k-Metric Antidimension: a Privacy Measure for Social Graphs

Rolando Trujillo-Rasua[a,*], Ismael G. Yero[b]

[a] *Interdisciplinary Centre for Security, Reliability and Trust*
University of Luxembourg
6, rue Richard Coudenhove-Kalergi L-1359 Luxembourg
[b] *Departamento de Matemáticas, Escuela Politécnica Superior de Algeciras*
Universidad de Cádiz, Spain
Av. Ramón Puyol s/n, 11202 Algeciras, Spain

Abstract

The study and analysis of social graphs impacts on a wide range of applications, such as community decision making support and recommender systems. With the boom of *online* social networks, such analyses are benefiting from a massive collection and publication of social graphs at large scale. Unfortunately, individuals' privacy right might be inadvertently violated when publishing this type of data. In this article, we introduce (k, ℓ)-anonymity; a novel privacy measure aimed at evaluating the resistance of social graphs to active attacks. (k, ℓ)-anonymity is based on a new problem in Graph Theory, the *k-metric antidimension* defined as follows.

Let $G = (V, E)$ be a simple connected graph and $S = \{w_1, \cdots, w_t\} \subseteq V$ an ordered subset of vertices. The metric representation of a vertex $u \in V$ with respect to S is the t-vector $r(u|S) = (d_G(u, w_1), \cdots, d_G(u, w_t))$, where $d_G(u, v)$ represents the length of a shortest $u - v$ path in G. We call S a k-antiresolving set if k is the largest positive integer such that for every vertex $v \in V - S$ there exist other $k - 1$ different vertices $v_1, \cdots, v_{k-1} \in V - S$ with $r(v|S) = r(v_1|S) = \cdots = r(v_{k-1}|S)$. The k-metric antidimension of G is the minimum cardinality among all the k-antiresolving sets for G.

We address the k-metric antidimension problem by proposing a true-biased algorithm with success rate above 80% when considering random graphs of size at most 100. The proposed algorithm is used to determine the privacy guarantees offered by two real-life social graphs with respect to (k, ℓ)-anonymity. We also investigate theoretical properties of the k-metric antidimension of graphs. In particular, we focus on paths, cycles, complete bipartite graphs and trees.

Keywords: anonymity, active attack, social network, graph, resolving set, k-metric antidimension

1. Introduction

Social networking services are widely used in modern society as illustrated by the Alexa's Top 500 Global Sites statistics[1] where *facebook* and *linkedin* rank 2nd and 11th respectively in 2014.

*Corresponding author. Phone: +352 466 644 5458. Fax: 466 644 3 5458. Email: rolando.trujillo@uni.lu
[1] http://www.alexa.com/topsites

Such popularity has enabled governments and third-party enterprises to massively collect social network data, which eventually can be released[2] for mining and analysis purposes.

The power of social network analysis is questionless. It might uncover previously unknown knowledge such as community-based problem, media use, individual engagement, amongst others. Sociology is a trivial example of a field that certainly benefits from social graphs publication. Many other fields (*e.g.,* economics, geography, or political science) and systems (*e.g.,* service-oriented systems, advertisers, or recommended systems) improve their decisions, processes, and services, based on users interaction.

However, all these benefits are not cost-free. An adversary can compromise users privacy using the published social network, which results in the disclosure of sensitive data such as e-mails, instant messages, or relationships. A simple and popular approach to prevent this privacy problem is *anonymization* by means of removing potential identifying attributes. Doing so, aggregate knowledge still can be inferred (*e.g.,* connectivity, distance, or node degrees) while the "who" information has been removed. In practice, however, this naive approach is not enough for protecting users' privacy.

What makes social network anonymization a challenging problem is the combination of the adversary's background knowledge with the released structure of the network. Considering a social network as a simple graph, in which individuals are represented by vertices and their bidirectional relationships by edges, the adversary's background knowledge about a victim may take many forms, *e.g.,* vertex degrees, connectivity, or local neighborhood. This structural knowledge, together with the released graph, is often enough to perform *passive attacks* where the users and their relationships are re-identified [12].

Other privacy attacks exist. In 2007, Backstrom *et al.* [1] introduced *active attacks* based on the creation and insertion in the network of *attacker nodes* controlled by the adversary. The attacker nodes could be either new accounts with pseudonymous or spoofed identities (Sybil nodes), or legitimate users in the network who collude with the adversary. Attacker nodes establish links with other nodes in the network (also between themselves) aiming at creating a sort of fingerprint in the network. Once the social graph is released, the adversary just need to retrieve such a fingerprint (the attacker nodes) and use it as a hub to re-identify other nodes in the network. Backstrom et. al. proved that $O(\sqrt{\log n})$ attacker nodes in the network can compromise the privacy of arbitrary targeted nodes with high probability, which makes active attack particularly dangerous.

1.1. Contribution and plan of the article

Several active attacks to social graphs have been proposed. They could even target random nodes in the network as recently shown in [15]. However, to the best of our knowledge, no privacy measure aimed at evaluating the resistance of a social graph to this kind of attack exists. The lack of such a measure prevents the development of privacy-preserving methods with theoretically proven privacy guarantees.

In this article we define (k,ℓ)-*anonymity*; a privacy measure that can be applied to real-life social graphs in order to measure their resistance to active attacks. The proposed privacy

[2] See for example http://snap.stanford.edu/data/

measure copes with adversaries whose background knowledge concerning a node u and a subset S of attacker nodes is the metric representation of u with respect to S. (k,ℓ)-anonymity turns out to be based on a new problem in Graph Theory: the *k-metric antidimension*. Recognizing the hardness of the k-metric antidimension problem (is NP-complete for $k = 1$ [7]), we propose a true-biased algorithm whose computational complexity and success rate can be adjusted. Empirical results show that our algorithm finds k-antiresolving basis in random graphs of order at most 100 with a success rate above 80%. Our algorithm has been also used to determine the privacy offered by two real-life social graphs against active attacks. Finally, we provide theoretical results on the k-metric antidimension of graphs, such as paths, cycles, complete bipartite graphs and trees.

The rest of this article is structured as follows. Section 2 briefly reviews the literature on privacy-preserving publication of social network data. Section 3 presents the metric representation as a reasonable definition of the adversary's background knowledge. It also introduces the k-metric antidimension as the basis for the privacy measure (k,ℓ)-anonymity. In Section 4 we present a true-biased algorithm for computing the k-metric antidimension of a graph, and evaluate the proposed algorithm through experiments. Preliminary results (mathematical properties) on the new problem (the k-metric antidimension) are provided in Sections 5 and 6 (the later specifically addresses the case of tree graphs). Section 7 draws conclusions and future work.

2. Related work

A social graph $G = (V,E)$ is a simple graph where V represents the set of social actors and $E \subseteq V \times V$ their relationships. Both vertices and edges could be enriched with attribute values such as weights representing trustworthiness or labels providing meaning. We consider, however, social network data in its most "simplest" form, *i.e.*, a simple graph without further annotation.

Privacy breaches in social networks are mainly categorized in identity disclosure or link disclosure [22]. To perform such attacks adversaries rely on background knowledge, which is usually defined as structural knowledge such as vertex degrees [9] or neighborhoods [28]. The assumptions on the adversary's background knowledge determine the type of privacy attacks and the corresponding countermeasures.

Privacy-preserving methods for the publication of social graphs are normally based on the well-known concept k-anonymity [16] adapted to graphs. k-anonymity, initially proposed for microdata, aims at ensuring that no record in a database can be re-identified with probability higher than $1/k$. To do so, identifying attributes should be obviously removed, and any combination of non-identifying attribute values should not be unique in the database. In practice, not all the attributes need to be combined, because they do not belong to the adversary's knowledge. This leads to the concept of *quasi-identifier*, that is, an attribute that can be found in external source of information and, combined with other quasi-identifiers, can uniquely identify a record in the database.

Even though graphs can be represented in tabular form and, thus, graph k-anonymity can be defined in terms of quasi-identifying attributes [18], graph k-anonymity is typically defined in terms of structural properties of the graph rather than on attributes. For instance in [6], the adversary's background knowledge is defined as a *knowledge query* $Q(x)$ evaluated for a given target node of the original graph G. The knowledge query $Q(x)$ allows the creation of a candidate set consisting of $\{y \in V | Q(x) = Q(y)\}$. In other words, all the nodes in the network matching the query $Q(.)$ are equally likely to be the target node x. This simple concept is the basis of several passive attacks and privacy-preserving methods in the publication of social graphs [9, 28, 30].

Other privacy notions based on entropy rather than on k-anonymity have been proposed [3]. This type of privacy measure is better suited for methods based on random addition, deletion, or switching of edges. The perturbation could be made in such a way that the number of edges or the degree of the vertices are preserved [25, 26]. However, empirical results obtained in [24, 26] suggest that random obfuscation poorly preserves the topological features of the network.

Passive attacks to social networks can be combined with active attacks. In addition to structural knowledge, in an active attack the adversary manages to control a subset of nodes (attacker nodes) of the original graph G [1]. The attacker nodes aim at creating links with their victims by either identity theft or cloning of existing users profiles [2]. They also establish links between themselves so as to build a subgraph H of attacker nodes with the following properties: i) H can be efficiently identified in G and ii) H does not have a non-trivial automorphisms. Once H has been identified, the adversary is able to re-identify neighbor nodes of H [1] or even arbitrary nodes in the network [12, 15].

Performing active attacks is not easy, given that there exist several detection mechanisms of attacker or Sybil nodes in a network [27]. However, such defenses strongly depend on assumptions on the topological structure of the social network, which does not hold in many real-world scenarios [11]. Actually, recent works aim at mitigating, instead of preventing, the impact of Sybil attacks [20]. Furthermore, a group of users who collude in order to breach the privacy of other users in the network can be also regarded as attacker nodes.

Other types of active attacks exist. For instance, the maximal vertex coverage (MVC) attack consists in attacking a few nodes so as to delete as many edges of the network as possible. In this attack, the attacker tries to convince some users to leave the social network in order to reduce the number of residual social ties. Metrics to quantify the impact of MVC attacks have been studied in [8]. MVC is not a privacy attack, though.

While there exist several published active attacks to social graphs, there does not exist yet a rational privacy metric for evaluating the resistance of social graphs to this type of privacy attack. To overcome this problem, in this article we introduce (k, ℓ)-anonymity; a privacy notion based on k-anonymity and the metric representation of nodes in a graph. Note that, privacy notions with the same name has been already proposed. For instance, Feder and Nabar proposed (k, ℓ)-anonymity where ℓ represents the number of common neighbors of two nodes [4]. This notion was later generalized by Stokes and Torra in [18]. In our privacy notion, however, ℓ represents an upper bound on the k-metric antidimension of the graph.

Interested readers could refer to [13, 22, 29] for further reading on privacy-preserving publication of social graphs.

3. Privacy against active attacks

In this section we define the metric representation of nodes with respect to a set of attacker nodes S as the adversary's background knowledge. We also introduce the resulting privacy measure, named (k, ℓ)-anonymity, and its related mathematical problem: the k-metric antidimension.

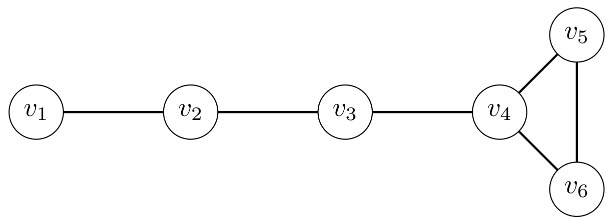


Figure 1: An example.

3.1. Adversary's background knowledge

Vulnerabilities in an anonymized social graphs are better understood once the adversary's knowledge has been properly modeled. This knowledge can be acquired from public information sources and through malicious actions. In practice, the adversary could even be a close friend, which makes the publication of social network where users cannot re-identify themselves a reasonable privacy goal.

Adversary's background information in passive attacks is typically modeled as structural knowledge on the network. This is a sort of *global* view that provides adversaries with the ability to partition the set of nodes into equivalence classes of structurally equivalent nodes. The strongest of those structural relations is automorphism [30]. Two vertices u and v are automorphically equivalent if there exists an isomorphism from the graph to itself such that u maps to v. Other types of structural relations are based on vertex degrees, connectivity, or local neighborhood. Intuitively, structurally equivalent vertices are indistinguishable with respect to the considered structural property.

However, adversaries controlling attacker nodes in a network are undoubtedly more powerful. In addition to the global view, they have a *local view* determined by the relationship of the attacker nodes with the network. To illustrate this let us consider the graph shown in Figure 1. With respect to the vertex degree property, v_2 and v_3 are indistinguishable. They are easily re-identifiable by either an adversary or a legitimate user owning the vertex v_4 and knowing its distance to v_2 and v_3, though.

A first step towards modeling such local view was given by Hay et al. [6], who defined the concept of *hub fingerprint queries*. A hub is a relevant node in the network with high degree and high centrality, and a hub fingerprint for a target node x is a vector of distances from x to hub vertices. Although not explicitly mentioned in [6], the largest hub fingerprint for a target node x is indeed the *metric representation* of x with respect to the hub vertices. We formally define this concept as follows.

Definition 1 (Metric representation). *Let $G = (V, E)$ be a simple connected graph and $d_G(u, v)$ be the length of the shortest path between the vertices u and v in G. For an ordered set $S = \{u_1, \cdots, u_t\}$ of vertices in V and a vertex v, we call $r(v|S) = (d_G(v, u_1), \cdots, d_G(v, u_t))$ the* metric representation *of v with respect to S.*

Similarly to Hay et al. work [6], we define the adversary's background knowledge about a target node u as the metric representation of u with respect to S. In this article, however, we assume S to be any subset of attacker nodes rather than hub vertices only.

It is worth mentioning that the concept of metric representation is also the basis of two weel-known concepts: resolving sets and metric dimension (cf. Definition 2). Both have been already motivated by problems related to unique recognition of an intruder position in a network [17], where resolving sets were called *locating sets*. The name “resolving set” is due to Harary and Melter [5], who introduced the concept in 1976.

Definition 2 (Resolving set and metric dimension)**.** *Let $G = (V, E)$ be a simple connected graph. A set $S \subset V(G)$ is said to be a* resolving set *for G if any pair of vertices of G have different metric representations with respect to S. A resolving set of the smallest possible cardinality is called a* metric basis*, and its cardinality the* metric dimension *of G.*

3.2. (k, ℓ)-anonymity

(k, ℓ)-anonymity is a privacy measure that evolves from the adversary's background knowledge defined previously. It is based on the concept of *k-antiresolving set* defined as follows.

Definition 3 (k-antiresolving set)**.** *Let $G = (V, E)$ be a simple connected graph and let $S = \{u_1, \cdots, u_t\}$ be a subset of vertices of G. The set S is called a k-*antiresolving set *if k is the greatest positive integer such that for every vertex $v \in V - S$ there exist at least $k - 1$ different vertices $v_1, \cdots, v_{k-1} \in V - S$ with $r(v|S) = r(v_1|S) = \cdots = r(v_{k-1}|S)$,* i.e.*, v and $v_1, \cdots, v_{k-1}$ have the same metric representation with respect to S.*

The following concepts derive from Definition 3, whose study is one of the goals of this article.

Definition 4 (k-metric antidimension and k-antiresolving basis)**.** *The* k-metric antidimension *of a simple connected graph $G = (V, E)$ is the minimum cardinality amongst the k-antiresolving sets in G and is denoted by* $\mathrm{adim}_k(G)$*. A k-antiresolving set of cardinality* $\mathrm{adim}_k(G)$ *is called a* k-antiresolving basis *for G.*

It is easy to prove that if the set of attacker nodes S is a k-antiresolving set, the adversary cannot uniquely re-identify other nodes in the network with probability higher than $1/k$. However, given that S is unknown, the privacy measure should quantify over all possible subsets S as follows.

Definition 5 ((k, ℓ)-anonymity)**.** *A graph G meets* (k, ℓ)-anonymity *with respect to active attacks if k is the smallest positive integer such that the k-metric antidimension of G is lower or equal than ℓ.*

In Definition 5 the parameter k is used as a privacy threshold, whilst ℓ is an upper bound on the expected number of attacker nodes in the network. Because attacker nodes are difficult to enrol in a network without been detected [27], ℓ can be estimated through statistical analysis. A fair assumption, for example, is that the number of attacker nodes is significantly lower than the total number of nodes in the network. To further explain the role of k and ℓ in Definition 5 we provide the following example result.

Theorem 1. *For every $n > 0$ and $0 < \ell < n$, the graph K_n meets $(n - \ell, \ell) - anonymity$.*

Proof. Since all the vertices in a complete graph K_n are connected, every subset S of vertices of K_n is an $(n-|S|)$-antiresolving set. Therefore, the k-metric antidimension of K_n is $n-k$.

According to Definition 5, the k-metric antidimension should be lower or equal than ℓ, which implies that $k \geq n-\ell$. Moreover, k should be the smallest positive integer satisfying the previous condition. Therefore, K_n holds $(n-\ell,\ell)$-*anonymity.* □

Corollary 2. *A social graph K_n guarantees that a user cannot be re-identified with probability higher than $\frac{1}{n-\ell}$ by an adversary controlling ℓ attacker nodes.*

These simple and intuitive result obtained in Theorem 1 and Corollary 2 shows the role of the privacy measure (k,ℓ)-anonymity in privacy-preserving publication of social graphs. Before releasing a social graph G, the goal is to find k such that G satisfies (k,ℓ)-anonymity. To do so, theoretical results and efficient algorithms on the k-metric antidimension of a graph need to be investigated.

4. Computing the k-metric antidimension

Computing the k-metric antidimension of a graph seems to be a challenging problem, whose hardness ought to be investigated. It should be remarked, however, that it becomes the well-known metric dimension problem for $k=1$, which has been proven to be NP-Complete [7]. The extension of the metric dimension (the k-metric dimension) problem turned out to be NP-complete [23] as well. Thus, we address the k-metric antidimension problem by proposing a true-biased algorithm whose success rate and computational cost can be balanced.

4.1. A true-biased algorithm

A true-biased algorithm is always correct when it returns `true`, it might fail with some small probability when its output is `false`. True-biased algorithms normally are Monte Carlo algorithms with deterministic running time and randomized behavior. The algorithm we introduce in this section resembles to a Monte Carlo algorithm in the sense that it is deterministic and has the true-biased property. The proposed algorithm is not randomized, though.

The mathematical foundation of our algorithm requires the introduction of notation as follows. For a given subset of vertices $X \subseteq V(G)$, we denote $\sim_X : V(G)\times V(G)$ to the symmetric, reflexive and transitive relation satisfying that $u \sim_X v \implies r(u|X) = r(v|X)$. The set of equivalence classes created by $\sim_X$ over the subset of vertices $V(G) - X$ is denoted as C_X. We deliberately abuse notation and use $\sim_v$ and C_v instead of $\sim_{\{v\}}$ and $C_{\{v\}}$ for every vertex $v \in V(G)$.

Proposition 3. *Let $S \subseteq V(G)$ and $S' \subseteq S$:*

- $u \sim_S v \implies u \sim_{S'} v$
- $\forall X \in C_S$ *there exists* $X' \in C_{S'}$ *such that* $X \subseteq X'$
- $\forall X \in C_S$ *and* $\forall X' \in C_{S'}$, $X \cap X' \neq \emptyset \implies X \subseteq X'$

Lemma 4. *Let S be a k-antiresolving set and let $S' \subseteq S$. Let $Y = \{X \in C_{S'} : |X| < k\}$, then $S' \cup (\bigcup_{y\in Y} y) \subseteq S$.*

Proof. By Proposition 3, for every $X \in C_S$ there exists $X' \in C_{S'}$ such that $X \subseteq X'$, which implies that $|X'| \geq |X| \geq k$ due to the definition of k-antiresolving set. Consequently, $|X'| < k$ implies that there does not exist $X \in C_S$ such that $X \subseteq X'$, meaning that there does not exist $X \in C_S$ such that $X \cap X' \neq \emptyset$ according to Proposition 3. Therefore, $X' \cap (V(G) - S) = \emptyset$ and thus $X' \subseteq S$. □

In the spirit of Lemma 4, let $f : V(G) \to V(G)$ be the function defined recursively as follows:

$$f(S) = \begin{cases} f(S \cup (\bigcup_{y \in Y} y)), & \text{if } Y = \{X \in C_S : |X| < k\} \text{ is not empty,} \\ S, & \text{otherwise.} \end{cases} \tag{1}$$

According to Lemma 4, if S is a subset of a k-antiresolving set, so is $f(S)$. We therefore give some useful properties of the function f in Theorem 5 below.

Theorem 5. *The function defined in Equation 1 satisfies the following properties.*

1. $f(f(S)) = f(S)$
2. $S' \subseteq S \implies f(S') \subseteq f(S)$
3. $\forall S' \subset S, f(S) = f(f(S - S') \bigcup f(S'))$
4. $S' \subseteq f(S) \implies f(S') \subseteq f(S)$

Proof. The first property comes straightforwardly from Equation 1. In order to prove the second property, let $S' \subseteq S$ and $u \in f(S')$. If $u \in S$, then $u \in f(S)$ by definition. Let us thus assume that $u \notin S$. Given that $u \in f(S')$, there exist $X' \in C_{S'}$ such that $|X'| < k$ and $u \in X'$. Let $X \in C_S$ such that $u \in X$. Note that, such an X exists because $u \notin S$. According to Proposition 3, since $X \cap X' \neq \emptyset$ and $S' \subseteq S$, then $X \subseteq X'$, which means that $|X| < k$ and that $X \subseteq f(S)$, which proves the second property.

The third property can be proven by using the first property. Given that $S - S' \subset S$ and $S' \subset S$, then $f(S - S') \subseteq f(S)$ and $f(S') \subseteq f(S)$, hence, $f(f(S - S') \bigcup f(S')) \subseteq f(S)$. Similarly, $S - S' \subseteq f(S - S')$ and $S' \subseteq f(S')$ by definition, which implies that $S \subseteq f(S - S') \bigcup f(S')$. Again, applying the first property we obtain that $f(S) \subseteq f(f(S - S') \bigcup f(S'))$. The two results lead to $f(S) = f(f(S - S') \bigcup f(S'))$.

Finally, the last property is proven as follows. If $S' \subseteq f(S)$, then $f(S') \subseteq f(f(S))$ by applying the second property. The proof is concluded by simply considering the first property. □

The function $f(.)$ is the basis of Algorithm 1, which aims to find a k-antiresolving set in a graph. Algorithm 1 is an optimized version supported by Theorem 5 of the following algorithm. Let us consider all subsets S of $V(G)$ with cardinality lower than or equal to m. If $f(S)$ is a k-antiresolving set, then a positive output is provided. If not, a proof that a k-antiresolving set does not exist is found when $f(S) = V(G)$ for every $S \subseteq V(G)$ such that $|S| = m$. Note that, this impossibility result comes from the monotonicity of the function f, *i.e.*, $S' \subseteq S \implies f(S') \subseteq f(S)$. Any other case leads to the *unknown* state where neither a proof nor a disproof of the existence of a k-antiresolving set can be found.

Algorithm 1 Given a positive integer k, this algorithms outputs: i) `true` if it finds a k-antiresolving set, ii) `false` if such a set does not exist, iii) `unknown` when neither a k-antiresolving set nor a proof that such a set does no exist was found.

Require: A graph G, an integer value m to control the exponential explosion, and the integer value k.

Let $V(G) = \{v_1, \cdots, v_N\}$
Let $C_1 = \{f(\{v_1\}), \cdots, f(\{v_N\})\}$
if $\exists S \in C_1$ that is a k-antiresolving set **then return** true
for $h = 2$ to m **do**
    Let C_h be an empty set
    for $i = 1$ to $|C_{h-1}|$ **do**
        Let S_i be the ith element of C_{h-1}
        for $j = i + 1$ to $|C_{h-1}|$ **do**
            Let S_j be the jth element of C_{h-1}
            if $S_i \nsubseteq S_j$ and $S_j \nsubseteq S_i$ **then**
                $S = f(S_i \bigcup S_j)$
                if S is a k-antiresolving set **then return** `true`
                Add S to C_h
if $\forall S \in C_m, S = V(G)$ **then return** `false`
else return `unknown`

Algorithm 1 can be considered a true-biased algorithm if the unknown state is regarded as a negative result. Its computational complexity is clearly exponential in terms of m. More precisely, for every $i \in \{2, \cdots, m\}$ we obtain that $|C_{i-1}| \leq |C_i| \leq |C_{i-1}|(|C_{i-1}| - 1)/2$, because C_i is formed by joining every pair of elements of C_{i-1}. This means that the computational complexity of Algorithm 1 is determined by the size of C_m. Given that, in the worst case, the cardinality of C_m quadratically increases with respect to C_{m-1}, we obtain that the worst-case computational complexity of this algorithm is $\mathcal{O}(N^{2^{m-1}})$.

Although $\mathcal{O}(N^{2^{m-1}})$ is double exponential in terms of m, when $m << N$ it becomes significantly lower than the computational complexity of a brute force algorithm that considers the 2^N subsets of $V(G)$. For example, for $m = 1$, $m = 2$, and $m = 3$, the computational complexity becomes $\mathcal{O}(N)$, $\mathcal{O}(N^2)$, and $\mathcal{O}(N^4)$, respectively. Moreover, given that the search space monotonically increases with m, the accuracy of the algorithm also increases with m. In this sense, m provides a trade-off between false negatives and computational cost.

It is worth remarking that a theoretical lower bound, although not considered in the analysis, of the computational complexity of Algorithm 1 is $\mathcal{O}(N^3)$, which is the computational complexity of the *classic* Floyd-Warshall algorithm required to compute the metric representation of all vertices. This prevents our method to be used on large graphs even when $m = 1$. In this case, more efficient implementations of both Algorithm 1 and the Floyd-Warshall algorithm ought to be considered, *e.g.,* [19].

Algorithm 1 can be adapted to find a k-antiresolving basis rather than a k-antiresolving set. To that aim, we rely on Proposition 6 below. Proposition 6 gives a sufficient condition for the presence of a k-antiresolving basis. This implies just a small modification to Algorithm 1. In particular, the conditional statements in lines 3 and 12 should take into account that such sufficient condition is satisfied. The full pseudo-code considering this modification is presented in Algorithm 2.

Proposition 6. *Let S be the subset of smaller cardinality in $V(G)$ such that $f(S)$ is a k-antiresolving set. Then, $f(S)$ is a k-antiresolving basis if $\forall S' \subseteq V(G)$ such that $|S'| = |S|$ it follows that $|f(S)| \leq |f(S')|$.*

Algorithm 2 Given a positive integer k, this algorithms outputs: i) `true` if a k-antiresolving basis is found, ii) `false` if a k-antiresolving basis does not exist, iii) `unknown` when neither a k-antiresolving basis nor a proof that it does not exist was found.

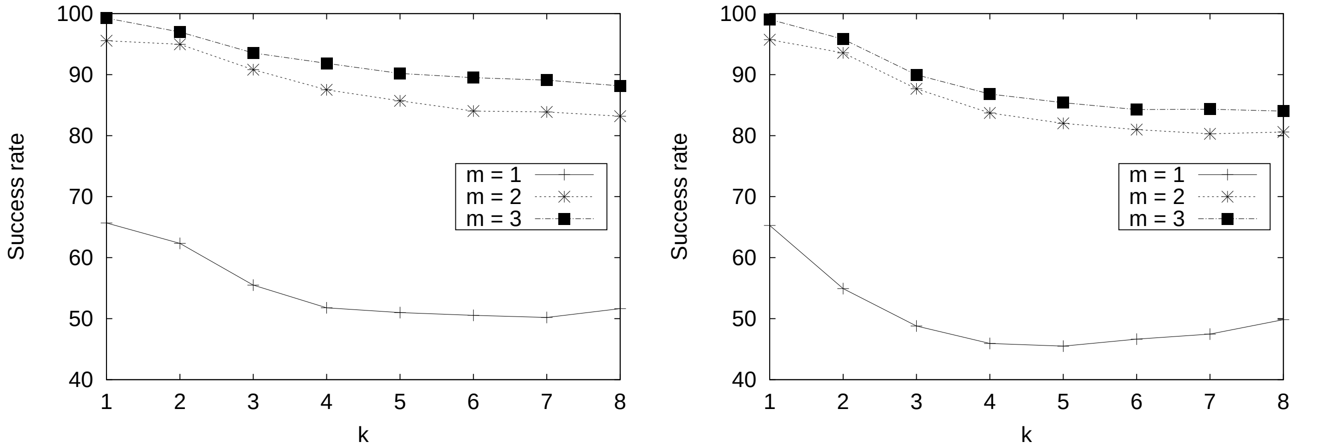
Require: A graph G, an integer value m to control the exponential explosion, and the integer value k.

Let $V(G) = \{v_1, \cdots, v_N\}$
Let $C_1 = \{f(\{v_1\}), \cdots, f(\{v_N\})\}$
Let $minSet = \min(|f(\{v_1\})|, \cdots, |f(\{v_N\})|)$
if $\exists S \in C_1$ such that S is a k-antiresolving set and $|S| \leq minSet$ **then return** `true`
for $h = 2$ to m **do**
    Let C_h be an empty set
    for $i = 1$ to $|C_{h-1}|$ **do**
        Let S_i be the ith element of C_{h-1}
        for $j = i + 1$ to $|C_{h-1}|$ **do**
            Let S_j be the jth element of C_{h-1}
            if $S_i \not\subseteq S_j$ and $S_j \not\subseteq S_i$ **then**
                Add $f(S_i \bigcup S_j)$ to C_h
    Let $minSet$ be the minimum cardinality of a set in C_h
    if $\exists S \in C_1 \cup \cdots \cup C_h$ such that S is a k-antiresolving set and $|S| \leq minSet$ **then return** `true`
if $\forall S \in C_m, S = V(G)$ **then return** `false`
else return `unknown`

4.2. Empirical evaluation on synthetic graphs

In order to show the feasibility of both Algorithm 1 and Algorithm 2, we ran experiments considering $m \in \{1, 2, 3\}$ and random graphs as input. The aim of the experiments is to provide statistically sound data on the ratio between positive, negative, and unknown results of the proposed algorithms. Further below in this section we also show results on real-life social graphs.

A random graph is created by choosing integer values uniformly distributed in the interval $[k + 2, 100]$ as the number of vertices N; where k is a privacy threshold. The number of edges also distributes uniformly in the interval $[0, N \times (N - 1)/2]$, and the edges are added randomly to the graph. For each pair $(m, k) \in \{1, 2, 3\} \times \{1, 2, 3, 4, 5, 6, 7, 8\}$, we created 10 000 random



(a) k-antiresolving set

(b) k-antiresolving basis

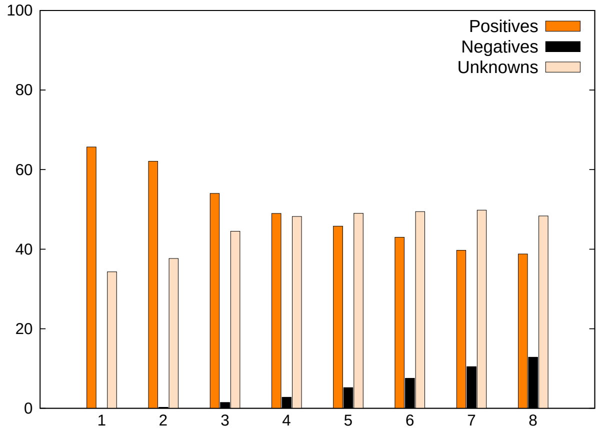
Figure 2: Figure 2(a) and Figure 2(b) depicts the success rates of Algorithm 1 and Algorithm 2, respectively. The considered values of m are $\{1, 2, 3\}$, while k varies from 1 to 8.

graphs and executed Algorithm 1 and Algorithm 2 in order to look for a k-antiresolving set and a k-antiresolving basis, respectively.

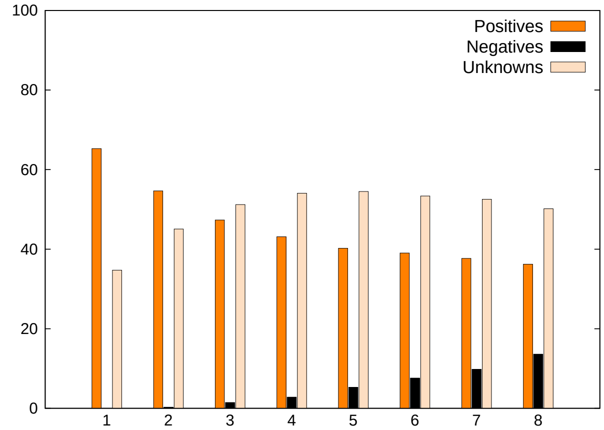
The success rate of both algorithms considering $m \in \{1, 2, 3\}$ is shown in Figure 2. We define a success as either a positive or a negative result, *i.e.,* whenever a k-antiresolving set (basis) or a proof that it does not exist is found. According to Figure 2, both algorithms perform poorly for $m = 1$. However, when $m = 2$ they already achieve a success rate above 80%, which is further improved by the more computationally demanding versions of Algorithm 1 and Algorithm 2 that consider $m = 3$.

The difference between Figure 2(a) and Figure 2(b) suggests, as expected, that finding a k-antiresolving basis is harder than finding a k-antiresolving set. Notwithstanding, Algorithm 2 performs above 80% when $m = 2$ and $m = 3$. It is also worth remarking that, even though Figure 2 hints that the success rate of both algorithms monotonically decreases with k, our algorithms have 100% of success rate if k is equal to the order of the graph. This is because $\forall u \in V(G)\, \forall X \in C_u(|X| < |V(G)|)$ and, thus, all the nodes in the graph should be contained in a k-antiresolving set according to Lemma 4.

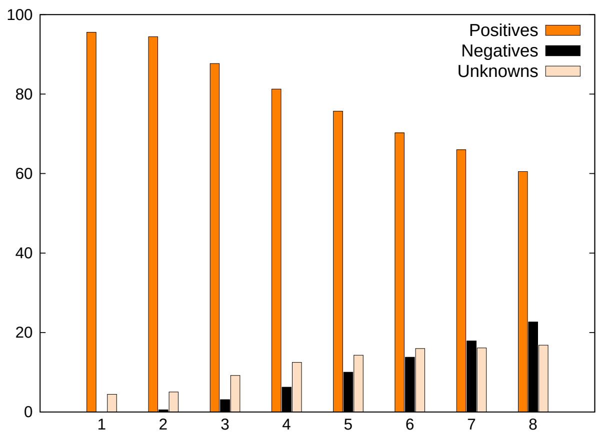
In the previous section we provided a theoretical impossibility result whereby a graph can be proven to not contain a k-antiresolving set. In Figure 3, we show that such impossibility result can be achieved by random graphs; with small probability though. It seems also that the percentage of negative results monotonically increases with k. Indeed, it is easy to prove that this percentage reaches its minimum (0%) and maximum (100%) when k takes its minimum ($k = 1$) and maximum ($k = N$) value, respectively. However, proving the monotonicity of the percentage of negative results with respect to k looks challenging and cumbersome. Figure 3 also shows that the increase of the success rate of both algorithms when m grows is due to an increase on both the number of positive and negative results.



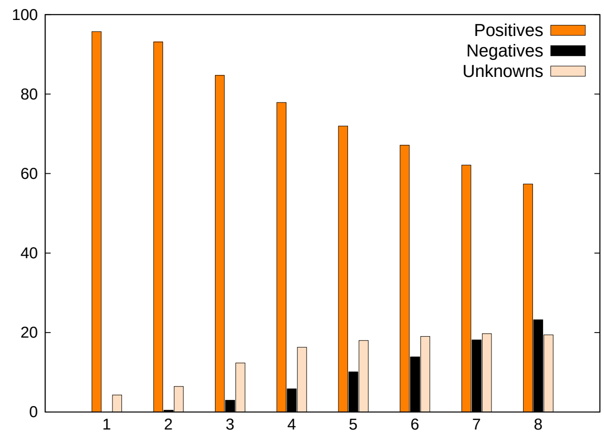
(a) k-antiresolving set ($m = 1$)



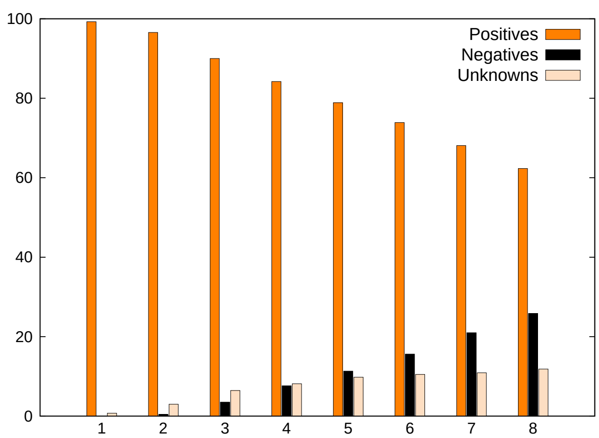
(b) k-antiresolving basis ($m = 1$)



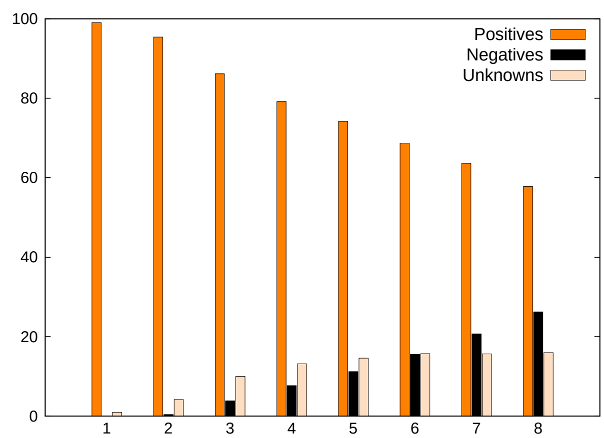
(c) k-antiresolving set ($m = 2$)



(d) k-antiresolving basis ($m = 2$)



(e) k-antiresolving set ($m = 3$)



(f) k-antiresolving basis ($m = 3$)

Figure 3: Six charts showing the ratio of `true`, `false`, and `unknown` results provided by Algorithm 1 and Algorithm 2 on different values for $m \in \{1, 2, 3\}$. Charts at the left are devoted to the algorithm aimed at finding a k-antiresolving set, charts at the right consider the algorithm that looks for a k-antiresolving basis.

4.3. Empirical evaluation on real-life social graphs

This section ends with the evaluation of two real-life social graphs with respect to the proposed privacy measure. The first graph, named *Facebook graph* in what follows, consists of 10 ego-networks from Facebook [10]. It contains 4039 users, 88234 edges, and 193 circles. The second graph describes an online community of students at the University of California [14]. In total, 1899 students were registered in the network and 13838 links were created. We refer to this graph as *Panzarasa graph*.

Both graphs have been analyzed in order to determine the values of k and ℓ such that they (k,ℓ)-anonymity. Taking into account the previously presented empirical results on synthetic data, we used for the analyses Algorithm 2 with $m=2$ as a good trade-off between performance and success rate. The results are as follows.

The Panzarasa graph contains a 1-antiresolving basis of size 1. This means that this graph does not satisfy (k,ℓ)-anonymity for $k>1$ unless $\ell<1$, which is a meaningless scenario. Similarly, the k-metric antidimension of the Facebook graph is 1 for $k=1$. Hence, it satisfies $(1,1)$-anonymity only; the lowest privacy guarantee with respect to our privacy measure.

Our results show that neither the Panzarasa nor the Facebook graph provide privacy guarantees against active attack. This is not surprising since these graphs have not been anonymized to prevent any type of structural attack. Future work thus should be oriented to anonymization methods that consider (k,ℓ)-anonymity as a privacy goal.

5. Mathematical properties on the k-metric antidimension of graphs

In the next two sections we provide some primary theoretical results on the k-metric antidimension problem. We focus on giving mathematical properties that, supported by the results in Section 4, can determine or bound the k-metric antidimension for some families of graphs. In particular in this section, we study the k-metric antidimension of cycles, paths, complete bipartite graphs, and other graph families satisfying some specific conditions. To do so, we first observe some basic properties of k-antiresolving sets, which will be further used.

Observation 1.

(i) *A 1-antiresolving set is also a resolving set.*

(ii) *There does not exist $k>1$ such that all the vertices of a graph form a k-antiresolving set.*

(iii) *There does not exist any n-antiresolving set in a graph of order n.*

(iv) *Not for every graph G of order n and every $k<n$, there exist a k-antiresolving set in G. For instance, if G is a path graph, for $k\geq 3$ there does not exist a k-antiresolving set in G.*

In order to continue with our study we need to introduce some terminology and notation. For a graph G and a vertex $v\in V(G)$, the set $N_G(v)=\{u\in V:\ uv\in E(G)\}$ is the *open neighborhood* of v and the set $N_G[v]=N_G(v)\cup\{v\}$ is the *closed neighborhood* of v. Two vertices x, y are called *(false) true twins* if $(N_G(x)=N_G(y))$ $N_G[x]=N_G[y]$. In this sense, a vertex x is a *twin* if there exists $y\neq x$ such that they are either true or false twins. The diameter of G is defined as $D(G)=\max_{u,v\in V}\{d_G(u,v)\}$.

As mentioned before, not for every integer k it is possible to find a k-antiresolving set in a graph G. Thus, it is desirable to analyze first the interval of suitable values for k satisfying that G contains at least one k-antiresolving set. According to Definition 3 we present the following concept, which is relevant in the study of the k-metric antidimension of graphs.

Definition 6 (k-metric antidimensional graph)**.** *A simple connected graph $G = (V, E)$ is k-*metric antidimensional*, if k is the largest integer such that G contains a k-antiresolving set.*

5.1. k-metric antidimensional graphs

In order to study the k-metric antidimension of graphs, we first focus into obtaining the values of k for which a given graph is k-metric antidimensional. First notice that any graph G is always k-metric antidimensional for some $k \geq 1$, and a natural upper bound for k which makes that G would be k-metric antidimensional is clearly the maximum degree of the graph, since the number of vertices at distance one from any vertex is at most the maximum degree of the graph.

Observation 2. *If G is a connected k-metric antidimensional graph of maximum degree Δ, then $1 \leq k \leq \Delta$.*

Since the maximum degree of a graph is at most the order of the graph minus one, a particular case of the above result is the next one.

Remark 7. *If G is any connected k-metric antidimensional graph of order n, then $1 \leq k \leq n-1$. Moreover, G is $(n-1)$-metric antidimensional if and only if G has maximum degree $n-1$.*

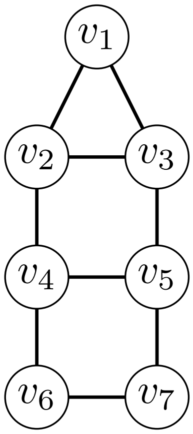
Proof. The upper bound is a particular case of Remark 2. Now, it is straightforward to observe that if v is a vertex of G of degree $n-1$, then for every vertex $u, w \neq v$ it follows that $r(u|\{v\}) = r(w|\{v\}) = 1$, *i.e.*, every vertex different from v has the same metric representation with respect to $\{v\}$. Thus, $\{v\}$ is a $(n-1)$-antiresolving set, since there exists no n-antiresolving sets in G. Thus, G is $(n-1)$-metric antidimensional.

On the contrary, we assume that G is $(n-1)$-metric antidimensional. Hence, if S is a $(n-1)$-antiresolving set, then $|S| = 1$. Thus, the only possibility is that S is formed by a single vertex and that every vertex is adjacent to it. □

To continue with our study we need some extra notation. The *eccentricity* $\epsilon(v)$ of a vertex v in a connected graph G is the maximum length of a shortest path between v and any other vertex u of G. Notice that the maximum of the eccentricities of any vertex of G is the diameter of G and the minimum of the eccentricities is the radius of G. For more information on vertex eccentricity in graphs, see for instance [21]. Figure 4 shows a graph G and a table with the eccentricities of all its vertices. Given a vertex u of a graph G, we consider the following local parameter. For every $i \in \{1, ..., \epsilon(u)\}$, let $d_i(u) = \{v \in V(G) : d(v, u) = i\}$. Now, for every $u \in V(G)$, let

$$\phi(u) = \min_{1 \leq i \leq \epsilon(u)} \{|d_i(u)|\}.$$

and for any graph G, let $\phi(G) = \max_{v \in V(G)}\{\phi(v)\}$. The table and the graph of Figure 4 clarify the notation above.



v_i	$\epsilon(v_i)$	$d_1(v_i)$	$d_2(v_i)$	$d_3(v_i)$	$\phi(v_i)$
v_1	3	$\{2,3\}$	$\{4,5\}$	$\{6,7\}$	2
v_2	3	$\{1,3,4\}$	$\{5,6\}$	$\{7\}$	1
v_3	3	$\{1,2,5\}$	$\{4,7\}$	$\{6\}$	1
v_4	2	$\{2,5,6\}$	$\{1,3,7\}$	$\emptyset$	3
v_5	2	$\{2,4,7\}$	$\{1,3,6\}$	$\emptyset$	3
v_6	3	$\{4,7\}$	$\{2,5\}$	$\{1,3\}$	2
v_7	3	$\{5,6\}$	$\{3,4\}$	$\{1,2\}$	2

Figure 4: Eccentricities of vertices of a graph G and a table which shows that $\phi(G) = 3$.

Theorem 8. *Any connected graph G is k-metric antidimensional for some $k \geq \phi(G)$.*

Proof. Assume x is a vertex of degree at least two in G such that $\phi(G) = \phi(x)$. Thus, for any vertex $y \neq x$, there exist at least $\phi(x) - 1$ vertices $v_1, v_2, ...v_{\phi(G)-1}$ in $V(G) - \{x, y\}$ such that $d(y, x) = d(v_1, x) = ... = d(v_{\phi(T)-1}, x)$. Moreover, since $\phi(G) = \phi(x)$, there exists at least one vertex y' such that there are exactly $\phi(G) - 1$ different vertices satisfying the above mentioned. So, $\{x\}$ is a $\phi(x)$-antiresolving set and G is k-metric antidimensional for some $k \geq \phi(G)$. □

If a graph G is 1-metric antidimensional, then its 1-antiresolving sets are standard resolving sets as defined in [5, 17] and this has been very frequently studied in the last years. According to that, in this work we are mainly interested in those graphs being k-metric antidimensional for some $k \geq 2$. An example of a graph being 1-metric antidimensional is for instance the path graph of even order.

5.2. Graphs that are k-metric antidimensional for some $k \geq 2$

To begin with the description of some families of graphs being k-metric antidimensional for some $k \geq 2$ we define the *radius* and the *center* of a graph as follows. The *radius* $r(G)$ of G is the minimum eccentricity of any vertex in G. The *center* of G is the set S of vertices of G having eccentricity equal to the radius of G.

Remark 9. *If the center of a graph G is only one vertex, then G is k-metric antidimensional for some $k \geq 2$.*

Proof. Let v be the center of G. Hence, there exist two diametral vertices u, w such that $d_G(v, u) = d_G(v, w) = \epsilon(v) = r(G)$. Since v has eccentricity $r(G)$, there is no vertex $z \neq u, w$ in G such that $d_G(v, z) > d_G(v, u) = d_G(v, w)$. Thus, it follows that for any vertex $x \neq v$ there exists at least a vertex y belonging to the $u - w$ path such that $d_G(x, v) = d_G(y, v)$. Therefore, $\{v\}$ is a k-antiresolving set in G for some $k \geq 2$. □

If a path graph has odd order, then its center is formed by only one vertex. Also, for every vertex of any path, there exists at most other different vertex having equal distance to a third vertex of the path. Thus, it is clear the following consequence of the Remark above.

Corollary 10. *If a path P_n has odd order, then it is 2-metric antidimensional.*

Another example of 2-metric antidimensional are the cycle graphs as we next see.

Remark 11. *Any cycle graph C_n is 2-metric antidimensional.*

Proof. We assume first that n is odd. Let v be any vertex of C_n. Hence, for any vertex $x \neq v$ of C_n, there exists only one $y \neq x, v$ such that $d_{C_n}(x, v) = d_{C_n}(y, v)$. Thus, $\{v\}$ is a 2-antiresolving set in C_n. Assume now that n is even and let $\{u, w\}$ be any two diametral vertices of C_n. We observe that for any vertex $x \neq u, w$ of C_n, there exists only one $y \neq x, u, w$ such that $d_{C_n}(x, u) = d_{C_n}(y, u)$ and $d_{C_n}(x, w) = d_{C_n}(y, w)$. Thus, $\{u, w\}$ is a 2-antiresolving set in C_n.

On the other hand, there does not exists $k > 2$ such that C_n contains a k-metric antiresolving set, since for any vertex of C_n, there exists at most another different vertex having equal distance to a third vertex of the path. Therefore, C_n is 2-metric antidimensional. □

If the vertices of a set S are pairwise twins in a graph G, then it is clear that they have the same distance to every other vertex $x \notin S$. So, $V(G) - S$ is a $|S|$-antiresolving set for G. Hence, the following result.

Observation 3. *If the vertices of a set S are pairwise twins in a graph G, then G is k-metric antidimensional for some $k \geq |S|$.*

Complete bipartite graph[3] are special kind of graphs, since they have a bipartition of the vertex set in which all the vertices belonging to one of the sets of the bipartition are pairwise twin vertices. Let $K_{r,t}$ be a complete bipartite graph. Next we analyze the suitable values k making a complete bipartite graph k-metric antidimensional.

Remark 12. *Any complete bipartite graph $K_{r,t}$ with $r \geq t$ is r-metric antidimensional.*

Proof. Let U and V be the two disjoint sets of $K_{r,t}$ with $|U| = r$ and $|V| = t$. Notice that U (respectively V) is a set of pairwise twin vertices. Thus, by Observation 3 we have that $K_{r,t}$ is k-metric antidimensional for some $k \geq |U| = r$. Suppose that $k \geq r + 1$ and let S be a k-antiresolving set for $K_{r,t}$. Since every vertex of $K_{r,t}$ has distance either one or two to any other vertex of $K_{r,t}$ it is not possible to find $k - 1$ vertices having the same distance to every vertex of S, a contradiction. So, $k = r$ and the proof is complete. □

5.3. The k-metric antidimension of graphs

In this subsection we compute the k'-metric antidimension of some graphs which were already described to be k-metric antidimensional for some value $k \geq k'$. It is clear that the first natural bound which follows for the k-metric antidimension of a graph of order n is $\mathrm{adim}_k(G) \leq n - k$. Such a bound is tight. It is achieved, for instance, for the complete bipartite graphs $K_{r,t}$ as we can see at next by taking the case $t < k \leq r$.

Proposition 13. *Let r, t be two positive integers with $r \geq t$.*

[3]A graph G is complete bipartite if its vertex set can be divided into two disjoint sets U and V such that every vertex in U is adjacent to every vertex in V and no more.

1. *If $t < k \le r$, then* $\operatorname{adim}_k(K_{r,t}) = r + t - k$.

2. *If $1 < k \le t$, then* $\operatorname{adim}_k(K_{r,t}) = r + t - 2k$.

Proof. From Proposition 12 we know that $K_{r,t}$ is a r-metric antidimensional graph. Let U and V be the two partite sets of $K_{r,t}$ with $|U| = r$ and $|V| = t$. We assume first that $t < k \le r$. Let $A \subseteq U$ with $|A| = k$ and let be $S = (V \cup U) - A$. Notice that if $k = r$, then $A = U$ and so, $S = V$. Since any vertex $v \notin S$ (or equivalently $v \in A$) is adjacent to every vertex of V and it has distance two to every vertex in $U - A$, we have that all the vertices of A have the same metric representation with respect to S. As $|A| = k$, it follows that S is a k-antiresolving set and $\operatorname{adim}_k(K_{r,t}) \le r + t - k$. Now, suppose $\operatorname{adim}_k(K_{r,t}) < r + t - k$ and let S' be a k-antiresolving set for $K_{r,t}$. So, we have either one of the following situations.

- There exist more than k vertices of U not in S'. Hence, for any vertex $u \in U - S'$ there exist at least k vertices not in S' which, together with u, have the same metric representation with respect to S'. So, S' is not a k-antiresolving set, but a k'-antiresolving set for some $k' \ge k + 1$, a contradiction.

- There exists at least one vertex of V not in S'. It is a direct contradiction, since $|V| = t < k$.

Therefore, we obtain that $\operatorname{adim}_k(K_{r,t}) = r + t - k$.

On the other hand, we assume that $1 < k \le r$. Let $X \subseteq U$ with $|X| = k$, let $Y \subseteq V$ with $|Y| = k$ and let $Q = (V - Y) \cup (U - X)$. Hence, for any vertex $v \notin Q$ (or equivalently $v \in X \cup Y$), there exist exactly $k - 1$ vertices, such that all of them, together with v, have the same metric representation with respect to Q. Thus, Q is a k-antiresolving set and $\operatorname{adim}_k(K_{r,t}) \le r + t - 2k$.

Now, suppose that $\operatorname{adim}_k(G) < r + t - 2k$ and let Q' be a k-antiresolving set in $K_{r,t}$. Hence, either there exist more than k vertices of U not in Q' or there exist more than k vertices of V not in Q'. As above, in any of both possibilities we obtain that Q' is not k-antiresolving set, but a k'-antiresolving set for some $k' \ge k + 1$, a contradiction. As a consequence, we obtain that $\operatorname{adim}_k(K_{r,t}) = r + t - 2k$. □

Next we study the k-metric antidimension of some other families of basic graphs. According to Remark 11 we know that the cycles C_n are 2-metric antidimensional and, by Corollary 10, that the paths P_n are 2-metric antidimensional only in the case n is odd. Next we compute its 2-metric antidimension.

Proposition 14. *Let $n \ge 2$ be an integer. Then*

$$\operatorname{adim}_2(P_{2n+1}) = 1 \ \textit{and} \ \operatorname{adim}_2(C_n) = \begin{cases} 1, & \textit{if } n \textit{ is odd,} \\ 2, & \textit{if } n \textit{ is even.} \end{cases}$$

Proof. If v is the center of a path P_{2n+1}, then for any other vertex $u \ne v$ there exists exactly one vertex $w \ne v, u$ such that w, u have the same metric representation with respect to $\{v\}$. Thus $\operatorname{adim}_2(P_{2n+1}) = 1$.

Suppose n is even and let u, w be two diametral vertices in C_n. We observe that for any vertex $x \ne u, w$ there exists exactly one vertex $y \ne x, u, w$ in C_n, such that x, y have the same metric representation with respect to $\{u, w\}$. Thus, $\mathrm{adim}_2(C_{2n}) \leq 2$. To see that $\mathrm{adim}_2(C_{2n}) = 2$ we can observe that any set with only one vertex h is not a 1-antiresolving set, since for the vertex f being diametral with h there does not exist any other vertex f' having the same metric representation with respect to $\{h\}$. On the other hand, if n is odd and a is any vertex of C_n, then we can check that for any vertex $b \neq a$, there exists exactly one vertex $c \neq a, b$, such that b, c have the same metric representation with respect to $\{a\}$. Thus, $\mathrm{adim}_2(C_{2n+1}) = 1$. □

6. The particular case of trees

Let T be a tree and let u be a vertex of T of degree at least two. Let v be a neighbor of u. A v-branch of T at u is the subtree $T_{u,v}$ obtained from the union of all length maximal paths beginning in u, passing throughout v and finishing at a vertex of degree one in T. Given a y-branch $T_{x,y}$ at x, we say that $\xi(T_{x,y})$ is the eccentricity of the vertex x in the y-branch $T_{x,y}$. Two branches T_{x,y_1} and T_{x,y_2} at x are ξ_x-*equivalent* if $\xi(T_{x,y_1}) = \xi(T_{x,y_2})$. For every vertex x of T, let $\xi(x)$ represents the maximum number of pairwise ξ_x-equivalent branches at x and let $l_\xi(x)$ equals the length of any ξ_x-equivalent branch. Now, for any tree T, we define the following parameter:

$$\xi(T) = \max_{x \in V(T):\delta(x) \geq 2} \{\xi(x)\}.$$

An example which helps to clarify the above definitions is given in Figure 5. There we have a tree T satisfying the following. The vertex v_5 has 4 branches: T_{v_5,v_6}, $T_{v_5,v_{10}}$, $T_{v_5,v_{15}}$ and T_{v_5,v_4}. For instance $V(T_{v_5,v_{15}}) = \{v_5, v_{15}, v_{16}, v_{17}, v_{11}, v_{18}, v_{14}, v_{13}\}$. We observe that $\xi(T_{v_5,v_6}) = 2$, $\xi(T_{v_5,v_{10}}) = 1$, $\xi(T_{v_5,v_{15}}) = 3$ and $\xi(T_{v_5,v_4}) = 4$. So, v_5 has no ξ_{v_5}-equivalent branches and $\xi(v_5) = 0$. Similarly, it can be noticed that v_3 and v_{15} are the only vertices of T which have equivalent branches. That is, T_{v_3,v_2} and T_{v_3,v_9} are ξ_{v_3}-equivalent, since $\xi(T_{v_3,v_2}) = \xi(T_{v_3,v_9}) = 2$. Thus $\xi(v_3) = 2$ and $l_\xi(v_3) = 2$. Analogously, $\xi(T_{v_{15},v_{11}}) = \xi(T_{v_{15},v_{10}}) = \xi(T_{v_{15},v_{14}}) = 2$ and $\xi(v_{15}) = 3$, $l_\xi(v_{15}) = 2$. Therefore $\xi(T) = 3$.

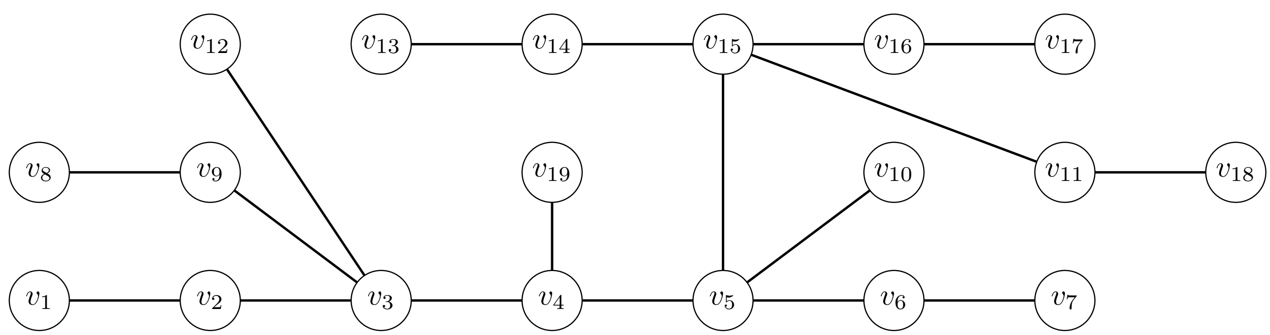


Figure 5: A 3-metric antidimensional tree T.

Now, for the particular case of trees, we next use the definition of $\phi(G)$ already presented in Section 5. As an example, for the tree of Figure 5 we have that, for instance, $\phi(v_3) = 3$, $\phi(v_4) = 3$ and $\phi(v_5) = 2$. Also, some calculations give that $\phi(T) = 3$.

Now, with the definitions above we present the following result.

Theorem 15. *Any tree T is k-metric antidimensional for some $k \geq \max\{\phi(T), \xi(T)\}$.*

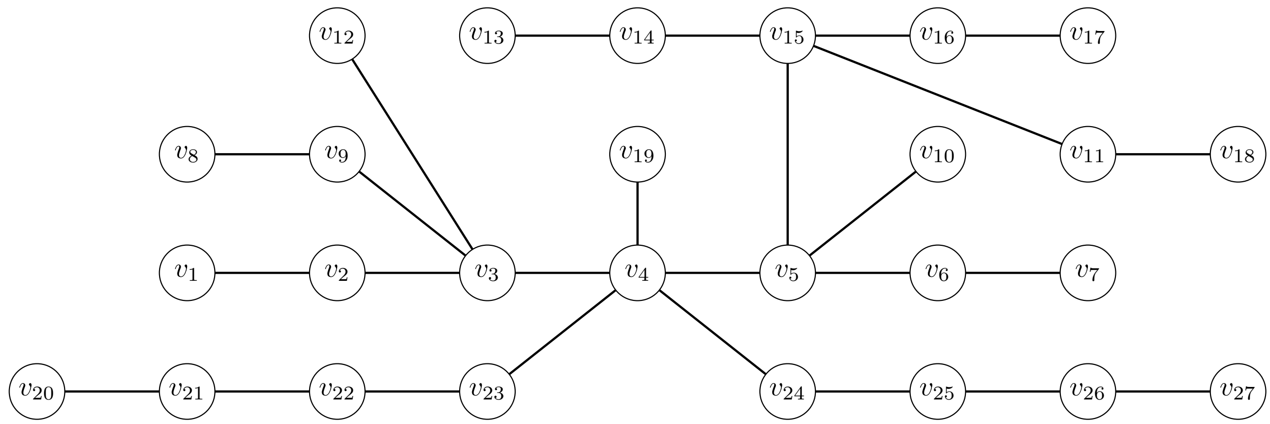


Figure 6: A 5-metric antidimensional tree T.

Proof. From Theorem 8 it follows that $k \geq \phi(T)$. Now, let x be a vertex of degree at least two in T such that $\xi(T) = \xi(x)$. Hence, there exist $\xi(T)$ disjoint paths beginning in x, passing throughout a vertex y_j (neighbor of x), and ending in a vertex w_j of degree one in T with $j \in \{1, ..., \xi(T)\}$. Moreover, every y_j-branch T_{x,y_j} does not contain any other vertex further away from x than w_j.

We consider now the set

$$A = V(T) - \left(\bigcup_{i=1}^{\xi(T)} V(T_{x,y_j}) \right) \bigcup \{x\}.$$

Notice that for any vertex $u \notin A$, there exist at least $\xi(T) - 1$ different vertices $v_1, v_2, ...v_{\xi(T)-1}$ in $V(T) - A$ such that $d(u, z) = d(v_1, z) = ... = d(v_{\xi(T)-1}, z)$ for every $z \in A$. Moreover, since $\xi(T) = \xi(x)$, it follows that there exists a vertex u' such that there are exactly $\xi(T) - 1$ different vertices satisfying the above mentioned. Thus, A is a $\xi(T)$-antiresolving set and, as a consequence, T is k-metric antidimensional for some $k \geq \xi(T)$.

Therefore we obtain that T is k-metric antidimensional for some $k \geq \max\{\phi(T), \xi(T)\}$ and the proof is complete. $\square$

According to Theorem 15, we conclude that the tree T in Figure 5 is k-metric antidimensional for some $k \geq 3$, since $\phi(T) = 3$ and $\xi(T) = 3$; that tree is indeed 3-antidimensional. If we add some extra vertices to this mentioned tree, like in Figure 6, we obtain that $\phi(T) = 5$ (since $\phi(v_4) = 5$), and it remains $\xi(T) = 3$. Thus, this new tree is k-metric antidimensional for some $k \geq 5$, and by Remark 2 we have that $k = 5$.

Notice that $\xi(T) > \phi(T)$ holds for some trees. For instance, if we take a star graph $S_{1,n}$, $n \geq 4$, and we add an extra vertex x connected by an edge with one leaf y of $S_{1,n}$, then we have a tree T such that $\phi(T) = 2$ (for the vertex y, $\phi(y) = 2$) and $\xi(T) = n - 1$.

Moreover, there are graphs in which the bound of Theorem 15 is not achieved. An example of this appears in Figure 7. There we have a tree T such that $\xi(T) = 3$ ($\xi(v_7) = 3$) and $\phi(T) = 3$ ($\phi(v_{12}) = 3$). Nevertheless the set $\{v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{18}, v_{19}\}$ is a 4-antiresolving set.

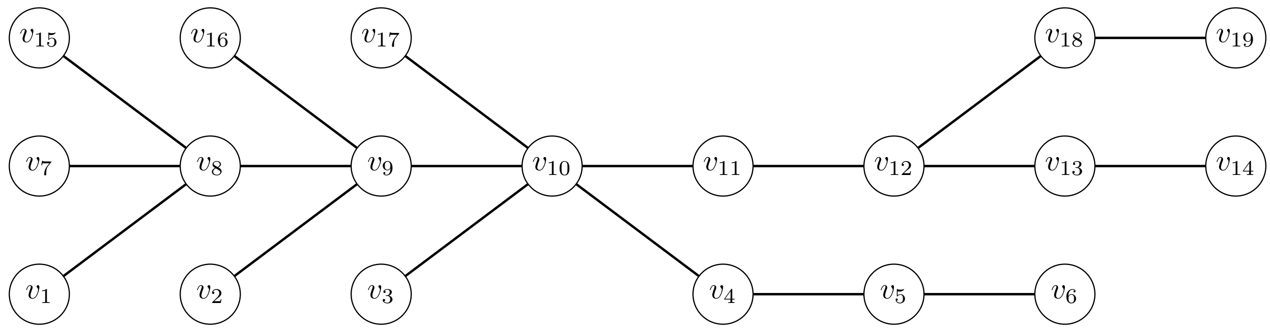


Figure 7: The set $\{v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{18}, v_{19}\}$ is a 4-metric antiresolving set.

6.1. The k-metric antidimension of trees

Once we have a lower bound for the integer k' for which a given tree T is k'-metric antidimensional, we are able to compute its k-metric antidimension for a suitable value $k \leq k'$. We first notice that if T is 1-metric antidimensional, then any 1-metric antiresolving set is an standard resolving set, as defined in [5, 17] and, in such a case, $\mathrm{adim}_1(T) = dim(T)$. Since it is not our goal to study such a case, from now on we consider only those trees being k-metric antidimensional for some $k \geq 2$.

Remark 16. *Let T be a k-metric antidimensional tree and let $x \in V(T)$ such that $\phi(x) = t$. If $t \geq 2$, then* $\mathrm{adim}_t(T) = 1$.

Proof. Since $t \geq 2$, then for every vertex $y \neq x$, there exist at least $t-1$ vertices $v_1, ..., v_{t-1} \in V(T) - \{x, y\}$ such that $d(y, x) = d(v_1, x) = ... = d(v_{\phi(T)-1}, x)$. Moreover, there exists a vertex y' such that there are exactly $t-1$ different vertices satisfying the above mentioned. Thus, $\{x\}$ is a t-metric antiresolving set and, as a consequence, $\mathrm{adim}_t(T) = 1$, since on the other hand, $\mathrm{adim}_t(G) \geq 1$ for any graph G. □

Corollary 17. *For any tree T such that $\phi(T) \geq 2$,* $\mathrm{adim}_{\phi(T)}(T) = 1$.

According to the results above, it remains to study the k-metric dimension of trees for the case in which every vertex v of T satisfies that $\phi(v) \neq k$. To do so, we need to introduce some notations.

We denote by $\Xi_k(T)$, for some $k \in \{2, ..., \xi(T)\}$, the set of vertices $v \in V(T)$ such that $\xi(v) \geq k$. Now, for every $v \in \Xi_k(T)$, let

$$N_{<l_\xi}(v) = \{x \in V(T_{v,u}) \,:\, u \in N(v) \text{ and } \xi(T_{v,u}) < l_\xi(v)\} - \{v\},$$

and for the set of vertices $u \in N(v)$ such that $\xi(T_{v,u}) = l_\xi(v)$, let $N^k_{=l_\xi}(v)$ be the maximum cardinality among all possible sets obtained as the union of k vertex sets of the branches $T_{v,u}$ where $u \in N(v)$ minus the vertex v itself. As an example, we consider the tree of Figure 5. For $k = 2$, there we have that $\Xi_k(T) = \{v_3, v_{15}\}$, $N_{<l_\xi}(v_3) = \{v_{12}\}$, $N_{<l_\xi}(v_{15}) = \emptyset$, $N^2_{=l_\xi}(v_3) = \{v_1, v_2, v_8, v_9\}$ and $N^2_{=l_\xi}(v_{15}) = \{v_{13}, v_{14}, v_{16}, v_{17}\}$ (notice that $N^2_{=l_\xi}(v_{15})$ can be different from this set, but it always has five vertices).

With this definition we are able to present the following result, where we analyze only those graphs being k'-metric antidimensional for $k' = \max\{\phi(T), \xi(T)\}$.

Theorem 18. *Let T be a k'-metric antidimensional of order n with $k' = \max\{\phi(T), \xi(T)\}$. Then for any $k \le k'$,*

$$\mathrm{adim}_k(T) \le n - \left| \bigcup_{v \in \Xi_i(T)} N_{<l_\xi}(v) \right| - \left| \bigcup_{v \in \Xi_i(T)} N^k_{=l_\xi}(v) \right|.$$

Proof. We consider a set $S \subset V(T)$ given by

$$S = V(T) - \bigcup_{v \in \Xi_k(T)} N_{<l_\xi}(v) - \bigcup_{v \in \Xi_k(T)} N^k_{=l_\xi}(v).$$

In this sense, for any vertex $x \notin S$, there exists at least $k-1$ vertices $y_1, y_2, ..., y_{k-1}$ not in S such that $d(x, w) = d(y_1, w) = ... = d(y_{k-1}, w)$ for every $w \in S$. Moreover, if there exists at least one vertex $x' \notin S$ for which there are exactly $k-1$ vertices not in S satisfying the above mentioned, then S is a k-metric antiresolving set and the result follows since the cardinality of S is given by the formula of the theorem. On the contrary, if such a vertex does not exist, then S is a k''-metric antiresolving set for G for some $k'' \ge k$. Since in this case, $\mathrm{adim}_{k''}(G) \ge \mathrm{adim}_k(G)$ we obtain the result. □

Consider now the example of Figure 5. According to the result above, we have that the set $S = \{v_3, v_4, v_5, v_6, v_7, v_{10}, v_{11}, v_{15}, v_{18}, v_{19}\}$ is a 2-metric antiresolving set for such a tree T and $\mathrm{adim}_2(T) \le 10$. Nevertheless, since $\phi(v_5) = 2$, from Remark 16 we have that $\mathrm{adim}_2(T) = 1$. Next we present a family of trees, where the bound of Theorem 18 is achieved.

We consider the family $\mathcal{F}$ of trees T_r satisfying the following conditions.

- The center of T_r is formed by two adjacent vertices, say x, y.
- T_r is "rooted" in x, y.
- T_r is a complete r-ary tree (each vertex of degree greater than one has r children)
- Any two leaves being descendants of the same root (x or y), have the same distance to this root.

An example of a tree T_3 of the family $\mathcal{F}$ is given in Figure 8.

It is straightforward to observe that $\xi(T_r) = r$ and $\phi(T_r) = r + 1$. Thus, T_r is k'-metric antidimensional for some $k \ge r + 1$, and by Remark 2 we have that $k = r + 1$. Now on, we compute the r-metric antidimension of T_r. According to the construction of the family $\mathcal{F}$, we see that the root vertices x, y of a tree $T_r \in \mathcal{F}$ satisfy that $x, y \in \Xi_i(T_r)$. Also, $N_{<l_\xi}(x) = N_{<l_\xi}(y) = \emptyset$ and the sets $N^r_{=l_\xi}(x)$, $N^r_{=l_\xi}(y)$ are formed by the set of all their corresponding descendants (this fact makes unnecessary to consider other vertices of T_r). As a consequence of this, by Theorem 18 we have that $\{x, y\}$ is a r-metric antiresolving set and $\mathrm{adim}_r(T_r) \le 2$. Since, for any non-leaf vertex u of T_r satisfies that $\phi(u) = 4$, we have that any singleton vertex (being not a leaf) is a $(r + 1)$-metric antiresolving set. Thus, $\mathrm{adim}_r(T_r) \ge 2$ and we have that $\mathrm{adim}_r(T_r) = 2$, which makes that the bound of Theorem 18 is tight.

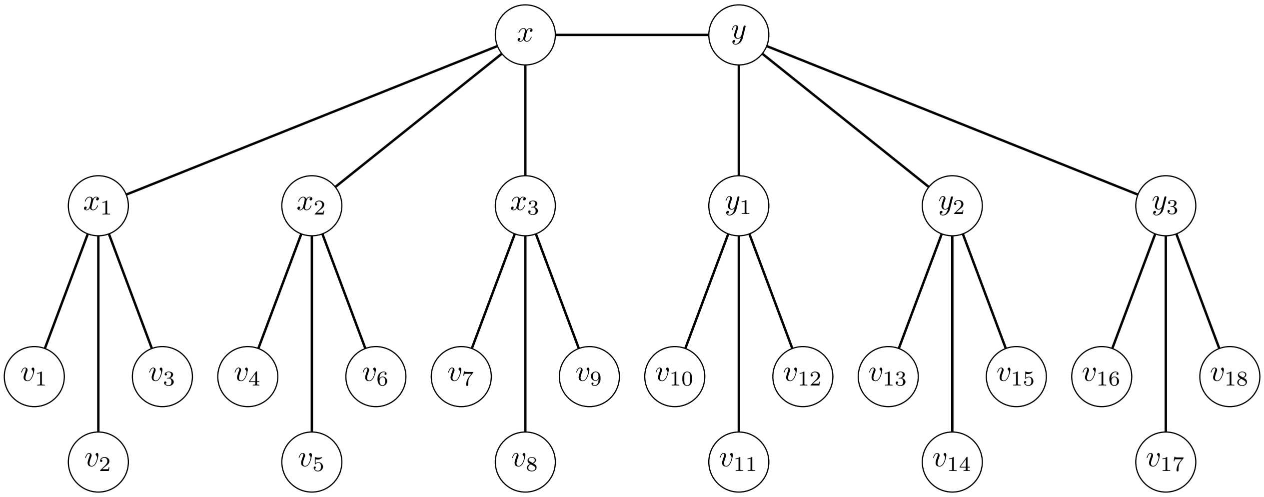


Figure 8: A tree T_3 of the family $\mathcal{F}$. The set $\{x, y\}$ is a 3-metric antiresolving set of minimum cardinality.

7. Discussion and conclusions

In this article we have introduced a new problem in Graph Theory (the k-metric antidimension problem) that resembles to the well-known metric dimension problem. The k-metric antidimension is the basis of our novel privacy measure (k, ℓ)-anonymity. This measures quantifies the level of privacy offered by an outsourced social graph against active attacks. Consequently, privacy-preserving methods for the publication of social networks ought to consider (k, ℓ)-anonymity as one of their privacy goal.

We have proposed a true-biased algorithm aimed at finding both a k-antiresolving set and a k-antiresolving basis in a graph. The algorithm, although computationally demanding, reached a success rate above 80% during the executed experiments when looking for a k-antiresolving basis. We expect future experiments to be conducted over real-life social graphs so that privacy-preserving methods satisfying (k, ℓ)-anonymity can be empirically evaluated in terms of utility and resistance to active attacks.

We have also began the study of mathematical properties of the k-antiresolving sets and the k-metric antidimension of graphs. We have studied some particular graph families like cycles, paths, complete bipartite graphs and trees. For instance, we have obtained that for any path P_n of odd order, $\mathrm{adim}_2(P_n) = 1$ and for any cycle C_n it follows that $\mathrm{adim}_2(C_n) = 1$ if n is odd, and $\mathrm{adim}_2(C_n) = 2$ if n is even. Also, for every complete bipartite graph $K_{r,t}$, $\mathrm{adim}_k(K_{r,t}) = r + t - k$ if $t < k \leq r$, and $\mathrm{adim}_k(K_{r,t}) = r + t - 2k$ if $1 < k \leq t$. For the case of trees we have presented a tight lower bound for its k-metric antidimension in terms of the order of the tree and the order of some subtrees satisfying some specific conditions. We have also described an infinite family of k-ary trees which achieve this bound.

Finally, this article opens new and challenging open problems related to the k-metric antidimension of graphs and the privacy concept (k, ℓ)-anonymity. For instance, it would be interesting to characterize the family of graphs such that they are 1-metric antidimensional, as well as looking for a close relationship between the k-metric antidimension and the k-metric dimension of a graph. In particular, those families of graphs that resemble to social graphs must be considered. The computational complexity of computing the k-metric antidimension should also be addressed. In case the problem is NP-complete, efficient heuristics and privacy-preserving methods need to be developed so as to compute the k-metric antidimension and, ultimately, transform a social graph into a (k, ℓ)-anonymous graph for given values of k and ℓ.

Bibliography

[1] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. *The 16th International Conference on World Wide Web*, pages 181–190, 2007.

[2] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: Automated identity theft attacks on social networks. *The 18th International Conference on World Wide Web*, pages 551–560, 2009.

[3] Francesco Bonchi, Aristides Gionis, and Tamir Tassa. Identity obfuscation in graphs through the information theoretic lens. *Information Sciences*, 275(0):232 - 256, 2014.

[4] Tomás Feder, Shubha U. Nabar, and Evimaria Terzi. Anonymizing graphs. *CoRR*, abs/0810.5578, 2008.

[5] Frank Harary and Robert A. Melter. On the metric dimension of a graph. *Ars Combinatoria*, 2:191–195, 1976.

[6] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. Resisting structural re-identification in anonymized social networks. *The VLDB journal*, 1(1):102–114, August 2008.

[7] Samir Khuller, Balaji Raghavachari, and Azriel Rosenfeld. Landmarks in graphs. *Discrete Appl. Math.*, 70(3):217–229, October 1996.

[8] Rong-Hua Li, Jeffrey Xu Yu, Xin Huang, Hong Cheng, and Zechao Shang. Measuring the impact of MVC attack in large complex networks. *Information Sciences*, 278(0):685 - 702, 2014.

[9] Kun Liu and Evimaria Terzi. Towards identity anonymization on graphs. *The 2008 ACM SIGMOD International Conference on Management of Data*, pages 93–106, 2008.

[10] Julian J. McAuley and Jure Leskovec. Learning to Discover Social Circles in Ego Networks. *The 26th Annual Conference on Neural Information Processing Systems*, pages 548–556, 2009.

[11] Abedelaziz Mohaisen, Tamer AbuHmed, Hun Jeong Kang, Yongdae Kim, and DaeHun Nyang. Mistaking friends for foes: An analysis of a social network-based sybil defense in mobile networks. *The 5th International Conference on Ubiquitous Information Management and Communication*, pages 28:1–28:8, 2011.

[12] A. Narayanan and V. Shmatikov. De-anonymizing social networks. *The 30th IEEE Symposium on Security and Privacy, 2009*, pages 173–187, 2009.

[13] M. Netter, S. Herbst, and G. Pernul. Analyzing privacy in social networks–an interdisciplinary approach. *Third International Conference on Social Computing*, pages 1327–1334, Oct 2011.

[14] Pietro Panzarasa and Tore Opsahl and Kathleen M. Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *JASIST*, 60(5):911–932, 2009.

[15] Wei Peng, Feng Li, Xukai Zou, and Jie Wu. A two-stage deanonymization attack against anonymized social networks. *IEEE Transactions on Computers*, 63(2):290–303, Feb 2014.

[16] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, 1998.

[17] Peter J. Slater. Leaves of trees. *Congressus Numerantium*, 14:549–559, 1975.

[18] Klara Stokes and Vicenç Torra. Reidentification and k-anonymity: a model for disclosure risk in graphs. *Soft Comput.*, 16(10):1657–1670, 2012.

[19] Andrey Gubichev, Srikanta Bedathur, Stephan Seufert, and Gerhard Weikum. Fast and Accurate Estimation of Shortest Paths in Large Graphs. *The 19th ACM International Conference on Information and Knowledge Management (CIKM '10).*, pages 499–508, 2010.

[20] Bimal Viswanath, Mainack Mondal, Krishna P. Gummadi, Alan Mislove, and Ansley Post. Canal: Scaling social network-based sybil tolerance schemes. *The 7th ACM European Conference on Computer Systems*, pages 309–322, 2012.

[21] Douglas B. West. Introduction to Graph Theory. *Prentice Hall, 2nd Edition*, 2000.

[22] Xintao Wu, Xiaowei Ying, Kun Liu, and Lei Chen. A survey of privacy-preservation of graphs and social networks. *Managing and Mining Graph Data*, pages 421–453, 2010.

[23] Ismael G. Yero, Alejandro Estrada-Moreno, and Juan A. Rodríguez-Velázquez. The k-metric dimension of a graph: Complexity and algorithms. *arXiv:1401.0342 [math.CO].*

[24] Xiaowei Ying, Kai Pan, Xintao Wu, and Ling Guo. Comparisons of Randomization and K-degree Anonymization Schemes for Privacy Preserving Social Network Publishing. *The 3rd Workshop on Social Network Mining and Analysis*, pages 10:1–10:10, 2009.

[25] Xiaowei Ying and Xintao Wu. Randomizing Social Networks: a Spectrum Preserving Approach. *SIAM International Conference on Data Mining*, pages 739–750, 2008.

[26] Xiaowei Ying and Xintao Wu. On Link Privacy in Randomizing Social Networks. *Advances in Knowledge Discovery and Data Mining*, pages 28–39, 2009.

[27] Haifeng Yu, Phillip B. Gibbons, Michael Kaminsky, and Feng Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. *IEEE/ACM Trans. Netw.*, 18(3):885–898, June 2010.

[28] Bin Zhou and Jian Pei. Preserving privacy in social networks against neighborhood attacks. *IEEE 24th International Conference on Data Engineering*, pages 506–515, April 2008.

[29] Bin Zhou, Jian Pei, and WoShun Luk. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *SIGKDD Explor. Newsl.*, 10(2):12–22, December 2008.

[30] Lei Zou, Lei Chen, and M. Tamer Özsu. K-automorphism: A general framework for privacy preserving network publication. *The VLDB Journal*, 2(1):946–957, August 2009.