Using Identity Separation Against De-anonymization of Social Networks

Gábor György Gulyás[*,**], **Sándor Imre**[**]

[*]Laboratory of Cryptography and Systems Security, BME, Hungary

[**]Mobile Communications and Quantum Technologies Laboratory, BME, Hungary

E-mail: gulyas@crysys.hu, imre@hit.bme.hu

Abstract. Due to the nature of the data that is accumulated in social networking services, there are a great variety of data-driven uses. However, private information occasionally gets published within sanitized datasets offered to third parties. In this paper we consider a strong class of de-anonymization attacks that can re-identify these datasets using structural information crawled from other networks. We provide[*] the model level analysis of a technique called identity separation that could be used for hiding information even from these attacks. We show that in case of non-collaborating users ca. 50% of them need to adopt the technique in order to tackle re-identification over the network. We additionally highlight several settings of the technique that allows preserving privacy on the personal level. In the second part of our experiments we evaluate a measure of anonymity, and show that if users with low anonymity values apply identity separation, the minimum adoption rate for repelling the attack drops down to $3 - 15\%$. Additionally, we show that it is necessary for top degree nodes to participate.

Keywords. social networks, privacy, de-anonymization, identity separation

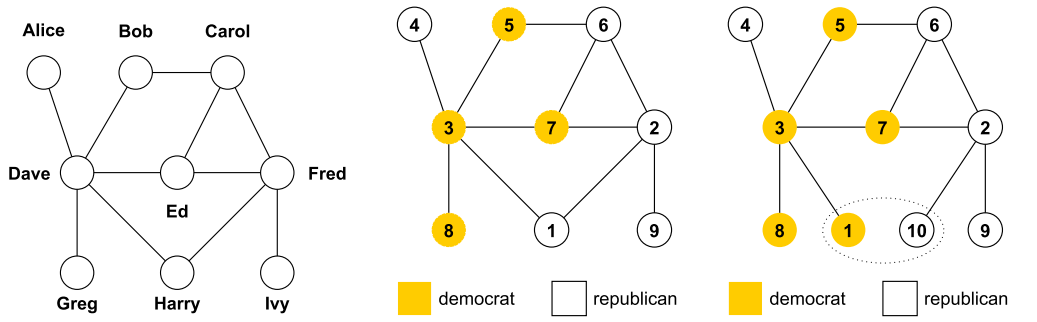
1 Introduction

Social network based services provide interfaces for managing social relationships, while others focus on enhancing collaboration between their users. A common and useful feature of these services is that they are supported by an underlying graph structure. However, social media also serves as an optimal platform for commercial surveillance, and recent cases of government surveillance have also been confirmed [4]. Therefore it is crucial to investigate privacy issues beyond the use of related settings.

Here, we consider how the meta-data of relationships can be abused. For example, someone can obtain a sanitized dataset containing private attributes without explicit identifiers, which was previously released for business or research purposes. For a malicious party this is an opportunity to re-identify nodes in the dataset by using their relationships, in order to monetize the private data or use it another way.

The basic idea for performing this type of attack is to use structural data from another (social) network to execute an iterative re-identification algorithm. While it is not trivial how to execute network alignment efficiently for large networks (over tens of thousands

[*]The current work is the extended version of [15].



(a) Public network (as auxiliary data)

(b) Anonymized network

(c) Anonymized network with identity separation

Figure 1: Datasets for the example of de-anonymization.

of nodes), several attacks have been published recently that are able to breach user privacy at large-scale [5, 20–24]. Then the attacker can learn private attributes, relationships, membership statuses and can use this information without limitations.

We illustrate how these attacks work on a simple example. Let us consider an attacker, who obtains datasets as depicted on Fig. 1a (background knowledge) and Fig. 1b (sanitized dataset), wishing to learn an otherwise inaccessible private attribute by structural de-anonymization: who is a democrat or republican voter in the public network. Initially, the attacker re-identifies (or maps) $v_{Dave} \leftrightarrow v_3$ and $v_{Fred} \leftrightarrow v_2$ as they have globally the highest matching degree values in both networks. Then he continues with local re-identification by inspecting nodes related to the ones already re-identified. Therefore, he picks v_{Ed}, who is the highest degree common neighbor of (v_{Dave}, v_{Fred}), and then it is mapped as $v_{Ed} \leftrightarrow v_7$, as v_7 is the only node neighboring v_2, v_3 and have a degree of 3. This algorithm can continue iterating through unmapped nodes, resulting in discovering further possible mappings (e.g., $v_{Harry} \leftrightarrow v_1, v_{Carol} \leftrightarrow v_6$).

In our work, we consider how a privacy-enhancing method related to the identity partitioning technique [10, 17, 25, 26], called identity separation, can be used to tackle such re-identification attacks. Identity separation allows a user to have multiple unlinkable profiles in the same network, which results in multiple unlinkable nodes in sanitized graphs also (i.e., as the service provider should also be unaware of the link between the identities). The effect of identity separation is demonstrated by the example of Harry on Fig. 1c. We use simulation evaluation with the attack algorithm presented in [21] (later referred to as Nar09) to provide the model level analysis of identity separation. However, as applying the proposed strategies manually is difficult, this should be supported by software (e.g., browser extension or standalone application) that allows the ease of use while following the necessary steps in order to maintain network or user privacy. Designing such a system is possible and feasible, but a complex task; the detailed elaboration of such a system is beyond the scope of this work.

This paper is the extended version of [15], where we provided the following contributions on the basic analysis of non-cooperative identity separation. By running realistic simulation measurements on datasets obtained from three real-world networks, we characterized how resistant the Nar09 attack is against different features of identity separation, i.e., splitting nodes or deleting edges. By evaluating multiple possible strategies we could achieve

information disclosure of 2.33% and lower for nodes using identity separation[1]. In addition, we showed the effectiveness of a simple decoying strategy against Nar09 that allows minimizing information disclosure in a user-controllable way. Unfortunately, it also turned out that at least $40 - 50\%$ of non-cooperating users need to participate in order to make identity separation repel the attack on the network level.

In this paper, we provide further details of the analysis of identity separation. In particular, we focus on the case when the participants of the network collude, and our main goal is to lower the minimum adoption rate of users for stopping the attack compared to the non-cooperative setting. Our main contributions in this paper are the following:

- We add multiple details regarding the settings of non-cooperative identity separation, and analyze the effectiveness of identity management patterns observed in real-world ego-networks (from Twitter).
- Based on recent results in [16], we evaluate non-cooperative identity separation against changing the initialization settings of the algorithm that is under control of the attacker. We furthermore analyze how the level of noise caused by identity separation influences the probability of large-scale propagation, and how this could be handled by the attacker.
- We evaluate two measures of importance in details, namely node degree and an anonymity measure (denoted LTA_A, that we introduced in [14]). We show both have strong levels of correlation with re-identification rates of nodes in large networks. Furthermore, in networks where the number of low degree nodes is proportionally smaller, LTA_A has significantly better correlation with re-identification rates.
- We analyze cooperation between users that have a low level of anonymity according to LTA_A. We manage to decrease the minimum adoption rate to $3 - 15\%$ (depending on network topology), in order to decrease the percent of correctly re-identified nodes under 5%. We furthermore analyze if the attacker could enhance results by changing the initialization method, but according to our measurements, this could add only a little advantage for the attacker.
- Finally, we show that it is essential that high degree nodes participate in cooperation.

The paper is organized as follows. In Section 2 we discuss related work, and background information is presented in Section 3. In Section 4 we provide methodology of our evaluation. In Section 5 we characterize the sensitivity of the attack algorithm against features of identity separation. In Section 6 and 7 we provide the detailed analysis of the non-cooperative and cooperative identity separation, and in Section 8 we conclude and discuss future work.

2 Related Work

2.1 De-anonymization Attack Algorithms

The first structural de-anonymization attack designed specifically for re-identifying a large fraction of nodes is the one proposed by Narayanan and Shmatikov in 2009 [21] (Nar09). The authors in their main experiment re-identified 30.8% of nodes being mutually present

[1] We had repeated the experiments in [15] which lead to similar, but slightly smaller disclosure rates.

in a Twitter and a Flickr crawl with a relatively low error rate of 12.1%. Since their work, several attacks with the same basic principles have been published [6, 19, 20, 22–24].

These attacks differ in many aspects, however, in general they consist of two sequentially executed phases: one for initialization (seed or global identification phase), and another that propagates re-identification in an iterative manner (propagation or local re-identification phase). The goal of the first phase is to find globally outstanding nodes (i.e., the seeds), for example by their degree. After the seed set reaches a sufficient size, the second phase starts to extend the seed set in an iterative way, locally comparing nodes being connected to others already re-identified.

Here we include the most relevant related works appeared since [21]. Narayanan et al. in 2011 presented another variant of their attack [20] specialized for the task of working on two snapshots of the same network (with a higher recall rate). Another proposal of Wei et al. [23] challenged Nar09; however, their attack is only evaluated on small networks and against a light edge perturbation procedure, instead of the more realistic one proposed in [21]. The latter deletes edges from both networks (e.g., node and edge overlaps can be as low as 25%), while in [23] edges are only added to the target network (up to 3%) without deletion; this is a remarkable deficiency. In addition, further experiments are needed to show if the algorithm in [23] performs better in large networks having tens of thousands of nodes – if the seed-and-grow attack is still feasible on such datasets.

It has been shown by Srivatsa and Hicks that location traces can also be re-identified with similar methods [24]. In their work they succeeded in identifying 80% of users by building anonymous networks by observing location traces, and using explicit social networks for de-anonymization. For supporting structural re-identification attacks, some works additionally involve user content and attributes in the process of re-identification [6, 9, 12, 19].

Pedarsani et al. proposed a novel type of attack that can work without any initial input such as seeds [22]. In fact, in their design, seeding is incorporated into the propagation phase, as the initial propagation step that starts identifying top nodes according to a given node fingerprint measure. This is very reminiscent of a seed based initialization mechanism where top nodes are selected, e.g., by their degree, betweenness or other characteristics. Their algorithm could reach acceptable error rates on a small test network (consisting of $2,024$ nodes) when the probability that an edge coexisted in both networks was $\beta = 0.85$. In our measurements this is equivalent for the cases with $\alpha_e = 0.75$ (edge overlap; see the detailed description in Section 4), but as the perturbation strategy we apply consists node deletion also, this β is significantly lower, smaller than what algorithm in [22] accepts.

2.2 User-Oriented Solutions for Preventing De-anonymization

We consider user centered privacy protection mechanisms for preventing de-anonymization, ones that can be applied to existing services (instead of graph sanitization applied by the service provider). For instance, Scramble is a good example for solutions being independent of the service provider and allowing a fine-grained access of social data [8]. Otherwise, one might consider using revised service models, such as distributed social networks like Safebook [11]; however, these services are more difficult to introduce.

Beato et al. proposed the friend-in-the-middle model, where proxy-like nodes serve as mediators to hide connections, enabling to repel the attack on a network level [7]. Their concept could be also implemented as an external tool that could be used in existing social networking services. The viability of the FiM model is presented (successfully) on two snapshots of the Slashdot network [3] (which we also used, see Section 4). However, identity separation allows more than hiding connections, even hiding profile information

beside relationships [10]. As this allows finer-grained management of information, with less cooperation – this can even enable the protection of a single individual.

The concept of privacy-enhancing identity management was developed in details within the framework of the PRIME Project [18], including how identity partitioning and separation could be implemented in various contexts and services. The possible use of identity separation in social networks was introduced in [17] by Gulyás et al., where the authors proposed a modified social network model with a non-flat structure. The works of van den Berg and Leenes in [25,26] provided further details on identity partitioning, especially focusing on access control and division of information shared.

Previously, we have analytically showed that identity separation is an effective tool against clique based seeding mechanisms [13]. More recently, we also analyzed the protective strength of non-cooperative identity separation against the propagation phase of Nar09 [15] with simulation on datasets obtained from three different social networks. We have shown that while almost half of the users are required to repel the attack (and retain network privacy); it is possible to effectively hide information from the attacker even for a few nodes if the proper settings are applied. In other words, with non-cooperative identity separation personal privacy can be effectively preserved.

In [14] we have characterized and evaluated a measure reflecting the anonymity level of nodes, called Local Topological Anonymity (LTA). From another point of view, LTA gives how well a node is hidden in its neighborhood including friends-of-friends. We evaluated three LTA variants on small networks (up to the size of $10,000$ nodes) and for the best candidate, we measured an average Pearson correlation of -0.421. However, the work in [14] lacked measurements for large networks of tens of thousands of nodes.

Several works in the literature focus on analyzing different properties structural de-anonymization attacks. Recently we have shown that seeding parameters are an important aspect of the de-anonymization procedure, as they have a significant effect on the overall results [16]. Thus, it should be detailed both for comparing new attack schemes (e.g., [23]) and for evaluating protection mechanisms (e.g., [7, 15]). Therefore we analyzed our findings from this aspect, too.

3 Background and Notation

3.1 Details of the Nar09 algorithm

In the original experiment the Nar09 attack used 4-cliques of high degree nodes as seeding [21]. Its local re-identification phase works similarly as described in the example of Section 1: it is based on a propagation step which is iterated on the neighbors of the seed set until new nodes can be re-identified (already identified nodes are revisited). In every iteration, candidates for the currently inspected source node are selected from target graph nodes, sharing at least a common mapped neighbor with it. At this point the algorithm calculates a score based on cosine similarity for each candidate. If there is an outstanding candidate, before it is selected as a match, a reverse checking is executed to verify the proposed mapping from a reversed point of view. If the result of reverse checking equals the source node, the new mapping is registered.

Probably the most important parameter of Nar09 is the Θ threshold, controlling the ratio of true positives (recall rate) and false positives (error rate). The lower Θ is the less accurate mappings Nar09 is willing to accept, as Θ controls how outstanding the best candidate should be from the others.

Seeding method and size also has a great effect on overall results. In our previous work in [16], we provided details for various methods, and showed that the overall recall rate is influenced by several properties of seed nodes, such as the structural relation between them (e.g., cliquish structure or neighboring), and their global properties (e.g., node degree, betweenness centrality score). Results are also shown to be dependent on network size and structure. Our experiments highlighted seeding methods that were top performers on the large networks, regardless of network structure (e.g., nodes with highest degree and betweenness centrality scores). We used these in our experiments.

3.2 Notation and Definitions

Given a sanitized graph G_{tar} (target graph) to be de-anonymized by using an auxiliary data source G_{src} (where node identities are known), let $\tilde{V}_{src} \subseteq V_{src}, \tilde{V}_{tar} \subseteq V_{tar}$ denote the set of nodes mutually existing in both. Ground truth is represented by mapping $\mu_G : \tilde{V}_{src} \to \tilde{V}_{tar}$ denoting relationship between coexisting nodes, and $\lambda_G : \tilde{V}_{src} \rightrightarrows \tilde{V}_{tar}$ denote mappings between nodes in G_{src} and the sets of their separated identities in G_{tar}. Running a deterministic re-identification attack on (G_{src} , G_{tar}) initialized by seed set $\mu_0 : V_{src} \to V_{tar}$ results in a re-identification mapping denoted as $\mu : V_{src} \to V_{tar}$. We denote the set of nodes adopting identity separation as $V_{ids} \subseteq V_{tar}$.

We use two measures for evaluating simulation results. The *recall rate* reflects the extent of re-identification (this can be used due to constantly negligible error rates), describing success from an attacker point of view. Identity separation is a tool for separating sensitive attributes or relationships, hence the quantity of information the attacker gained access to should also be concerned, which is quantified by the *disclosure rate*. This describes overall protection efficiency from a user point of view.

Now we describe the method of calculation of these rates. The *recall rate* is calculated by dividing the number of correct identifications with the number of mutually existing nodes (seeds are excluded from the results). The score of a node regarding a given re-identification mapping μ can be expressed as:

$$s(v,\mu) = \begin{cases} 0 & \text{if } \nexists \mu(v) \\ 1 & \text{if } \mu(v) = \mu_G(v) \vee \mu(v) \in \lambda_G(v) \\ -1 & \text{if } \mu(v) \neq \mu_G(v) \wedge \mu(v) \notin \lambda_G(v) \end{cases} . \tag{1}$$

We can now quantify the *recall rate* of an attack based on the score function. During the evaluation we excluded seed nodes from the results, as such mappings are not additionally discovered ground truth mappings. Thus, the recall rate of the attack resulting in mapping μ as

$$R(\mu) = \sum_{\forall v \in \tilde{V}_{src}} \frac{s(v,\mu) \cdot \max(0, s(v,\mu))}{|\tilde{V}_{src}|}. \tag{2}$$

We denote the maximum of recall as R_{max}.

The *disclosure rate* can be calculated in a similar manner. As current identity separation models are bond to structural information, we use a measure reflecting the average percent of edges that the attacker successfully revealed (this can be extended for other types of information in other experiments, e.g., sensitive profile attributes). This can be quantified for an individual node $v_n \in \tilde{V}_{ids}$ (before identity separation) where the partial identity $v_{n\setminus i}$ could be identified in mapping μ as

$$d(v_n, \mu) = \begin{cases} 1 & \text{if } \exists \mu(v) \wedge \mu(v) = \mu_G(v) \\ \frac{deg(v_{n \setminus i})}{deg(v_n)} & \text{if } \exists \mu(v_n) \in \lambda_G(v_n) \\ 0 & \text{otherwise} \end{cases} . \quad (3)$$

By using this function we can now define the disclosure rate of the attacker over the nodes applying identity separation in order to evaluate the efficiency of the protective method. Thus w.r.t. mapping μ let the disclosure rate be defined as

$$D(\mu) = \sum_{\forall v_n \in \tilde{V}_{ids}} \frac{d(v_n, \mu)}{|\tilde{V}_{ids}|}. \quad (4)$$

We also consider the *re-identification rate of a node* v in a series of experiments ν as:

$$S(v) = \sum_{\forall \mu \in \nu} s(v, \mu), \quad (5)$$

where $s(v, \mu)$ can theoretically take arbitrary values in the series of ν. However, as the Nar09 algorithm is quite deterministic, negative and positive values of $s(v, \mu)$ almost never occur for the same node in a series of experiments.

In this paper we also use *measure of importance* for nodes. We consider a node to be important, if it can be re-identified with high probability. For this reason we build on the concept of anonymity, where it can be assumed that a node with a low level of anonymity can be re-identified with high probability. To the best of our knowledge, the only anonymity definition in the current context is given in [14], which is called *Local Topological Anonymity* (LTA). LTA value of a node aims to reflect how structurally hidden the node is in its neighborhood (i.e., compared to friends-of-friends), as the Nar09 algorithm in [21] examines this property for making new matches. The lower the LTA of a node is, the easier it can be re-identified.

4 Methodology

In this section, we present the methodology we use in our work.

4.1 Social Network Data and Modeling Identity Separation

During our experiments we used multiple datasets with different characteristics in order to avoid related biases. In addition, we used large networks, as brute-force attacks can be mounted against smaller ones. We obtained two datasets from the SNAP collection [3], namely the Slashdot network crawled in 2009 (82,168 nodes, 504,230 edges) and the Epinions network crawled in 2002 (75,879 nodes, 405,740 edges). The third dataset is a subgraph exported from the LiveJournal network crawled in 2010 (at our dept.; consisting of 66,752 nodes, 619,512 edges). All datasets were obtained from real networks in order to maintain our measurements being realistic.

For modeling identity separation, it would be desirable to analyze real-world data on user behavior, but to our knowledge, such datasets are unavailable and there are no trivial ways of crawling one. However, some data on ego networks, which is a similar functionality to identity separation, is available from Google+, Twitter and Facebook [3]. By analyzing this

		Slashdot				Epinions				LJ66k			
		α_e				α_e				α_e			
		0.25	0.5	0.75	1.0	0.25	0.5	0.75	1.0	0.25	0.5	0.75	1.0
α_v	0.25	0.6	2.6	11.7	19.8	1.1	5.1	10.9	14.8	0.8	6.3	19.5	27.6
	0.5	0.4	19.8	36.5	47.5	0.9	17.2	25.9	32.6	0.6	24.8	35.7	54.4
	0.75	0.3	33.6	50.7	60.4	0.6	25.3	36.1	44.4	0.7	33.9	57.9	78.7
	1.0	0.3	30.1	58.9	68.3	0.4	31.2	43.2	52.5	1.5	37.7	75.2	88.5

Table 1: Recall rates were proportional to the overlap between G_{src} and G_{tar}: the less perturbation is used (resulting higher overlaps) the higher recall rates are.

data, we found that the number of circles has a power-law distribution, and duplication of connections across contact groups is not widely used [15].

As we cannot draw strong conclusions from these observations regarding identity separation (as the data lacks patterns on the use of hidden connections), we use the probability based models we previously introduced in [13] for deriving test data from real-world datasets featuring identity separation. These models capture identity separation as splitting a node, and assigning previously existing edges to the new nodes. The number of new identities is modeled with a random variable Y (with no bounds on distribution), which we either set to a fixed value, or model it with a random variable having a power-law-like distribution. In our work it is assumed, that the identity separation is done in secret, and cannot be learned by the attacker from auxiliary sources.

For edge sorting, there are four models in [13] regarding whether it is allowed to delete (i.e., an edge becomes private) or to duplicate edges, from which we used three in our experiments. The basic model is simple and easy to work with, as it consists a simple redistribution of edges between the new identities (no edge deletion or duplication allowed). We also used the realistic model to capture real-life behavior, too (both operations are allowed). We additionally used the best model describing privacy oriented user behavior (no edge duplication, but deletion allowed). While we explicitly omitted the worst model (edge duplication only), we must note that the behavior patterns from the Twitter network are the closest to this model, as these patterns contain duplication but no deletion.

4.2 Data Preparation

In order to generate the test data, first we derived a background knowledge (G_{src}) and a target graph (G_{tar}), having desired overlap of nodes and edges, and then modeled identity separation on a subset of nodes in the target graph. For the first part, we used the perturbation strategy proposed by Narayanan and Shmatikov [21], as we found their method to be producing fairly realistic test data. Their algorithm takes the initial graph to derive G_{src}, G_{tar} with the desired fraction of overlapping nodes (α_v), and then edges are deleted independently from the copies to achieve edge overlap α_e. By knowing the original graph, the ground truth μ_G can be easily created at this point.

We found $\alpha_v = 0.5$, $\alpha_e = 0.75$ to be a good trade-off at which a significant level of uncertainty is present in the data (thus life-like), but the Nar09 attack is still capable of identifying a large ratio of the co-existing nodes. Fraction of correctly identified nodes are presented for various settings in all the test network in Table 1.

Identity separation is then modeled on the target graph by uniformly sampling a given percent of nodes with at least $deg(v) = 2$ (this ratio is maintained for the ground truth

nodes), and then nodes are split and their edges are sorted according to the settings of the currently used model. This results in extending the ground truth mapping μ_G with λ_G by recording identity separation operations.

4.3 Calibrating Simulations

By comparing the directed and undirected versions of Nar09, we found little difference in results. Therefore, due to this reason and for sake of simplicity, in our experiments we used undirected networks. Additionally, in each experiment we created two random perturbations, and run simulations three times on both with a different seed set (it is indicated when other settings were used). We observed only minor deviations in results, usually less than a percent.

In accordance with our earlier experiments [15, 16], we measured fairly low error rates even for small values of Θ, therefore we worked with $\Theta = 0.01$. The ground truth error rate stayed around 1-2% for large networks, usually less than 5%.

In simulations, by default we applied random seed selection with high degree nodes, where nodes are selected from the top 25% by degree (denoted as `random.25`). Seed set size was selected constantly for a thousand nodes, as this proved to be robust in all networks [16]. For the simulation of stronger attackers, top degree nodes were selected as seeds (denoted as `top`) or nodes with the highest betweenness values also having degree in the top 10% (denoted as `betwc.1`), as these methods proved to be the most effective in our datasets [16].

5 Sensitivity Measurement of the Nar09 Algorithm

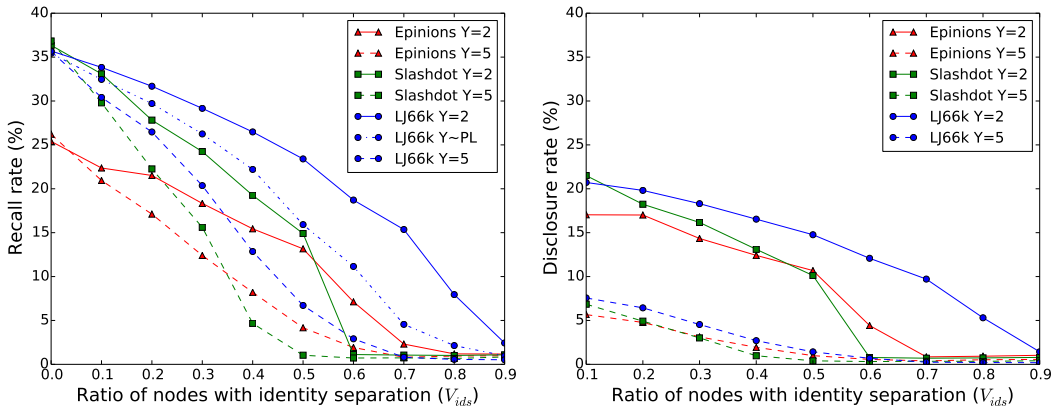
In order to discover the strongest privacy-enhancing identity separation mechanisms, we investigated the efficiency of features in different models against the Nar09 algorithm. Initially we executed measurements with different parameter settings of the perturbation algorithm provided in [21] (see recall rates in Table 1). As these small matrices are not diagonal symmetric, the algorithm seems to be more sensitive to edge deletion than to node deletion. Thus, in our experiments, we analyzed these features separately to see their effect.

5.1 Characterizing Sensitivity to the Number of Identities

First, we tested the Nar09 algorithm against the *basic model with uniform edge sorting probability* on all networks having a ratio of users applying identity separation of $V_{ids} \in [0.0, 0.9]$. For the selected users a fixed number of new identities were created ($Y \in \{2, 5\}$). We summarized results on Fig. 2, and omitted results for $Y \in \{3, 4\}$ which can be easily interpolated.

Against our expectations, the basic model with $Y = 2$ and uniform edge sorting probability turned out to be ineffective in stopping the attack. For the Epinions and Slashdot networks the recall rate mildly decreased until the ratio of privacy-protecting users reached circa $V_{ids} = 0.6$. For the LiveJournal graph the recall rate shows relevant fault tolerance of the attack (probably because of network structure, see Fig. 14), e.g., 15.36% are still correctly identified for $V_{ids} = 0.7$. When participating users had five new identities, recall rates dropped below 10% at $V_{ids} = 0.5$ for all networks.

Edges sorting was also tested with a power-law distribution having $Y = 5$. These experiments resulted in a slightly higher true positive rate, which is understandable: if edges



(a) Recall rates

(b) Disclosure rates

Figure 2: Experimental results gained by using the basic identity separation model.

are not uniformly distributed it is more likely for an identity to have more of the original edges than the others (with higher chances to be re-identified). In another comparative experiment we modeled a variable number of new identities with power-law-like distribution with $Y \in [2, 5]$ and uniform edge sorting probability. Results were properly centered between cases $Y = 2$ and $Y = 5$ as the LiveJournal example shows on Fig. 2a.

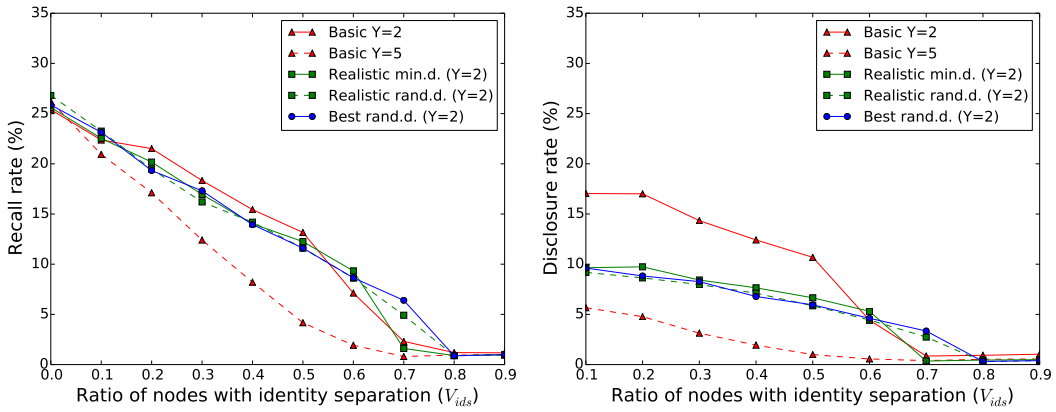
Although by inspecting recall rates the basic model seems ineffective in impeding the attack, the disclosure rates yield better results. As shown on Fig. 2b, disclosure rates are significantly lower compared to recall rates[2]. From this point of view using the basic model with $Y = 5$ and uniform edge sorting probability provides strong protection for even a small ratio of applying users: the disclosure rate is at most 7.56% when $V_{ids} = 0.1$. By comparing the results of the two measures, we conclude that by using the basic model it is not feasible to repel the attack. However, by using a higher number of identities the access of the attacker to information can be effectively limited.

5.2 Sensitivity to Edge Deletion

The realistic and best models were used to test the Nar09 against additional edge perturbation by identity separation [13]. We used three different settings in our experiments, as details are not explicitly defined in [13]. For all of them edge sorting probabilities are calculated according to multivariate normal distribution as $P(X_1 = x_1, \ldots, X_y = x_y) \sim \mathcal{N}_y(\boldsymbol{\eta}, \boldsymbol{\Sigma})$, where y denotes the current number of identities. The value of $\boldsymbol{\eta}$ was set to $(y)^{-1}$ and $\boldsymbol{\Sigma}$ was configured in a way to have higher probabilities for events when the sum of the degrees of the new identities are close to original node degree. In the best model when the sum was higher than the original degree, the distribution was simply recalculated.

We used the following settings. The *realistic model with minimal deletion*, in which every edge is assigned to one identity, and if there is still ample space left for edges, edges are assigned again randomly. In this setting edges are not deleted unless it is necessary. In the setting of the *realistic model with random deletion* new identities take a portion of edges proportional to $(x_1, \ldots, x_y)$, leading to delete unassigned edges proportionally to $\prod(1 -$

[2]Note: as the disclosure rate is measured for $\forall v \in dom(\lambda_G)$, results start from $V_{ids} = 0.1$.



(a) Recall rates

(b) Disclosure rates

Figure 3: Effect of edge deletion compared to the basic model (Epinions dataset).

$\frac{x_i}{deg(v)}$) , where v denotes the node before identity separation. We also included a setting with the best model called the *best model with random deletion*. We emphasize that none of these settings capture aggressive edge deletion, but it might be interesting to investigate such settings in the future.

We executed simulations for these models with $Y = 2$, and found that recall rates strongly correlate with results of the basic model (although being slightly better); thus, these models are also incapable of repelling the attack on the network level (see Fig. 3a). Fortunately, disclosure rates are better compared to the basic model, e.g., results for the Epinions network are depicted on Fig. 3b. We conclude that while these models are also incapable of stopping large-scale propagation, they yet perform better in privacy protection.

6 Non-cooperative Identity Separation for Preventing Information Disclosure

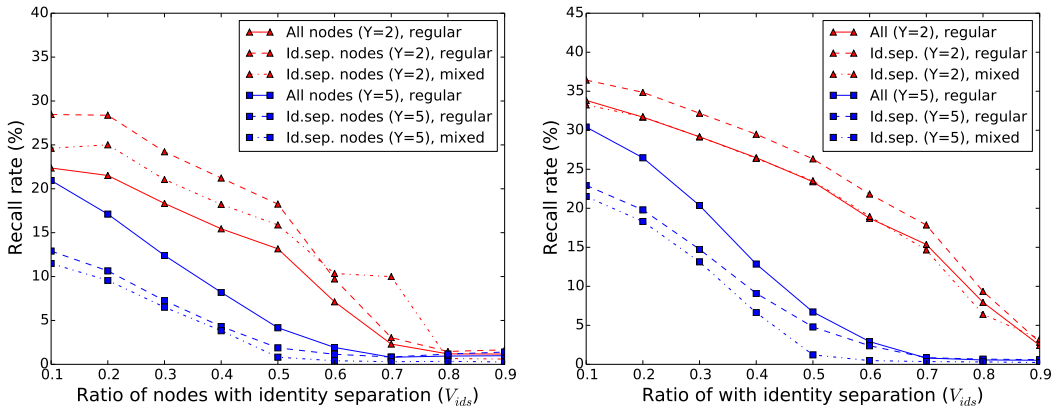
We analyze the limits of non-cooperative privacy-enhancing identity separation on our datasets.

6.1 In the Search of Privacy-Enhancing Strategies

Identity separation can be applied in multiple ways, and we analyze how these affect overall results.

6.1.1 Why the Use of the Basic Model with $Y = 2$ Should be Reconsidered

While conducting our analysis, we found a case when the recall rate was notably higher for users of identity separation ($\forall v \in dom(\lambda_G)$) compared to the overall recall ($\forall v' \in \tilde{V}_{src}$). For low values of V_{ids} this difference in the recall was almost constant and decreased for higher values (see values for regular seeding on Fig. 4).



(a) Epinions network

(b) LiveJournal network

Figure 4: Comparison of recall rates for all nodes and ones using identity separation shows that using a low number of new identities leads is counterproductive and leads to higher recall rates than average ($Y = 2$). Therefore users should be advised to use a higher number of new identities as results suggest.

Eventually, this turned out to be caused by the seeding strategy. Throughout our experiments we used seeds that were not affected by identity separation, but after changing to mixed seeding with an equal ratio of seeds selected from $dom(\mu_G)$ and $dom(\lambda_G)$, while the overall recall rate remained unchanged, the difference disappeared for the LiveJournal and Slashdot networks, and significantly decreased for the Epinions (examples showed on Fig. 4). From the user perspective, the disclosure rates did not vary much by changing the seeding strategy.

This finding has an interesting impact for the attacker on choosing the seeding strategy. Using a regular seeding mechanism is a natural choice, and adding fault tolerance against identity separation is not trivial. Therefore, by using the natural choice of seed identification, the attacker will also have a higher rate of correct identification for nodes protecting their privacy (with a low number of new identities). We note that the seeding mechanism should be chosen with caution. Analysis in [13] shows that the clique-based seeding method [21] is not resistant to identity separation.

The core message of this finding is important for users aiming to protect their privacy: they should use higher number of new identities. As examples on Fig. 4 shows that recall rates for users with $Y = 5$ was lower than the network average, and even if an attacker uses a mixed seeding mechanism it is also counterproductive.

6.1.2 Recall Rate Comparison for Multiple Models in Parallel

Previously, we described experiments in which settings of different identity separation models were used homogeneously. We investigated if the observed differences remain when multiple settings are allowed in the same network. The following models were used: basic model with uniform edge sorting probability (34% of V_{ids}), realistic model with random deletion (33% of V_{ids}), best model with random deletion (33% of V_{ids}). We found that for the users of each setting was proportional to results measured in previous experiments,

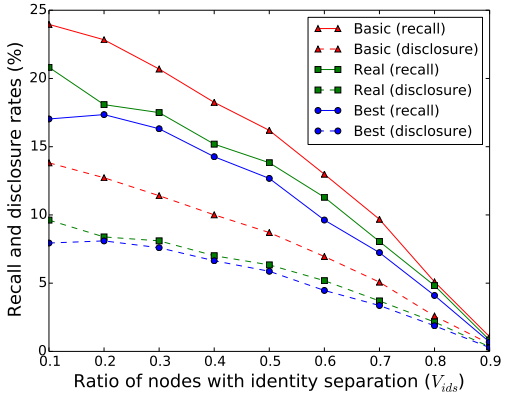


Figure 5: When multiple models are used in parallel, users get results that are proportional to the model they use. Experiments are run on the LJ66k dataset.

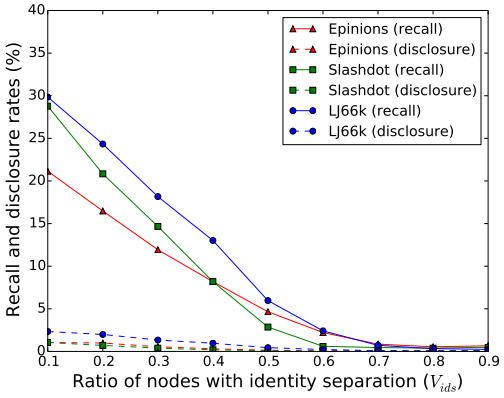


Figure 6: Even the best model with $Y = 5$ cannot repel the attacker on a network level, just by involving the majority of users; however, privacy-enhancing users still achieve low disclosure rates.

$P(X_0, X_1)$	$x_0 = 0$	$x_0 = 1$
$x_1 = 0$	0.00	0.16
$x_1 = 1$	0.82	0.02

(a) Distribution for a node v' we picked from our sample. The node had $deg(v') = 50$, $y_{v'} = 2$.

	$y = 2$	$y = 3$	$y = 4$	$y = 5$
$P(Y = y)$	0.49	0.27	0.14	0.08

(b) Distribution of new identities applied from the Twitter dataset.

Table 2: Characteristics for applying behavior patters from the Twitter dataset.

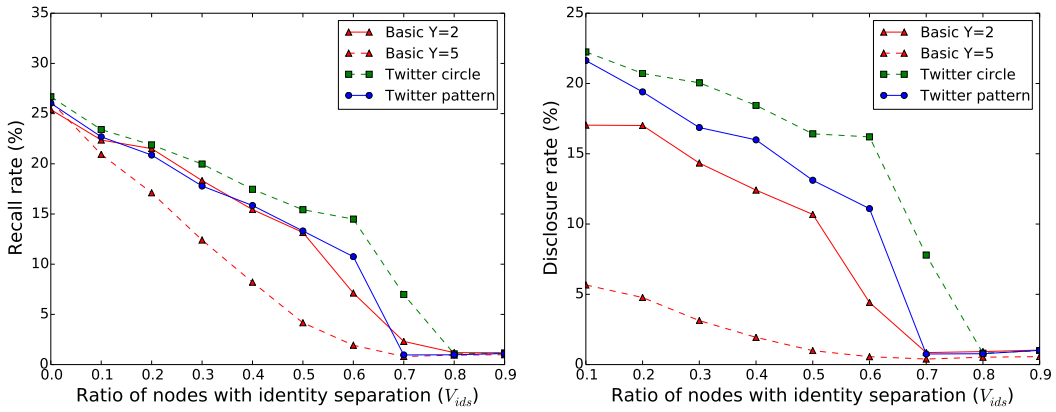
for instance, users of the best model achieved the lowest recall and disclosure rates. Simulation results in the LiveJournal graph are plotted on Fig. 5 for demonstration (results were measured for homogeneous groups consisting of nodes having the same setting).

6.1.3 Applying Behavioral Patterns from the Twitter dataset

We applied patterns from the Twitter dataset according to two strategies, in order to see how observed behavior tackles re-identification. In the case we call the *Twitter patterns*, for a given node we randomly select behavior patterns from nodes with a similar degree, which determines the number of new identities and how edges need to be sorted. For an example, see Table 2a.

In the case of the *Twitter circle* strategy, first we calculated the number of new identities $y_{v'}$ of node v' in coherence with the distribution of new identities observed in the data, where Y was limited for rational considerations (see Table 2b). Next, the pattern is selected randomly, with a probability proportional to its relative frequency in the dataset we used. Patterns distributed similarly as in Table 2a.

Our simulation experiments proved these strategies to be less useful against de-anonymization. However, this is not surprising, as there is no edge deletion, but duplication, thus patterns resemble the worst model. We summarized examples of our results in the Epinions dataset on Fig. 7. In addition, it should also be considered that these results show a higher level



(a) Recall rates

(b) Disclosure rates

Figure 7: User behavior patterns in the Twitter egonet dataset are closer to the worst model, thus also results in the Epinions dataset are worse than the basic model.

of privacy-protection based on the Twitter dataset than expected in reality (with the same patterns). We reckoned that a user have a separated identity for each circle in the dataset, however, this is an overestimation of the use of identity separation.

6.1.4 Applying the Best Model with $Y = 5$

We have previously showed that while none of the previously analyzed defense strategies can effectively stop the attack, identity separation can reduce disclosure rates. It also turned out that increasing the number of new identities has a powerful impact on the disclosure rate, while edge perturbation has little, but notable effect. Therefore, from the user point of view, the best model with a high number of identities seems to be a quite effective setting.

We run the best model with $Y = 5$ on all test networks. Results show that even this strategy cannot prevent large-scale re-identification when only a minority of users applies the technique. Instead, for all networks the re-identification rate constant monotone decreased as V_{ids} increased (see Fig. 6). Fortunately, the setting had more convincing results for disclosure rates: even for $V_{ids} = 0.1$ the disclosure rates topped at 2.33%, but were typically around or under 1%. Disclosure values also continued to fall as the ratio of defending users increased.

6.2 Enhancing the Attacker Model

As it turned out recently, the seed selection method is an important part of the attacker model [16]. In this section, the stability property based on the seeding [16] is further elaborated. We show that as the perturbation level induced by identity separation increases, the attacker needs larger seed sets to keep the probability of large-scale propagation high. We also show this probability also depends on the seeding method.

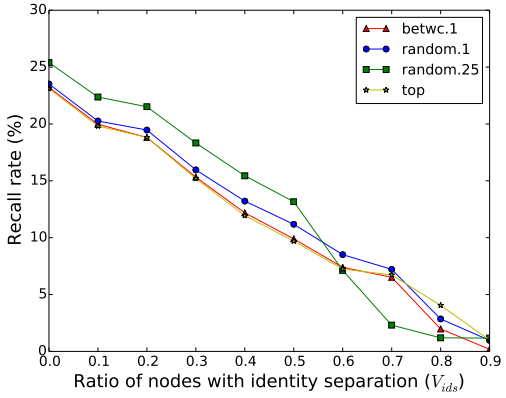


Figure 8: Comparison of advanced seeding methods against `random.25` – using other seeding measures did not lead to significantly different results. Results from the Epinions with basic model (Y=2) are shown.

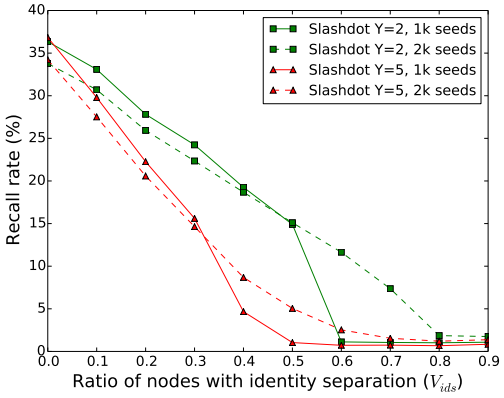


Figure 9: We tested the `random.25` method with higher seed values on the LJ66k dataset against different levels of perturbation created by the basic model with $Y = 2$ and $Y = 5$. Higher numbers of seeds led to better results when V_{ids} was also large.

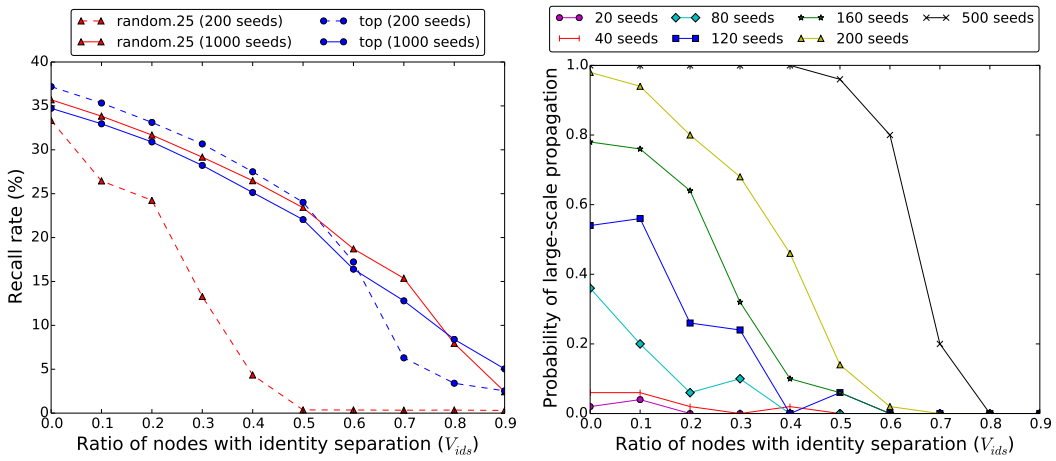
6.2.1 Using Different Seeding Methods

We compared the seeding method used in our experiments (i.e., `random.25`) in order to see if other methods are more robust against the perturbation caused by identity separation. Based on the results of [16], we selected the `betwc.1`, `random.1` (nodes randomly selected from the top 10% by degree) and `top` as advanced methods for comparison. For these only a handful of seed nodes are enough for large-scale propagation. We used a constant seed size of a thousand nodes, similarly to previous experiments.

Results in the Epinions network (with basic model $Y = 2$) are show on Fig. 8. We experienced minor differences only while using different seeding methods. However, it should be noted that advanced methods seem to be a better choice when a higher ratio of users apply identity separation ($V_{ids} \geq 0.6$). This was also true for the recall rates in Slashdot network, and for the disclosure rates in both networks, but only when using the basic model with $Y = 2$. In case when we modeled identity separation with the best model and $Y = 5$ (in any of the datasets), or if we considered the LJ66k network with either $Y = 2$ or $Y = 5$, recall and disclosure rates showed only very minor differences. In these cases the `random.25` seeding method seemed to provide only slightly better results than the others, but essentially there were no differences.

These results alone would not justify the use of seeding methods other than `random.25`. However, when it is not possible to re-identify a large number of seeds initially, other methods should be considered. As Fig. 10a shows, the `top` method was more robust against identity separation than `random.25` when only 200 seeds were available. Results with higher seed number are also plotted for comparison, and the results are from the LJ66k dataset, using basic model with $Y = 2$. Here, `top` have enough seeds even for large V_{ids} values, while `random.25` would need more nodes for seeding as the level of perturbation increases

This is for a simple reason: the greater the level of perturbation is in the current experi-



(a) Recall with lower number of seeds

(b) Seeding stability with `random.25`

Figure 10: If the number of seeds is only 200 nodes (considered stable for $V_{ids} = 0.0$), `top` turned out to be more resistant against identity separation than `random.25`. Probability measurements show that for larger perturbation more seeds are needed for stable seeding, and from this aspect `top` is more redundant than the other causing the difference visible on (a). Experiments were run on the LJ66k dataset with basic model (Y=2).

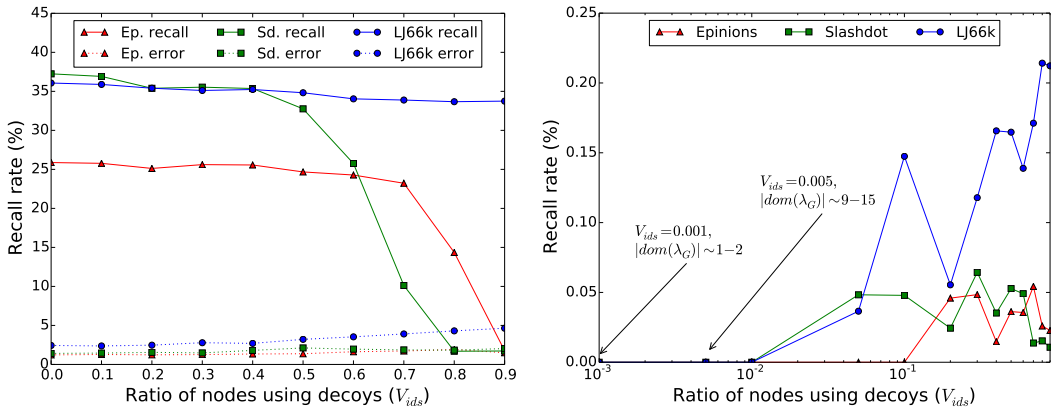
ment, the more seed nodes are required to have stable seeding. In [16] we showed that the `top` selection method needs significantly less nodes for stable seeding than the `random.25` method. We demonstrate the connection between the number of seeds and stability for various V_{ids} on Fig. 10b. Each experiment was run 25 times with different random seed sets for balanced results. Here, we consider seeding to be stable, when the probability of large-scale propagation is approximately 1.0. We consider propagation large-scale when $R(\mu) \geq 0.75 \cdot R_{max}$, i.e., recall rate reaches 75% of the highest observed recall for the given V_{ids}. By running measurements with identical parameters and datasets with $V_{ids} = 0.0$, in [16], we denoted the minimum seed set size required for stable seeding ca. 80 for `top`, and ca. 180 for `random.25` for the LJ66k network. This is coherent with our current measurements.

Further corollary of this finding is that another attacker type needs to be considered: who can search for a seed set consisting of a low number of nodes on a trial-and-error basis until large-scale propagation appears. Even in this case, due to the design of Nar09, the error rate is likely to be low. For example, we managed to reach $R(\mu) = 27.88\%$ ($R_{max} = 36.41\%$) with only 20 nodes selected by `random.25` in the LJ66k network.

6.2.2 Increasing the Seed Size

Another possibility of the attacker to have better results is to use larger seeds sets. We tested the `random.25` seeding method with higher number of seeds (2000 nodes) on the LJ66k dataset against the basic model with $Y = 2$ and $Y = 5$. Our results are shown on Fig. 9.

We found that increasing the number of seed nodes only help when the seeding method is unstable at the given perturbation rate (see Fig. 10b for details), e.g., when using `random.25`



(a) Recall sensitivity to the use of decoys

(b) Recall rate for hidden nodes

Figure 11: The use of decoy nodes only affected de-anonymization when it was used in large-scales, and only a tiny fraction of hidden nodes was re-identified.

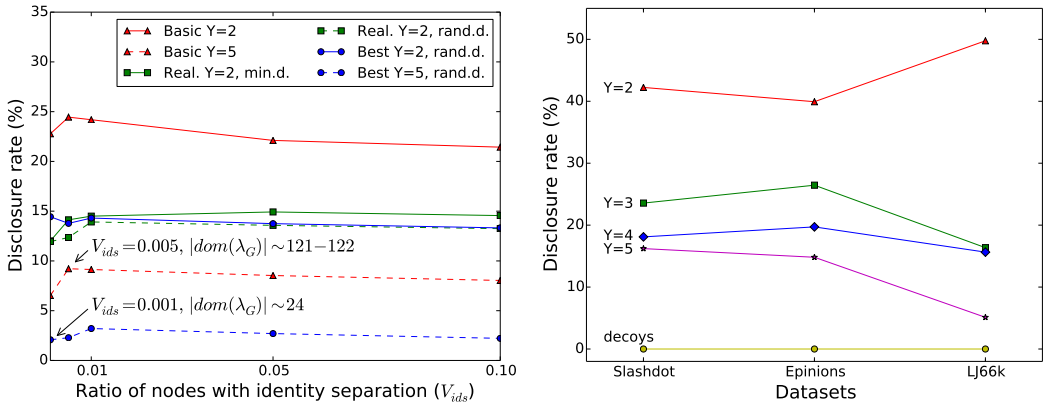
with only 200 seed nodes in the LJ66k network (Fig. 10a). However, we must note that a seed size of 2000 nodes is irrationally large in proportion to the overlap between the anonymized and known datasets in our experiments. For example, in the $Y = 2$ case the overlap size is just around 6000 nodes, which is only three times larger than the seed size.

6.3 Placing the User in the Decision-Making Position

Strategies discussed so far work on statistical basis, and lack user control: it mostly depends on the attacker what he can find. In [15] we proposed a simple model that puts the user into the decision-making position by utilizing decoy identities. We applied the following strategy on nodes $v_i \in \tilde{V}_{tar}$ that have at least 30 neighbors, which narrowed down possible nodes significantly. For example, in LJ66k, even for $R = 0.9$ only $|dom(\lambda_G)| \approx 11.2\%$ of $\tilde{V}_{tar}$. First we create a decoy node v_i^P (public profile) having 90% of acquaintances, representing non-sensitive connections with the goal of capturing the attention of attacker algorithm. Next, a hidden node v_i^H is created having the rest 10% of neighbors for modeling sensitive relationships, and an overlap of 10% with the neighbors of v_i^P.

We rerun simulations with this model subsequently to experiments in [15] (15 runs on every dataset). From the attacker point of view the algorithm achieved misleadingly high recall rates until large number of decoys appeared, while error rates constantly stayed lower than 5%. From the user perspective, privacy-protecting nodes achieved of revealing little sensitive information as shown on Fig. 11b. Recall rates were negligible for hidden nodes, less than 0.25% within all networks in our measurements.

This simple model can be defeated when the attacker optimizes for user strategy. For instance, a new algorithm could be able to discover both v_i^P and v_i^H, or at least one of them with high probability. In that case, given the background knowledge with no identity separation, the attacker can compare neighborhood sizes to decide if the discovered identity has the sensitive attributes. One possible solution for the user is to decrease the certainty of the sensitive attribute; for example, randomly assigning privacy sensitive values to v_i^P and v_i^H.



(a) Low participation rates in LJ66k

(b) Single user with best model or using decoy

Figure 12: In the search of the most effective privacy-enhancing strategies when only adopted by a few. Disclosure rates in the best model are quite competitive (a), even when only a handful of users apply identity separation in such a way. In general, we found that disclosure rates are proportional to the number of identities even when a single user adopts identity separation (b).

Or in another strategy, both nodes v_i^P and v_i^H could be aligned to their context regarding the structure of their neighborhood or sensitive values (e.g., similarly as node v_1 used identity separation on Fig. 1c). This would mean a strategy that would allow one to manage different identities in different networks in a way, that it would not matter which anonymized dataset the attacker has access to, he will learn no sensitive information. We leave these issues for future work.

6.4 Can Small-Groups and Sole Participants Protect Their Privacy?

We also examined disclosure rates for cases when participation rates were low such as $1‰$ of V_{tar}, meaning only a few tens or hundreds of users using identity separation from $\tilde{V}_{tar}$. As seen on Fig. 12a, our experiments resulted in approximately constant disclosure rates for all models (simulations were run 15 times on each dataset).

Furthermore, we analyzed the case when only a single user uses identity separation (best model with random deletion) or a decoy identity. Within these experiments 20 different perturbed datasets were created in which only a single user applied privacy protection (15 runs). Our results are summarized on Fig. 12b. For the nodes using identity separation the disclosure rate was somewhat proportional with the number of identities used, furthermore, for the users using decoys it was constantly zero. Therefore we conclude that even if only a few users use the best model with $Y = 5$, their privacy is protected as the attacker can reveal only a few percent of sensitive information. While in general, the use of the decoy model is advised, where novel strategies need to be researched in future work.

7 Cooperative Strategies for Tackling De-anonymization

It turns out from our previous measurements that while identity separation can effectively hide information from an attacker on the individual level, it cannot effectively tackle de-anonymization on a network-wide scale: in order to achieve decreasing the recall rate significantly (e.g., between 5-10%), a large fraction of the users need to participate (around 50%, see Fig. 6).

In this section, we investigate how this can be improved by introducing cooperation between nodes. In particular, we consider globally organized cooperation between nodes. We could also investigate local cooperation, but we do not expect that such methods would achieve significantly better results compared to the non-cooperative identity separation, as the extent of the cooperation is quite small compared to network sizes.

We must note that there are even further possible strategies, depending how well the nodes (i.e., the identity partition supporting software they run) know the network structure or consider information from other sources. For instance, strategies could aim to find different cuts in order to make the structure disconnected. We leave the investigation of these strategies as future work.

7.1 Evaluation of Local Topological Anonymity in Large Networks

For realizing global cooperation, we aim using measures capturing the importance of nodes from the re-identification attack point of view. Measuring anonymity could be appropriate for doing that: the lower level of anonymity a nodes has, the easier it can be re-identified.

Local Topological Anonymity (LTA) is a measure that predicts the level of anonymity against re-identification attacks capable of achieving large-scale re-identification (such as Nar09). For the Nar09 attack, we evaluated multiple LTA metrics in [14] in smaller and mid-sized networks (up to 10k nodes), and found that some of the evaluated measures have Pearson correlation [27] around 0.4-0.5 with re-identification. However, this paper is the first work to evaluate LTA in networks that are larger of an order of magnitude.

Large-scale structural re-identification attacks compare nodes against their 2-neighborhoods in their local re-identification phase, therefore, nodes that are more similar to their neighborhood, have a lower chance of being re-identified. This property is also captured by LTA measures. In other attack algorithms node neighborhood may be inspected more deeply at the expense of larger node fingerprints and increased run-time. However, the concept of LTA can be easily adopted for these cases, where the proposed measures should be evaluated similarly.

Our previous work [14] proposes the following measures. LTA_A specifies the average similarity of a node compared to its 2-neighborhood. The measure can be used with arbitrary similarity measures, but here we used cosine similarity[3] as scoring in the Nar09 is based on that. LTA_B uses a different normalization scheme than LTA_A, i.e., the degree of the node (or at least two). LTA_C further divides LTA_A with the standard deviation of the difference in degree values between v_i and members of V_i^2. In the following, we denote V_i as the set of neighbors of a node, and V_i^2 for the neighbors within a distance of 2 (friends-of-friends). The LTA measures can be written as follows:

$$\text{LTA}_A(v_i) = \sum_{\forall v_k \in V_i^2} \frac{CosSim(v_i, v_k)}{|V_i^2|}, \tag{6}$$

[3] $CosSim(v_i, v_j) = \frac{|V_i \cap V_j|}{\sqrt{|V_i| \cdot |V_j|}}$

$$\text{LTA}_B(v_i) = \sum_{\forall v_k \in V_i^2} \frac{CosSim(v_i, v_k)}{max(|V_i|, 2)}, \tag{7}$$

$$\text{LTA}_C(v_i) = \sum_{\forall v_k \in V_i^2} \frac{CosSim(v_i, v_k)}{|V_i^2| \cdot max(\sigma_{deg}(\Delta V_i^2), 1)}. \tag{8}$$

Recent work in [16] shed light on the importance of degree regarding the Nar09 attack, and we found that the nodes with lowest LTA_A values (the strongest candidate proposed in [14]) and with top degrees have a high overlap as denoted on Fig. 13. Thus we additionally compare degree as a heuristic for predicting re-identification rates. Later, this is denoted as LTA_{deg} for keeping our notation simple.

We compared these measures as follows. We measured correlation between node reidentification rates and anonymity measures in two test sets. In the first set, we created 32 perturbations for each network we used (with 16 perturbation settings). Table 1 contains the recall rates of our measurements by running the attack with similar parameters as before, but increasing the number of simulations up to 10. In the second experiment, we run identity separation with the basic ($Y = 2$) and the best models ($Y = 5$, random edge deletion) and measured correlation values afterwards.

For the correlation measurement here we used the Spearman's rank correlation [2] instead of Pearson correlation, as it is more important to see if an LTA metric correctly orders nodes in a decreasing or increasing order according to $S(v)$, but the exact difference between rankings is not that important. An example for the ranking is shown on Fig. 16. The mean value clarifies the trend that the majority of results follow.

We must note although that the distribution of $S(v)$ is quite unique. For the majority of nodes, it is quite deterministic (regardless of the current instance of seed nodes) which nodes the algorithm can and which it cannot find (see Fig. 15). For all the experiments enlisted in Table 1, 84.63% of nodes either had $S(v) = 0$ or $S(v) = 10$, and for cases with higher recall 93.83% of all nodes fall into this category.

The results of our test cases are shown on Fig. 17a. As for our experiments both correlation values closer to 1.0 (ordered by decreasing anonymity) and to -1.0 (vice-versa) are considered to be appropriate, we displayed the absolute value of the correlations. We could do this, as for the correlation values were consequently positive or negative for a given measure.

While LTA_A and LTA_{deg} stand out as the most competitive measures on Fig. 17a, results on Fig. 17b shows a case where LTA_A is clearly better. It can be further observed that correlation values are constant-like when recall rates achieve a fair level (e.g., $5\% \leq R(\mu)$). One should also note that we omitted LTA_B from these figures as it had significantly worse correlation values compared to the others, and results were almost randomly scattered around zero correlation.

When recall rates are low, only a handful of nodes are re-identified that are connected to seeds. These can have a wide range of LTA values, and this issue should account for the randomness on the lower end of recall rates (for all measures). On the higher end, the drop in correlation is caused by the fact that the majority of nodes re-identified with all kinds of LTA values, even with high values. However, correlation values we measured are still satisfactory (starting from 0.3-0.4).

While comparing the two most competitive measures, we found that it is not the perturbation setting which seem to differ for the correlation values, but the network structure. LTA_{deg} has better results in Slashdot, Epinions, while LTA_A produced better results in LJ66k. We calculated the differences on the first test set as $|\rho_{Spearman}(S(v), LTA_A(v))| -$

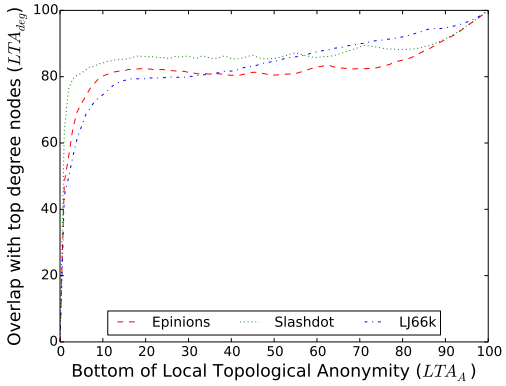


Figure 13: Overlap of top degree nodes and bottom LTA$_A$. The difference between these properties showed to be important depending on the degree distribution of the network.

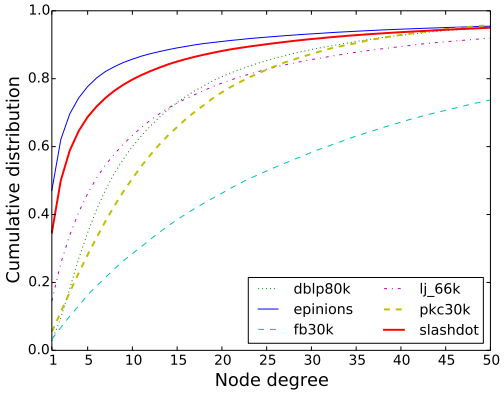


Figure 14: Networks covered a wide range of degree distribution types. This property turned out to be important from the aspect of selecting a measure of importance.

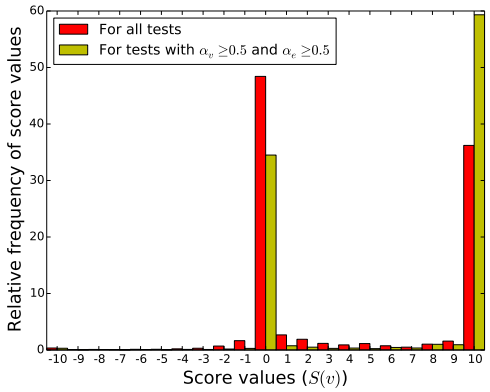


Figure 15: Distribution re-identification rates ($S(v)$) is quite unique of Nar09, as in most cases either a nodes is found or not found, but uncertain cases are uncommon.

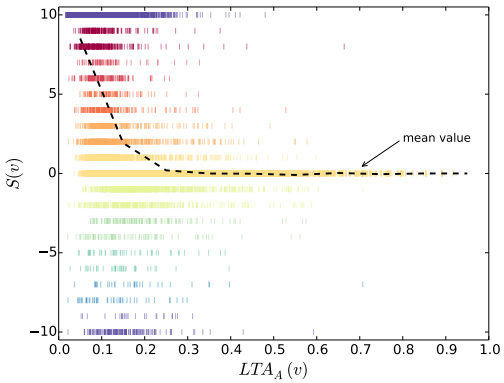
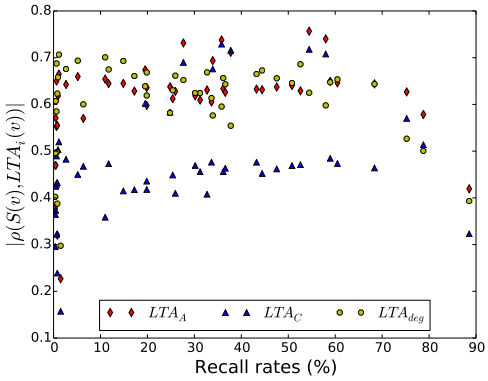
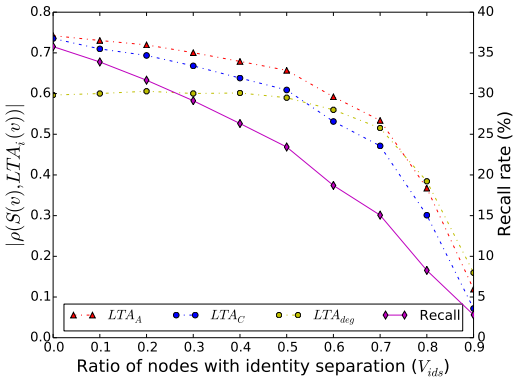


Figure 16: $S(v)$ ordered by LTA$_A$ scores (LJ66k). Beside correlation values this visualization also shows that LTA$_A$ is a promising measure of importance.



(a) Comparison of LTA variants with different perturbation settings, and their relation to recall.



(b) Correlation of LTA variants with identity separation in the LJ66k dataset (basic model, $Y = 2$).

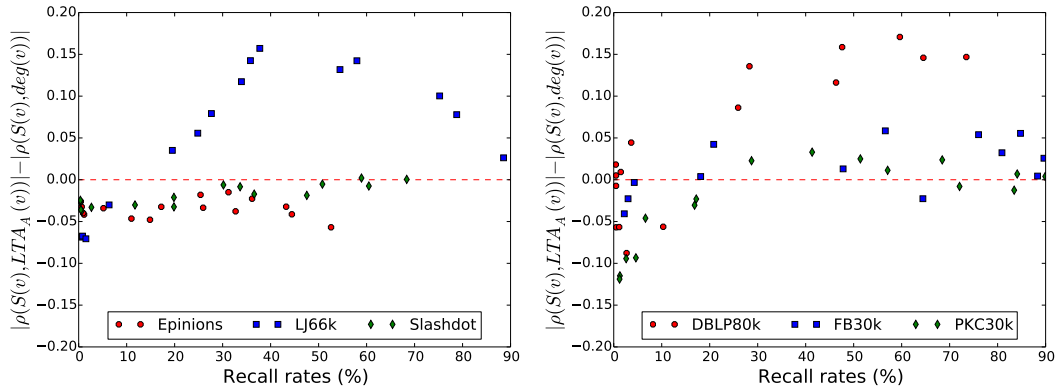
Figure 17: Results of the first set of experiments depicted on (a) with different perturbation settings (see Table 1), and results of the second set of experiments plotted on (b) with different ratio of users applying identity separation. While LTA$_A$ and LTA$_{deg}$ have the most competitive correlation values in (a), in other cases LTA$_A$ is clearly the best choice (b).

$|\rho_{Spearman}(S(v), deg(v))|$, and plotted results on Fig. 18a, now the network is indicated as well. In cases with $5\% \leq R(\mu)$ the LTA$_A$ measure is better in 13 cases with the average difference of 0.0821, while the LTA$_{deg}$ is better in 22 cases with the average difference 0.0270.

We compared the structure of the differing networks, and found that in Slashdot and Epinions the vast majority of nodes have very low degree values, e.g., the ratio of nodes with $deg(v) \leq 3$ is 58.8% and 69.6% respectively. While the degree distribution of LJ66k seems to be more balanced; the same ratio is just 33.9%. Our hypotheses was that in networks with a degree distribution similar to LJ66k, LTA$_A$ captures the difference between nodes more precisely than node degree. Therefore, for verification, we compared our results in additional networks that have similar degree distribution to LJ66k.

In order to do this, we downloaded additional datasets from the SNAP [3] and the Koblenz [1] collections. We used an export of the Pokec social network denoted as PKC30k (30,002 nodes, 245,790 edges), an export of the Facebook social network denoted as FB30k (30,002 nodes, 593,476 edges), and another from the DBLP co-author network denoted as DBLP80k (80,002 nodes, 602,096 edges). We plotted the degree distribution of all the included networks on Fig. 14, and differences of LTA$_A$ and LTA$_{deg}$ are displayed on Fig. 18b.

These measurements verify our intuition on why LTA$_A$ provided better performance in the DBLP80k, FB30k, PKC30k, LJ66k datasets. To put the difference between LTA$_A$ and LTA$_{deg}$ into another perspective, let us reinterpret measures as node fingerprints. Node degree is a first level node fingerprint with a limited information on the node neighborhood, while LTA$_A$ is a second level node fingerprint incorporating more information of the neighborhood. Thus we can expect the latter to perform better where more high degree nodes are present. Regarding our measurements presented in this section, we concluded to work with LTA$_A$ as a measure of importance in further experiments.



(a) (b)

Figure 18: Difference between the correlation values of LTA_{deg} and LTA_A. From (a) it is visible that in LJ66k LTA_A has better results than in the other networks. Comparing correlation in further networks having similar degree distribution to LJ66k is shown on (b). In these cases LTA_A proves to be a better measure for anonymity.

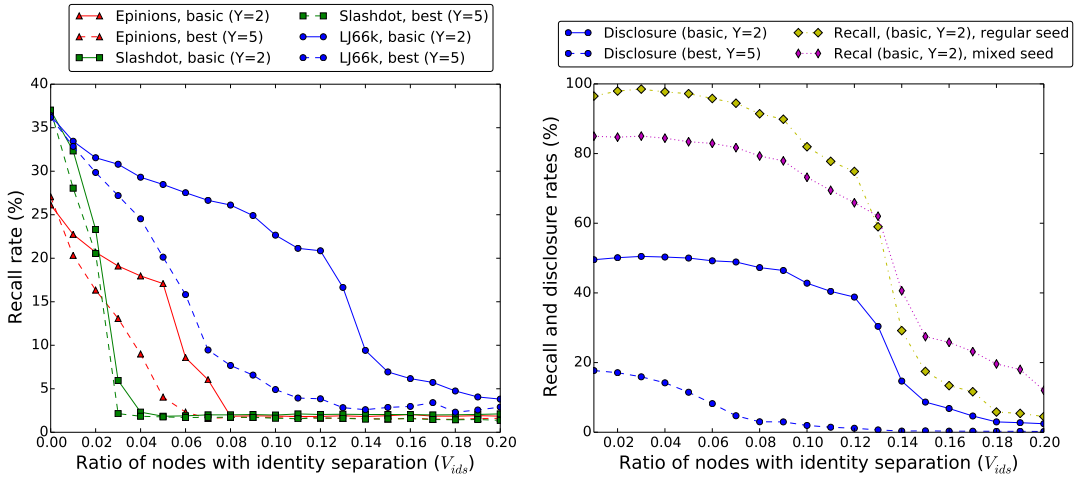
7.2 Analysis of LTA-based Global Cooperation

In this section we evaluate an LTA-based cooperative method involving nodes that have low LTA_A values, and we run simulated cooperation on the datasets we used earlier. In these measurements nodes using identity separation were not selected randomly, but the ones that had lowest LTA_A values. Thus $V_{ids} = 0.01$ means that 1% of nodes were selected to apply identity separation that had the lowest LTA scores among all nodes (this is maintained for overlapping nodes).

We applied this scheme with the basic model ($Y = 2$, uniform edge sorting) and also the best model ($Y = 5$, random edge deletion). Our results are displayed on Fig. 19a. The figure shows that in this case the attack fails even if a low number of users involved. For example, when users were selected randomly in our experiments, in the Slashdot dataset 60% of them needed to use identity separation in order to tackle the attack, while in the cooperative case only $V_{ids} = 4\%$ is enough (basic model). For the LJ66k network, this was as high as 90% in the non-cooperative case, while in the cooperative case for $V_{ids} = 15\%$ recall rates drop as $R(\mu) < 7\%$.

We displayed the disclosure rates for the LJ66k network on Fig. 19b. While disclosure rates are the quite promising here from the attacker point of view, they are less important here: the goal of the users here is to minimize recall rate. It is clearly visible from the figure that disclosure rates are highest for the bottom LTA nodes. We also included the recall rates for these nodes, that is $R(\mu) \geq 95\%$ when $V_{ids} \leq 5\%$, and remains rather high even after. This is likely because that identity separated nodes retain strong similarity with their match in the auxiliary dataset causing one of the new identities always to be found.

In addition, the seeding method also plays a significant role in this case as well (similarly as described in Section 6.1.1). For a mixed seeding method recall rates dropped for nodes using identity separation, while still staying high due to the ease of identification. For these reasons discussed above, disclosure rates stayed similarly high to this case throughout the cooperative experiments discussed in this paper.



(a) Recall rates

(b) Rates on nodes using id.sep. in LJ66k

Figure 19: After targeting users with lowest LTA_A values a significantly lower number of users is enough to tackle the attack, e.g., the recall rate in the Slashdot network drops below 5% even when only 3-4$\%$ of users participate (this is 50%-60% in the non-coordinated case). However, disclosure rates are rather high compared to the non-cooperative case: high degree nodes using identity separation are easier to be identified even despite defensive measures.

7.3 Enhancing the Seeding Method Against Cooperative Defense

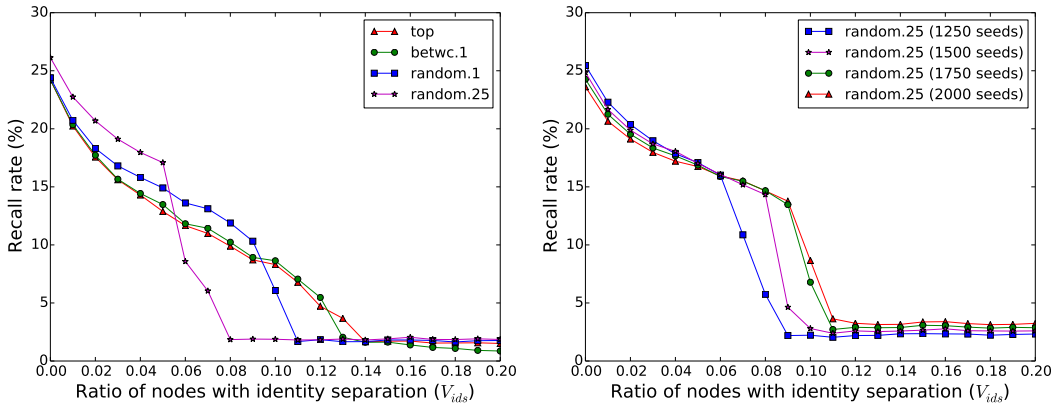
Similarly to our experiments in the non-cooperative case, we compared the `random.25` seeding method to others in order to measure their robustness against identity separation. Here, we also used the `betwc.1`, `random.1` and `top` for comparison (based on the results of [16]) with a seed set size of a thousand nodes. Additionally, we tested `random.25` with larger seed set sizes of $1250, 1500, 1750$ and 2000.

We highlighted examples of the results of these experiments on Fig. 20. Similar behavior were observed in other cases. Results indicated on the figures clearly show that the attacker has only a little control over the overall results: neither using different seeding methods (Fig. 20a), nor increasing the seed set size could improve recall rates significantly (Fig. 20b).

The drop in the recall rate of `random.25` after $V_{ids} = 0.05$ on Fig. 20a is also caused by the sensitivity to the number of seeds, which is further detailed with an example on Fig. 21. In this cooperative case the seed stability also depends on the number of seeds, similarly to the case of 200 seeds on Fig. 10b: with a higher number of seeds, large-scale propagation can be achieved with a higher probability for each perturbation settings (i.e., different values of V_{ids}). Thus, we conclude that this seed size dependency causes this minor difference between the results we measured.

7.4 Partial Participation in Global Cooperation

Cases we analyzed until this point are based on the assumption that all users cooperate to tackle the attack. However, in a real life scenario it is likely that only a subset of the



(a) Using different seeding methods

(b) `Random.25` with increased seed set

Figure 20: The attacker has little control over results. Neither using other seeding methods (a), nor increasing the seed set size (b) can significantly improve recall rates. (examples are from the Epinions dataset)

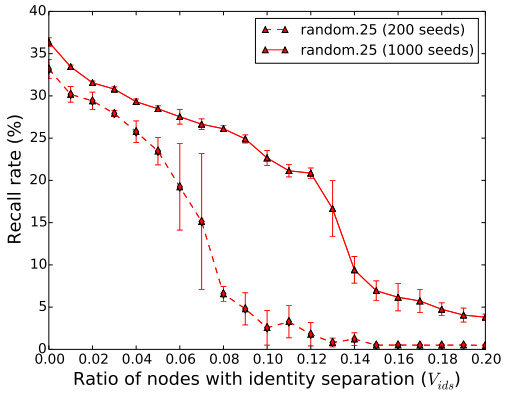


Figure 21: Recall dependency demonstrated on seed set size (LJ66k). Redundancy of seeding matters in case of cooperative identity separation as well.

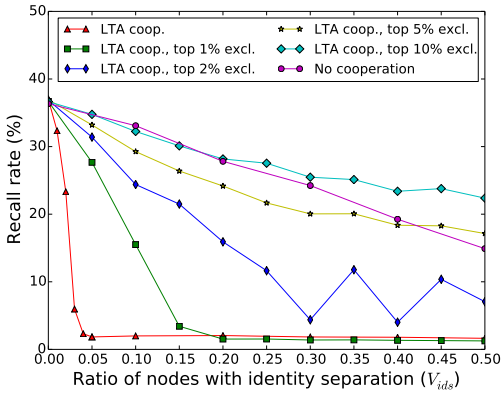


Figure 22: It significantly affects the success of defense when top degree users not participate (Slashdot). Success of defense proportionally decreases to the amount of top users omitting participation.

selected users would participate. Furthermore, the high degree nodes are the ones that are more likely to omit cooperation, e.g., because such users do not want to divide their audience. On the contrary, we expect that these users to use less visible solutions, such as decoys to hide their more privacy-sensitive activities.

Thus we analyzed how it affects the overall results if a given percent of the top degree nodes do not cooperate with others. Our results for the Slashdot network are shown on Fig. 22. Compared to when all users participate, it turns out that even if only 1% of top degree users omits cooperation a significantly larger ratio of users need to be involved for successfully tackling the attack (4% and 15% respectively).

Comparing the results of non-cooperative behavior to the cases when a large fraction of top nodes avoid cooperation, the importance of top degree nodes becomes clear. When 5%, 10% of top degree users are excluded from identity separation, overall results get similar to the non-cooperative case. This finding leads to a notable conclusion: identity separation cannot tackle structural re-identification successfully but only if top degree nodes cooperate. From the user point of view this means that the best strategy to seek personal privacy protection (e.g., using decoys) if high degree nodes are not accountable for participation.

8 Conclusion and Future Work

In this paper, we analyzed how identity separation can be used to tackle re-identification in social networks. Based on probabilistic models, we analyzed this problem from several aspects, and showed that if identity separation is applied in a non-cooperative manner, a very large fraction of users (around 50%) need to participate to repel re-identification of the whole network. Therefore we also investigated a cooperative identity separation model in which participants are selected accordingly to their level of anonymity.

We showed that node degree, and an anonymity measure we previously introduced in [14] (called LTA) are both good measures for predicting which nodes are easy preys for de-anonymization. After a comprehensive comparison, we selected LTA for the evaluation of cooperation, selecting nodes with low anonymity values for identity separation. Eventually this lead to a decrease in the required participation rates as low as $3 - 15\%$. We have also showed that it is hard for the attacker to increase this rate by using different seeding methods, but it is also crucial that top degree nodes do not omit participation.

We provided a simple strategy for hiding sensitive information in hidden identities by using public identities as decoys. We have shown that given the Nar09 attack the discovery probability of hidden identities regarding this strategy is negligible. However, it is possible for the attacker to adapt, thus more sophisticated strategies need to be provided and analyzed, which we leave for future work. Both these findings lead us to the conclusion that identity separation can be an effective tool for tackling re-identification; however, software support is needed for consequently following the strategies proposed in this paper. Elaboration of requirements and design are also left for future work.

We see further interesting research issues for future work. In this paper we have selected LTA as the best measure of importance, but the analysis of degree-based global cooperation would be also important. It would also worth checking if Nar09 could find multiple partial identities given the setting used in this paper (i.e., when a non-separated identity is in the auxiliary network). As Nar09 could only find one mapping at a time, we need a simple circumvention: when checking if a partial identity could be found or not, we remove all the others to ensure the possibility of success. After running this modified algorithm on 100 nodes (basic model, $Y = 2$, LJ66k dataset), we found that both identities could be

revealed for only 11 of them. In future work, it needs to be investigated if this ratio could be increased with new algorithms.

Furthermore, the question also rises what happens if an attacker obtains background knowledge that contains identity separated users. Regarding this case, as the identity separation process is assumed to be done secretly, the attacker could use this background knowledge to reveal hidden attributes in the identity separated anonymous network; however, this information could not be linked to the real identity of the user. Here the future work should focus on finding appropriate strategies for using identity separation in order to prevent such information leaks.

Acknowledgements

The authors would like to thank Tamás Holczer for reviewing draft versions of this paper, and for the fruitful discussions.

References

[1] The koblenz network collection. http://konect.uni-koblenz.de/. Accessed: 2014-04-28.

[2] Spearman's rank correlation. http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient. Accessed: 2014-04-22.

[3] Stanford network analysis platform (snap). http://snap.stanford.edu/. Accessed: 2014-04-22.

[4] What nsa's prism means for social media users. http://www.techrepublic.com/blog/tech-decision-maker/what-nsas-prism-means-for-social-media-users/. Accessed: 2014-05-26.

[5] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 181–190, New York, NY, USA, 2007. ACM.

[6] Sergey Bartunov, Anton Korshunov, Seung-Taek Park, Wonho Ryu, and Hyungdong Lee. Joint link-attribute user identity resolution in online social networks. In *Proceedings of the sixth Workshop on Social Network Mining and Analysis*, 2012.

[7] Filipe Beato, Mauro Conti, and Bart Preneel. Friend in the middle (fim): Tackling deanonymization in social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pages 279–284, 2013.

[8] Filipe Beato, Markulf Kohlweiss, and Karel Wouters. Scramble! your social network data. In Simone Fischer-Hbner and Nicholas Hopper, editors, *Privacy Enhancing Technologies*, volume 6794 of *Lecture Notes in Computer Science*, pages 211–225. Springer Berlin Heidelberg, 2011.

[9] Danqi Chen, Botao Hu, and Shuo Xie. De-anonymizing social networks, 2012.

[10] Sebastian Clauß, Dogan Kesdogan, and Tobias Kölsch. Privacy enhancing identity management: protection against re-identification and profiling. In *Proceedings of the 2005 workshop on Digital identity management*, DIM '05, pages 84–93, New York, NY, USA, 2005. ACM.

[11] Leucio Antonio Cutillo, Refik Molva, and Thorsten Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *Communications Magazine, IEEE*, 47(12):94–101, 2009.

[12] Oana Goga, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. Exploiting innocuous activity for correlating users across sites. In *Proceed-*

ings of the 22Nd International Conference on World Wide Web, WWW '13, pages 447–458, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[13] Gábor Gy. Gulyás and Sándor Imre. Analysis of identity separation against a passive clique-based de-anonymization attack. *Infocommunications Journal*, 4(3):11–20, December 2011.

[14] Gábor Gy. Gulyás and Sándor Imre. Measuring local topological anonymity in social networks. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 563–570, 2012.

[15] Gábor Gy. Gulyás and Sándor Imre. Hiding information in social networks from de-anonymization attacks by using identity separation. In Bart Decker, Jana Dittmann, Christian Kraetzer, and Claus Vielhauer, editors, *Communications and Multimedia Security*, volume 8099 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013.

[16] Gábor Gy. Gulyás and Sándor Imre. Measuring importance of seeding for structural de-anonymization attacks in social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, 2014.

[17] Gábor Gy. Gulyás, Róbert Schulcz, and Sándor Imre. Modeling role-based privacy in social networking services. In *Emerging Security Information, Systems and Technologies, 2009. SECURWARE '09. Third International Conference on*, pages 173–178, June 2009.

[18] Marit Hansen, Peter Berlich, Jan Camenisch, Sebastian Clauß, Andreas Pfitzmann, and Michael Waidner. Privacy-enhancing identity management. *Information Security Technical Report*, 9(1):35–44, 2004.

[19] Paridhi Jain, Ponnurangam Kumaraguru, and Anupam Joshi. @i seek 'fb.me': identifying users across multiple online social networks. In *Proceedings of the 22nd international conference on World Wide Web companion*, WWW '13 Companion, pages 1259–1268, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[20] Arvind Narayanan, Elaine Shi, and Benjamin I. P. Rubinstein. Link prediction by de-anonymization: How we won the kaggle social network challenge. In *The 2011 International Joint Conference on Neural Networks*, pages 1825–1834, 2011.

[21] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 173–187, 2009.

[22] Pedram Pedarsani, Daniel R. Figueiredo, and Matthias Grossglauser. A bayesian method for matching two similar graphs without seeds. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pages 1598–1607, Oct 2013.

[23] Wei Peng, Feng Li, Xukai Zou, and Jie Wu. Seed and grow: An attack against anonymized social networks. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012 9th Annual IEEE Communications Society Conference on*, pages 587–595, 2012.

[24] Mudhakar Srivatsa and Mike Hicks. Deanonymizing mobility traces: using social network as a side-channel. In *Proceedings of the 2012 ACM conference on Computer and communications security*, CCS '12, pages 628–637, New York, NY, USA, 2012. ACM.

[25] Bibi van den Berg and Ronald Leenes. Audience segregation in social network sites. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 1111–1116. IEEE, 2010.

[26] Bibi van den Berg and Ronald Leenes. Keeping up appearances: Audience segregation in social network sites. In Serge Gutwirth, Yves Poullet, Paul De Hert, and Ronald Leenes, editors, *Computers, Privacy and Data Protection: an Element of Choice*, pages 211–231. Springer Netherlands, 2011.

[27] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.