V-Del is not $n^{1/2-\varepsilon}$-approximable in fpt-time w.r.t. parameter s, even on trees, unless FPT = W[2]. □

4 Inapproximability of edge deletion versions

In this section, we first show that Anonym-E-Del is NP-hard on caterpillars; the corresponding proof is an adaption of the reduction provided in the proof of Theorem 2. A caterpillar is a tree that has a dominating path [1], that is, a caterpillar is a tree such that deleting all leaves results in a path. Then we provide polynomial-time inapproximability results for Min Anonym-E-Del and Max Anonym-E-Del for bounded-degree graphs and parameterized inapproximability results for Max Anonym-E-Del on general graphs.

Theorem 3. Anonym-E-Del *is NP-hard on caterpillars, even if* $k = 2$.

Theorem 4. *For every* $0 < \varepsilon \leq 1$, Max Anonym-E-Del *is not* $n^{1-\varepsilon}$*-approximable even on bounded-degree graphs, unless P = NP.*

Theorem 5. *For every* $0 < \varepsilon \leq 1$, Min Anonym-E-Del *is not* $n^{1-\varepsilon}$*-approximable even on bounded-degree graphs, unless P = NP.*

Theorem 6. *For every* $0 < \varepsilon \leq 1$, Max Anonym-E-Del *is not* $n^{1-\varepsilon}$*-approximable in fpt-time w.r.t. parameter* s*, unless FPT = W[1].*

Proof. We provide an fpt gap-reduction from the W[1]-hard CLIQUE problem [7] parameterized by the solution size h. CLIQUE is defined as follows: given a graph $G = (V, E)$ and an integer $h \in \mathbb{N}$, the task is to decide whether there is a subset $V' \subseteq V$ of at least h pairwise adjacent vertices. Let $I = (G, h)$ be an instance of CLIQUE. Assume w.l.o.g. that $\Delta_G + 2h + 1 \leq n$, where $n = |V|$. If this is not the case, then one can add isolated vertices to G until the bound holds.

We construct an instance $I' = (G' = (V', E'), s)$ of MAX ANONYM-E-DEL as follows: First, copy G into G'. Then, add a vertex u and connect it to the n vertices in G'. Next, for each vertex $v \in V$ add to G' degree-one vertices that are adjacent only to v such that $\deg_{G'}(v) = n - h$. This is always possible, since we assumed $\Delta_G + 2h + 1 \leq n$. Observe that in this way at most $n(n-h)$ degree-one vertices are added. Now, set $x := \lceil (4n)^{3/\varepsilon} \rceil$ and add cliques with two, $n - 2h + 1$, and $n - h + 1$ vertices such that after adding these cliques the number of degree-d vertices in G', for each $d \in \{1, n-2h, n-h\}$, is between $x + h$ and $x + h + n$, that is, $x + h \leq |D_{G'}(d)| \leq x + h + n$. After inserting these cliques, the graph consists of four blocks: of degree one, $n - h$, $n - 2h$, and n, where the first three blocks are roughly of the same size (between $x + h$ and $x + h + n$ vertices) and the last block of degree n contains exactly one vertex. To finish the construction, set $s := \binom{h}{2} + h$.

Now we show that if I is a yes-instance, then $opt(I') \geq x$ and if I is a no-instance, then $opt(I') < 2s$.

Suppose that I contains a clique $C \subseteq V$ of size h. Then, deleting the $\binom{h}{2}$ edges within C and the h edges between the vertices in C and u does not exceed the budget s and results in an x-anonymous graph G''. Since h edges incident to u are deleted, it follows that $\deg_{G''}(u) = n - h$. Furthermore, for each clique-vertex $v \in C$ also h incident edges are deleted ($h - 1$ edges to other clique-vertices and the edge to u), thus it follows that $\deg_{G''}(v) = n - 2h$. Since the degree of the remaining vertices remain unchanged, and $|D_{G'}(n-h)| \geq x + h$, it follows that each of the three blocks in G'' has size at least x. Hence, G'' is x-anonymous.

For the reverse direction, suppose that there is a $2s$-deletion set S of size at most s in G'. Since u is the only vertex in G' with degree n, and all other vertices in G' have degree at most $n - h$, it follows that S contains at least h edges that are incident to u. Since $N_{G'}(u) = V$, it follows that the degree of at least h vertices of the block $D_{G'}(n-h)$ is decreased by one. Denote these vertices by C. Since $|S| \leq s$ and h edges incident to u are contained in S, it follows that at most $2s - h + 1$ vertices are incident to an edge in S. Furthermore, since S is a $2s$-deletion set, it follows that the vertices in C are in $G' - S$ either contained in the block of degree one or in the block of degree $n - 2h$. Thus, by deleting the at most $\binom{h}{2}$ remaining edges in S, the degree of each of the h vertices in C is decreased by at least $h - 1$. Hence, these $\binom{h}{2}$ edges in S form a clique on the vertices in C and thus I is a yes-instance. Therefore, it follows that if I is a no-instance, then there is no $2s$-deletion set of size s in G' and hence $opt(I') < 2s$.

Thus we obtain a gap-reduction with the gap at least $\frac{x}{2s}$. Set $n' := |V'|$. By construction we have $3x \leq n' \leq n^2 + 3x + 3h + 3n + 1$. By the choice of x it follows that $x > n'/4$, since

$$\frac{n'}{4} \leq \frac{1}{4}(n^2 + 3x + 3h + 3n + 1) = x + \underbrace{\frac{1}{4}(n^2 + 3h + 3n + 1 - x)}_{<0} < x.$$

Hence the gap is

$$\frac{x}{2s} > \frac{n'^{1-\varepsilon+\varepsilon}}{4(h^2 + h)} \geq n'^{1-\varepsilon} \frac{n'^{\varepsilon}}{8h^2} > n'^{1-\varepsilon} \frac{x^{\varepsilon}}{8n^2} = n'^{1-\varepsilon} \frac{(4n)^{3\varepsilon/\varepsilon}}{8n^2} > n'^{1-\varepsilon}.$$

□

5 Fixed-Parameter Tractability

In previous work, it was shown that Anonym-E-Ins and Anonym-V-Del are both fixed-parameter tractable with respect to the combined parameter budget s and maximum degree Δ [2, 11]. Here we generalize the ideas behind these results and show fixed-parameter tractability for the general problem variant where one might insert and delete specified numbers of vertices and edges.

k-Degree Anonymity Editing (Anonym-Edt)
Input: An undirected graph $G = (V, E)$ and five positive integers s_1, s_2, s_3, s_4 and k.
Question: Is it possible to obtain a graph $G' = (V', E')$ from G using at most s_1 vertex deletions, s_2 vertex insertions, s_3 edge deletions, and s_4 edge insertions, such that G' is k-anonymous?

Observe that here we require that the inserted vertices have degree zero and we have to "pay" for making the inserted vertices adjacent to the existing ones. In particular, if $s_4 = 0$, then all inserted vertices are isolated in the target graph. Note that there are other models where the added vertices can be made adjacent to an arbitrary number of vertices [3, 5]. Our ideas, however, do not directly transfer to this variant.

For convenience, we set $s := s_1 + s_2 + s_3 + s_4$ to be the number of allowed graph modifications.

Theorem 7. Anonym-Edt *is fixed-parameter tractable w.r.t.* (s, Δ).

Proof (sketch). Let $I = (G = (V, E), k, s_1, s_2, s_3, s_4)$ be an instance of Anonym-Edt. In the following we give an algorithm finding a solution if existing. Intuitively, the algorithm first guesses a "solution structure" and then checks whether the graph modifications associated to this solution structure can be performed in G. A solution structure is a graph S with at most $s(\Delta + 1)$ vertices where

1. each vertex is colored with a color from $\{0, \ldots, \Delta\}$ indicating the degree of the vertex in G and
2. each edge and each vertex is marked either as "to be deleted", "to be inserted", or "not to be changed" such that:
   (a) all edges incident to a vertex marked as "to be inserted" are also marked as "to be inserted",
   (b) at most s_1 vertices and at most s_3 edges are marked as "to be deleted", and
   (c) at most s_2 vertices and at most s_4 edges are marked as "to be inserted".

The intuition about this definition is that a solution structure S contains all graph modifications in a solution *and* the vertices that are affected by the modifications, that is, the vertices whose degree is changed when performing these modifications. Observe that any solution for I defines such a solution structure with at most $s(\Delta + 1)$ vertices as each graph modification affects at most $\Delta + 1$ vertices. This bound is tight in the sense that deleting a vertex v affects v and his up to Δ neighbors. Furthermore, observe that once given such a solution structure, we can check in polynomial time whether performing the marked edge/vertex insertions/deletions results in a k-anonymous graph G', since the coloring of the vertex indicates the degrees of the vertices that are affected by the graph modifications.

Our algorithm works as follows: First it branches into all possibilities for the solution structure S. In each branch it checks whether performing the graph modifications indicated by the marks in S indeed result in a k-anonymous graph. If yes, then the algorithm checks whether the graph modifications associated to S can be performed in G: To this end, all edges and vertices marked as "to be inserted" are removed from S and the marks at the remaining vertices and edges are also removed and the resulting "cleaned" graph is called S'. Finally the algorithm tries to find S' as an induced subgraph of G such that the vertex degrees coincide with the vertex-coloring in S'. This is done by a meta-theorem for bounded local tree-width graphs [8]. If the algorithm succeeds and finds S' as an induced subgraph, then the graph modifications encoded in S can be performed which proves that I is a yes-instance. If the algorithm fails in every branch, then, due to the exhaustive search over all possibilities for S, it follows that I is a no-instance. Thus, the algorithm is indeed correct.

6 Conclusion

We have shown strong inapproximability results for the optimization variants of Anonym-E-Del and Anonym-V-Del. We leave two major open questions concerning polynomial-time approximability and parameterized approximability: In all our gap reductions the value of k is in the order of n. This leads to the question whether with constant k Min Anonym-E-Del or Min Anonym-V-Del are constant-factor approximable in polynomial time? Second, we failed to transfer the inapproximability results to Anonym-E-Edt where we require that the number of edge insertions plus deletions is at most s. Here, handling the possibility to *revert* already changed degrees seems to be crucial in order to obtain any approximation result (positive or negative) for the optimization variants of Anonym-E-Edt. This leads to the question whether there are "reasonable" (parameterized) approximation algorithms for the optimization variants of Anonym-E-Edt?

Bibliography

[1] A. Brandstädt, V. B. Le, and J. P. Spinrad. *Graph Classes: a Survey*, volume 3 of *SIAM Monographs on Discrete Mathematics and Applications*. SIAM, 1999.

[2] R. Bredereck, S. Hartung, A. Nichterlein, and G. J. Woeginger. The complexity of finding a large subgraph under anonymity constraints. In *Proc. 24th ISAAC*, volume 8283 of *LNCS*, pages 152–162. Springer, 2013.

[3] R. Bredereck, V. Froese, S. Hartung, A. Nichterlein, R. Niedermeier, and N. Talmon. The complexity of degree anonymization by vertex

addition. In *Proc. 10th AAIM*, volume 8546 of *LNCS*, pages 44–55. Springer, 2014.

[4] S. Chester, B. Kapron, G. Srivastava, and S. Venkatesh. Complexity of social network anonymization. *Social Netw. Analys. Mining*, 3(2): 151–166, 2013.

[5] S. Chester, B. M. Kapron, G. Ramesh, G. Srivastava, A. Thomo, and S. Venkatesh. Why Waldo befriended the dummy? k-anonymization of social networks with pseudo-nodes. *Social Netw. Analys. Mining*, 3(3):381–399, 2013.

[6] K. L. Clarkson, K. Liu, and E. Terzi. Towards identity anonymization in social networks. In *Link Mining: Models, Algorithms, and Applications*, pages 359–385. Springer, 2010.

[7] R. G. Downey and M. R. Fellows. *Fundamentals of Parameterized Complexity*. Springer, 2013.

[8] M. Frick and M. Grohe. Deciding first-order properties of locally tree-decomposable structures. *J. ACM*, 48(6):1184–1206, 2001.

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

[11] S. Hartung, A. Nichterlein, R. Niedermeier, and O. Suchý. A refined complexity analysis of degree anonymization on graphs. In *Proc. 40th ICALP*, volume 7966 of *LNCS*, pages 594–606. Springer, 2013.

[12] S. Hartung, C. Hoffmann, and A. Nichterlein. Improved upper and lower bound heuristics for degree anonymization in social networks. In *Proc. 13th SEA*, volume 8504 of *LNCS*, pages 376–387. Springer, 2014.

[13] K. Liu and E. Terzi. Towards identity anonymization on graphs. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '08)*, pages 93–106. ACM, 2008.

[14] D. Marx. Parameterized complexity and approximation algorithms. *Comput J*, 51(1):60–78, 2008.

[15] X. Wu, X. Ying, K. Liu, and L. Chen. A survey of privacy-preservation of graphs and social networks. In *Managing and mining graph data*, pages 421–453. Springer, 2010.

A Proofs

A.1 Proof 1 (Theorem 1)

Proof. We establish a gap-reduction from the NP-hard Vertex Cover problem to Min Anonym-V-Del. Given a graph G and a positive integer h, the task in the Vertex Cover is to decide whether there are at most h vertices such that each edge in G is incident to at least one of the h vertices [10, GT1].

Now, given an instance $I = (G = (V, E), h)$ of Vertex Cover, we construct an instance $I' = (G' = (V', E'), k)$ of Min Anonym-V-Del as follows: G' contains a copy of the graph G, an independent set IS of size x, and a complete graph of size $n + x - h$, where x is an integer greater than or equal to $n^{\frac{1}{\varepsilon}} 2^{\frac{1-\varepsilon}{\varepsilon}} - n$. Further, set $k := n + x - h$. Denote by n' the number of vertices of G', thus $n' = 2n + 2x - h$.

If G contains a vertex cover S of size h, then S is a k-deletion set for G' since removing S from G' results in a graph with two blocks, one of degree 0 and one of degree $n + x - h - 1$. Thus $opt(I') \leq h$.

If G has no vertex cover of size h, in order to obtain a graph that is k-anonymous, we have to remove at least $n + x$ vertices corresponding to vertices of G and IS. In this case $opt(I') = n + x$.

Thus we obtain a gap-reduction with the gap at least $\frac{n+x}{h} \geq (n + x)^{1-\varepsilon} 2^{1-\varepsilon} \geq (n')^{1-\varepsilon}$. □

A.2 Proof 2 (Theorem 3)

Proof. We provide a polynomial-time reduction from Exact Cover by 3-Sets. Let $I = (U, \mathcal{C}, h)$ be an instance of Exact Cover by 3-Sets. We assume without loss of generality that for each element $e_i \in U$ there exists a set $S_j \in \mathcal{C}$ with $e_i \in S_j$. To reduce the amount of indices in the construction given below we introduce the function $f\colon U \to \mathbb{N}$ that maps an element $e_i \in U$ to $f(e_i) = (2h + 3)i$.

The instance I' of Anonym-E-Del is defined on a graph $G = (V, E)$ constructed as follows. For each element $e_i \in U$ add a star $K_{1,f(e_i)}$ with the center vertex v_i^e. Denote with $V_U = \{v_1^e, \ldots, v_{3h}^e\}$ the set of all these center vertices. Furthermore, for each element $e_i \in U$ add two stars $K_{1,f(e_i)+1}$ and two stars $K_{1,f(e_i)+2}$.

For each set $S_j \in \mathcal{C}$ with $S_j = \{e_\alpha, e_\beta, e_\gamma\}$ add a *set-gadget* consisting of the stars $K_{1,f(e_\alpha)}$, $K_{1,f(e_\beta)}$, and $K_{1,f(e_\gamma)}$. Denote with v_α^j, v_β^j, and v_γ^j the center vertices of these stars and denote with $V_{\mathcal{C}}$ the set of all these center vertices, formally $V_{\mathcal{C}} = \{v_i^j \mid 1 \leq i \leq 3h \wedge 1 \leq j \leq n \wedge e_i \in S_j\}$. Next, add the edges $\{v_\alpha^j, v_\beta^j\}$ and $\{v_\beta^j, v_\gamma^j\}$ to E. Observe that $\deg(v_\alpha^j) = f(e_\alpha) + 1$, $\deg(v_\beta^j) = f(e_\beta) + 2$, and $\deg(v_\gamma^j) = f(e_\gamma) + 1$. To end up with one caterpillar instead of a forest of caterpillars, do the following exhaustively: Take two different connected components (caterpillars) C_1 and C_2, let v_1 be an endpoint of a dominating path in C_1, and let v_2 be an endpoint of the dominating path in C_2, such that $\deg_G(v_1) = \deg_G(v_2) = 1$. Then, add the edge $\{v_1, v_2\}$ to reduce the number of connected components by one. The resulting graph is clearly a caterpillar. We complete the construction of I' by setting $s = 2h$ and $k = 2$.

We now prove that I is a yes-instance of Exact Cover by 3-Sets if and only if $I' = (G, k, s)$ is a yes-instance of Anonym-E-Del.

Let $\mathcal{C}' \subseteq \mathcal{C}$ be an exact cover of size h. Then we construct a 2-deletion set $S \subseteq E$ of size $2h$ as follows: For each set $S_j \in \mathcal{C}'$ with $S_j = \{e_\alpha, e_\beta, e_\gamma\}$ insert the edges $\{v_\alpha^j, v_\beta^j\}$ and $\{v_\beta^j, v_\gamma^j\}$ into S. First, observe that $|S| = 2h$. Next, we show that S is indeed a 2-deletion set. Suppose towards a contradiction that there exist a vertex $v \in V$ such that there is no further vertex of the same degree in $G - S$. Then, by construction of G, it follows that $v = v_i^e \in V_U$ for some $i \in \{1, \ldots, 3h\}$ and, by construction of S, it follows that $e_i \notin \bigcup_{S_j \in \mathcal{C}'} S_j$, a contradiction.

Let S be a 2-deletion set of edges of size at most $2h$. Observe that the only vertices in G that violate the 2-anonymous property are the vertices in V_U. Furthermore, for each $e_i \in U$ there is exactly one vertex in G with a degree d between $f(e_i) - 2h \leq d \leq f(e_i)$, namely v_i^e. Since S is a 2-deletion set, it follows that for each $v_i^e \in V_U$ there is a vertex $v \in V(S)$ having the same degree as v_i^e in $G - S$. Since $|V_U| = 3h$ and $|\deg(v_i^e) - \deg(v_{i'}^e)| > 2h$ for all $i, i' \in \{1, \ldots, 3h\}$, it follows that $|V(S)| \geq 3h$. For the further argumentation we need some notation. A vertex $v \in V$ is a *type-ℓ* vertex, $\ell \in \mathbb{N}$, if there exist a vertex $v_i^e \in V_U$ such that $\deg_G(v) = \deg_G(v_i^e) + \ell$. Now, observe that in G the type-1 vertices are all pairwise non-adjacent. Since $|V(S)| \geq 3h$, this implies that $V(S)$ contains $2h$ type-1 vertices and h type-2 vertices and that $|V(S)| = 3h$. Thus, for each edge in S it follows that one endpoint is a type-1 vertex and the other endpoint a type-2 vertex. Note that the only edges fulfilling this requirement are the ones making two vertices in $V_{\mathcal{C}}$ adjacent and, thus, $V(S) \subseteq V_{\mathcal{C}}$. Thus, each type-2 vertex of $V(S)$ is contained in some set-gadget. Denote with $\mathcal{C}'$ the set of h sets corresponding to the set-gadgets that contain the h type-2 vertices in $V(S)$. We now prove that $\mathcal{C}'$ is an exact cover. Suppose towards a contradiction that there is an element $e_i \notin \bigcup_{S_j \in \mathcal{C}'} S_j$. This implies, that no vertex v_i^j such that $j \in \{1, \ldots, n\}$ and $e_i \in S_j$ is contained in $V(S)$. However, as $V(S) \subseteq V_{\mathcal{C}}$, this means that v_i^e has a unique degree in $G - S$, a contradiction to the fact that S is a 2-deletion set. Finally, since $|\mathcal{C}'| = h$, $\bigcup_{S_j \in \mathcal{C}'} S_j = U$, each set contains exactly three elements, and $|U| = 3h$, it follows that no element is covered twice. Hence, $\mathcal{C}'$ is an exact cover and, thus, I is a yes-instance. □

A.3 Proof 3 (Theorem 4)

Proof. We provide a gap-reduction from Exact Cover by 3-Sets proved NP-hard even when no element occurs in more than three subsets [10, SP2]. For these instances we have $h \leq n \leq 3h$.

Let $I = (U, \mathcal{C}, h)$ be an instance of Exact Cover by 3-Sets where no element occurs in more than three subsets. We construct an instance $I' = (G, s)$ of Max Anonym-E-Del as follows. The graph G contains a vertex u_i for each element e_i from U and a vertex c_j for each subset S_j from $\mathcal{C}$. There is an edge in G between u_i and c_j if S_j contains e_i. For each vertex c_j we add four degree-one vertices that are adjacent to c_j, thus the degree of each vertex c_j is 7. For each vertex u_i we add up to three degree-one vertices that are adjacent to u_i such that the degree of u_i is 3 (observe that each element occurs in at most three sets). Let x be an integer greater than or equal to $(3h40^{1-\varepsilon})^{1/\varepsilon}$. We add in G x stars $K_{1,7}$, x stars $K_{1,4}$ and x stars $K_{1,2}$. Consider $s = 3h$. Thus graph G has $n + x$ vertices of degree 7, x of degree 4, $3h$ of degree 3, x of degree 2 and many vertices of degree 1, and so G is $3h$-anonymous. The number of vertices in G is $n' \geq 16x + 3h + 5n$.

We now show that if I is a yes-instance then $opt(I') \geq x$ and if I is a no-instance then $opt(I') = 3h$.

Suppose that I contains an exact cover $\mathcal{C}' \subseteq \mathcal{C}$ of size h. Then removing from G the $3h$ edges between $c_j \in \mathcal{C}'$ and $u_i \in U$, we obtain a graph G' that is x-anonymous, since all vertices from the block of degree 3 from G are in G' in the block of degree 2.

Suppose now that I does not contain any exact cover of size h. There are two possibilities to try to increase the anonymity of the graph G. The first one consists of moving some vertices from the block of degree 4 to the block of degree 3, for this we have to remove $3h$ edges from the stars $K_{1,4}$ but then we have $3h$ vertices of degree 0. The other possibility is to move all vertices from the block of degree 3 to the block of degree 2, for this we have to remove an edge incident to each vertex u_i. Since there is no exact cover in I, removing such a set of edges will create some new blocks of degree 5 or/and 6. So in this case $opt(I') \leq 3h$.

Thus we obtain a gap-reduction with the gap at least $\frac{x}{3h} \geq 8^{1-\varepsilon}(5x)^{1-\varepsilon} \geq 8^{1-\varepsilon}(2x+3h)^{1-\varepsilon} = (16x+24h)^{1-\varepsilon} \geq (n')^{1-\varepsilon}$, where for the last inequalities we use that $x \geq h$, $h \leq n \leq 3h$, and $16x + 9h + 5n \leq n'$. □

A.4 Proof 4 (Theorem 5)

Proof. We provide a gap-reduction from EXACT COVER BY 3-SETS to MIN ANONYM-E-DEL. Let $I = (U, \mathcal{C}, h)$ be an instance of EXACT COVER BY 3-SETS where no element occurs in more than three subsets. We provide an instance $I' = (G, k)$ of MIN ANONYM-E-DEL where the graph is constructed as in the proof of Theorem 4 and $k := x$.

We now show that if I is a yes-instance then $opt(I') = 3h$ and if I is a no-instance then $opt(I') \geq x$.

Suppose that I contains an exact cover $\mathcal{C}' \subseteq \mathcal{C}$ of size h. Then removing from G the $3h$ edges between $c_j \in \mathcal{C}'$ and $u_i \in U$, we obtain a graph G' that is x-anonymous, since all vertices from the block of degree 3 from G are in G' in the block of degree 2.

Suppose now that I does not contain any exact cover of size h. We can prove, as in Theorem 4, that in order to obtain a k-anonymous graph, we have to remove x edges from the stars $K_{1,4}$ (one edge by star). Thus $opt(I') \geq x$.

Thus we obtain a gap-reduction with the gap at least $\frac{x}{3h} \geq (n')^{1-\varepsilon}$. □

A.5 Proof 5 (Theorem 7)

Proof. Let $I = (G = (V, E), k, s_1, s_2, s_3, s_4)$ be an instance of ANONYM-EDT. In the following we give an algorithm finding a solution if existing. Intuitively, the algorithm first guesses a "solution structure" and then checks whether the graph modifications associated to this solution structure can be performed in G. A solution structure is a graph S with at most $s(\Delta + 1)$ vertices where

1. each vertex is colored with a color from $\{0, \ldots, \Delta\}$ indicating the degree of the vertex in G and
2. each edge and each vertex is marked either as "to be deleted", "to be inserted", or "not to be changed" such that:
   (a) all edges incident to a vertex marked as "to be inserted" are also marked as "to be inserted",

(b) at most s_1 vertices and at most s_3 edges are marked as "to be deleted", and
(c) at most s_2 vertices and at most s_4 edges are marked as "to be inserted".

The intuition about this definition is that a solution structure S contains all graph modifications in a solution *and* the vertices that are affected by the modifications, that is, the vertices whose degree is changed when performing these modifications. Observe that any solution for I defines such a solution structure with at most $s(\Delta + 1)$ vertices as each graph modification affects at most $\Delta + 1$ vertices. This bound is tight in the sense that deleting a vertex v affects v and his up to Δ neighbors. Furthermore, observe that once given such a solution structure, we can check in polynomial time whether performing the marked edge/vertex insertions/deletions results in a k-anonymous graph G', since the coloring of the vertex indicates the degrees of the vertices that are affected by the graph modifications.

Our algorithm works as follows: First it branches into all possibilities for the solution structure S. In each branch it checks whether performing the graph modifications indicated by the marks in S indeed result in a k-anonymous graph. If yes, then the algorithm checks whether the graph modifications associated to S can be performed in G: To this end, all edges and vertices marked as "to be inserted" are removed from S and the marks at the remaining vertices and edges are also removed and the resulting "cleaned" graph is called S'. Finally the algorithm tries to find S' as an induced subgraph of G such that the vertex degrees coincide with the vertex-coloring in S'. If the algorithm succeeds and finds S' as an induced subgraph, then the graph modifications encoded in S can be performed which proves that I is a yes-instance. If the algorithm fails in every branch, then, due to the exhaustive search over all possibilities for S, it follows that I is a no-instance. Thus, the algorithm is indeed correct.

As to the running time: There are $s(\Delta+1)$ possibilities for the number of vertices in the solution structure. Hence, there are at most $s(\Delta + 1) \cdot 2^{\binom{s(\Delta+1)}{2}} < 2^{(s(\Delta+1))^2}$ graphs. Furthermore, there are at most $(\Delta+1)^{s(\Delta+1)}$ possibilities to color the vertices and $3^{s(\Delta+1)+\binom{s(\Delta+1)}{2}}$ possibilities to mark the vertices and edges. Overall, the algorithm branches into $2^{O((s\Delta)^2)}$ possibilities for the solution structure S. As mentioned above, checking whether performing the graph modifications indicated by S indeed results in a k-anonymous graph can be done in polynomial time.

Next, the algorithm checks for each S that may lead to a k-anonymous graph whether the cleaned graph S' occurs as an induced subgraph in G such that degree constraints given by the vertex coloring are fulfilled. Observe that since our input graph G has maximum degree Δ it also has a local tree-width of at most Δ [8]. Thus, for finding S' as induced subgraph, we can use a general result of Frick and Grohe [8, Theorem 1.2] showing that deciding whether a graph H of local tree-width at most ℓ satisfies a property ϕ definable in first-order logic is fixed-parameter tractable with respect to the combined parameter $(|\phi|, \ell)$. The subgraph isomorphism problem can be solved with this result on graphs with bounded local tree-width [8]. Thus it remains to specify the part of the formula ϕ that ensures the degree constraints. To this end, Frick and Grohe [8] gave as example the formula

$$x \in V \wedge \neg\exists y \exists z(\neg(y = z) \wedge (x, y) \in E \wedge (x, z) \in E)$$

to express that a vertex $x \in V$ has degree at most one. This formula can be extended to express that $x \in V$ has degree at most ℓ for some $1 \leq \ell \leq \Delta$ and the size of the formula is bounded in a function of Δ. Similarly, removing the first negation symbol yields the statement $x \in V$ has a degree of at least two (or at least $\ell + 1$ in the extended version). Hence, we can express the degree constraints and the formula-size is still bounded by a function of s and Δ (as there are up to $s(\Delta + 1)$ vertices in S'). Hence, applying the results of Frick and Grohe [8], shows that the overall algorithm runs in fpt-time w.r.t. to (s, Δ). □