Chapter 1
Links Reconstruction Attack

Using Link Prediction Algorithms to Compromise Social Networks Privacy

Michael Fire, Gilad Katz, Lior Rokach, and Yuval Elovici[1]

Abstract The explosion in the use of social networks has also created new kinds of security and privacy threats. Many users are unaware of the risks involved with exposing their personal information, which makes social networks a “bonanza” for identity thieves. In addition, it has already been proven that even concealing all personal data might not be sufficient for providing protection, as personal information can be inferred by analyzing a person’s connections to other users. In attempts to cope with these risks, some users hide parts of their social connections to other users. In this paper we present “link reconstruction attack”, a method that can infer a user’s connections to others with high accuracy. This attack can be used to detect connections that a user wanted to hide in order to preserve his privacy. We show that concealing one’s links is ineffective if not done by others in the network. We also provide an analysis of the performances of various machine learning algorithms for link prediction inside small communities.

Key words: Social Networks, Social Networks Privacy, Social Networks Analysis, Inference Attack, Link Prediction, Community Link Prediction

1.1 Introduction

In recent years there has been a surge in the use of social networks, smartphones, and other internet-enabled devices. Because of this trend, ever-growing amounts of data - both personal and financial - are available online. These data can be and are being collected by third parties. Companies can collect users’ data by various methods, including social web crawlers [1], website logs [2], social network applications [3], and smart phone applications [4].

[1] Deutsche Telekom Laboratories at Ben-Gurion University of the Negev, Department of Information Systems Engineering, Ben Gurion University, Be’er Sheva, 84105, Israel
Email: {mickyfi,katzgila,liorrk,elovici}@bgu.ac.il

In today's technological world, the loss of personal data is not only a financial risk, but can also lead to criminal charges. This problematic situation in which sensitive personal information is exposed to third parties has become worse in recent years due to the explosion in use of online social networks. The amount of personal information contained in such networks is enormous. For example, an analysis of the Facebook social network determined that it had more than 845 million registered users. According to recent statistics published by Facebook [5], 50% of Facebook users log onto this site on a daily basis via laptop or other mobile devices, and 30 billion pieces of content are shared each month (web links, news stories, blog posts, notes, photo albums, etc.). The average Facebook user has 130 friends and creates 90 pieces of content each month. Many Facebook users expose personal details, such as dates of birth, email addresses, high school names, and even their phone numbers [6, 7].

Usually, online social networks like Facebook provide their users with means to protect their personal information. This is done by allowing only one's "friends" (those who the user trusts and defines as such) to access one's personal information. However, limiting access to specific trust groups is not a perfect solution for privacy protection due to the fact that some users tend to accept unfamiliar users into their trust group. In so doing, they expose their personal data to third parties [7, 8]. Furthermore, even if a person takes almost every precaution and reveals nothing except links to other users in the social network, her personal data can still be inferred from her friends. This holds true for different types of social networks, including online social networks [9], mobile phone social networks [10], and real world student cooperation social networks [2]. Therefore, as suggested by Jianming et al. [12] and Lindamood et al. [9], in order to better protect their privacy, users should also conceal their links to other users, or at least make them accessible only to their "friends".

In this paper, we present a method for inferring hidden links within small communities that are part of large social networks. Our method is based on the link prediction algorithm that was first described by Fire et al. [13]. This new algorithm is based on a machine learning classifier trained on a small set of easy-to-compute topological features. We then use the classifier to predict hidden links inside different types of social network communities, each containing up to several hundred links: a Facebook group of people who work in the same company; an SMS social network from the Friends and Family study; the real world Students' Cooperation Network; and groups of researchers with the same affiliation that were collected from the Academia.edu social network. We demonstrate that, although our classifiers were trained only on small training sets, they can still infer hidden links within different types of communities with high rates of F-measure and AUC (Area under the ROC curve). Using our methods, it is possible, with a high degree of accuracy, to infer and reconstruct users' social links and personal information.

The remainder of this paper is organized as follows. In section 1.2, we provide a brief overview of previous studies on privacy protection in social networks and on different link prediction algorithms. In section 1.3, we describe the methods and experiments that were used during the construction and evaluation of our classifiers. In section 1.4, we describe the different social network communities that were used throughout this study. In section 1.5 we present our experimental results. Finally, in section 1.6, we present our conclusions and offer future research directions.

1.2 Related Work

In this section we describe previous work in the fields of social networks privacy and link prediction.

1.2.1 Privacy in Social Networks

In recent years, online social networks use has grown exponentially. Online social networks, such as Facebook [14], Twitter [15], LinkedIn [16], Flickr [17], YouTube [18], and LiveJournal [19], serve millions of users on a daily basis. With this increased use, new privacy concerns have been raised. These concerns results from the fact that online social network users publish personal information both about themselves and their friends; all of this information can be collected by a third party. Research by Acquisti and Gross [6] in the area of social network privacy attempted to evaluate the amount of personal information that was exposed by users on Facebook. They concluded that many Facebook users disclose personal information about themselves, including dates of birth, email addresses, relationship statuses, and even phone numbers.

Another interesting fact is that around 55% of users accept friend requests from people they do not know. By accepting these friend requests, users disclose their private information to strangers [8]. Moreover, studies on trust levels in social networks showed that 27.5% of Facebook users who participated in the study had met face-to-face with people who they had initially met through Facebook [20]. Recently, Boshmaf et al. [7] collected 250 GB of inbound traffic from Facebook using Socialbots. These Socialbots succeeded in harvesting data from Facebook users by using friend requests that originated from fake Facebook profiles. Another privacy related concern is that one's personal information can be inferred from one's links. Jianming et al. [12] and Lindamood et al. [9] demonstrated methods for inferring users' personal information by using that of their friends. This was done for social networks like Facebook and LiveJournal.

Similar privacy problems also exist in other types of social networks. In smartphones social networks, various applications were identified as collecting users' personal information. This personal information included data on such things as one's location [21] and user keystrokes [22]. An even greater threat was described by Altshuler et al. [11], who showed that attacks could steal one's social network and behavioral information. Moreover, Altshuler et al. demonstrated that other information, such as ethnicity, religion, origin and age could be inferred from social network links that were created through SMS messages [10]. Recently, Fire et al. showed that even complex attributes like academic course final test grades could be inferred from a student's links to other people who took the course. The social network that was used for this analysis was created by analyzing the course's assignments and the course web log [2].

1.2.2 Link Prediction

The link prediction problem (i.e., inferring the existence of unknown links based on known ones), has many applications outside the domain of social networks. In the bioinformatics domain, link prediction is used to identify interactions among proteins [23], while in the e-commerce domain it is used to provide recommendations to customers [24]. It is even applied in the area of homeland security, where its application attempts to detect terrorist cells [25]. The popularity of this method has generated a wide variety of possible solutions. However, in spite of their diversity, most algorithms rely on supervised machine learning and feature selection. A thorough review of previous work can be found in Al Hasan and Zaki n [26].

In this paper, we focus on the common approach of using supervised learning algorithms to solve the Link Prediction problem. This approach was introduced by Liben-Nowell and Kleinberg in 2003 [27]. They studied the utility of graph topological features by testing them on five co-authorship networks data sets, each containing several thousands of authors. In 2006, Al Hasan et al. [25] extended their work on the DBLP and BIOBASE coauthorship networks (each containing several hundreds of thousands of papers). Since its publication, the supervised learning approach has been implemented by several research groups [28, 29, 30]. Most solutions that these researchers proposed were tested on bibliographic or co-authorship data sets [25, 27, 28]. In 2009, Song et al. used matrix factorization to estimate the similarity of nodes in large scale social networks, such as Facebook and MySpace [31]. In 2011, several papers were published after the IJCNN social network challenge was issued [33]. Each of these papers proposed a different method for predicting links in social networks. Narayanan et al. won the challenge by using a method that combined machine-learning algorithms and de-anonymization [34]. Cukierski et al. [35] won second place by extracting 94 distinct graph features and using the Random Forest algorithm to analyze the training data (consisting of several thousands of edges). Recently, Fire et al. published a method for predicting links in large scale online social networks using easy-to-compute topological features. Their method used 50,000 links as a training set for the classifiers [13].

In this paper, we use the link prediction algorithm presented by Fire et al. [13] and test it on different types of social network communities in order to reconstruct users' links. We show that reconstructing users' links can compromise their privacy and render them vulnerable to inference attacks, as described by Jianming et al. [12] and Lindamood et al. [9].

1.3 Methods and Experiments

To identify hidden links inside different communities, we applied methods from the machine learning domain. For each community, we developed a dedicated link classifier capable of predicting the likelihood of the existence of a link between two members. For each community, we extracted a "positive" training set of links that exist in the communities' graphs and a set of "negative" links that do not exist in the graph.

Due to the small sizes of the communities, our "positive" links training set consisted almost entirely of links that connected members inside the community. Our "negative" links training set consisted of two types of links. The first was random links, where both nodes were chosen randomly (hereafter referred to as the "easy" train set). The second type of "negative" links was generated so that the two connected nodes were within a distance of two from each other (hereafter referred to as the "hard" train set).

Subsequently, for each positive and negative link we extracted a small set of easy-to-compute topological features, as suggested by Fire et al. [13]. We then used these extracted topological features to train several supervised learning classifiers. This was done using WEKA [32], a popular suite of machine learning. Finally, we used WEKA to evaluate the performance of each classifier.

The remainder of this section describes the small sets of features that were extracted to train our classifiers and the different machine learning algorithms used in our experiments.

1.3.1 Feature Extraction

This section describes the different features that were extracted in order to build our community link prediction classifiers. The extracted features are primarily based on the Friends-features subset, as suggested by Fire et al. [13].

Let $G =< V, E >$ be the graph representing the topological structure of a general social network community. Links in the graph are denoted by $e = (u, v) \in E$ where $u, v \in V$ are nodes in the community graph. Our goal is to construct simple classifiers capable of computing the likelihood of $(u, v) \in E$ or $(u, v) \notin E$ for every two nodes $u, v \in V$. To achieve this goal, we extracted the following features for each link, (u, v), from our training sets.

1. **Vertex degree:** Let be $v \in V$, we can define the neighborhood of v by:
$$\Gamma(v) := \{u|(u, v) \in E \text{ or } (v, u) \in E\} \tag{1.1}$$
If G is a directed graph we can also define the following neighborhoods:
$$\begin{aligned} \Gamma_{in}(v) &:= \{u|(u, v) \in E\} \\ \Gamma_{out}(v) &:= \{u|(v, u) \in E\} \end{aligned} \tag{1.2}$$
Using the above defined neighborhoods, we can define the following degree feature:
$$degree(v) := |\Gamma(v)| \tag{1.3}$$
If G is a directed graph, we can also define the following degree features:
$$degree_{in}(v) := |\Gamma_{in}(v)| \tag{1.4}$$
$$degree_{out}(v) := |\Gamma_{out}(v)| \tag{1.5}$$
The degree features are used to measure the number of friends v has inside the community. If we look at a directed graph of a community such as Twitter, the meaning of the degree feature is how many other members of the community v follows (out-degree), and how many members of the community follow v (in-degree).
2. **Common Friends**: Let $u, v \in V$; we define the common friends of u and v to be all the members in the community that are friends both of u and v. The formal definition of the number of common friends is: Let be $u, v \in V$ then
$$common\text{-}friends(u, v) := |\Gamma(v) \cap \Gamma(u)| \tag{1.6}$$
The common-friends feature was widely used in previous work in to predict links in different datasets [13, 24, 27, 29, 31, 35].
3. **Total Friends**: Let $u, v \in V$; we can define the number of distinct friends of u and v by:
$$total\text{-}friends(u, v) := |\Gamma(u) \cup \Gamma(v)| \tag{1.7}$$
4. **Jaccard's coefficient**: Jaccard's-coefficient is a well-known feature for link prediction [13, 24, 27, 29, 31, 35]. This feature, which measures the

similarity among sets of nodes, is defined as the size of the intersection divided by the size of the union of the sample sets. The formal definition of Jaccard's coefficient can be written in the following way.

$$Jaccard's\text{-}coefficient(u, v) := \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|} \tag{1.8}$$

In our approach, this measure indicates whether two community members have a significant number of common friends regardless of their total number of friends. A higher value of Jaccard's-coefficient indicates a stronger connection between two nodes in the community.

5. **Preferential-attachment-score**: The preferential-attachment score is defined as the multiplication of the number of friends of u and v.

$$preferential\text{-}attachment\text{-}score(u, v) := |\Gamma(u)| \cdot |\Gamma(v)| \tag{1.9}$$

The Preferential-attachment score is a well-known concept in social networks. It measures how "connected" each user is and also provides a strong indication of how likely (and at what rate) the user is likely to create additional connections [13, 25].

6. **Opposite direction friends**: For a directed graph G, we created a specific measure that indicates whether reciprocal connections exist between each pair of nodes

$$opposite\text{-}direction\text{-}friends(u, v) := \begin{cases} 1 & \text{if } (v, u) \in E \\ 0 & \text{otherwise} \end{cases} \tag{1.10}$$

7. **Shortest path**: We define the shortest path measure between nodes u and v in the following manner: $shortest - path(u, v)$. This measure represents the length of the shortest path between u and v inside the community. If the community graph is directed, this measure will not necessarily be symmetrical. The shortest path feature has been explored in several papers [13, 25] and found to be one of most significant features for the predicting hidden links.
8. **Friends Measure**: The friends measure is a private case of the Katz measure [36], and was first presented by Fire et al. [13]. The formal definition of the friends measure is: Let be $u, v \in V$, then

$$friends - measure(u, v) := \sum_{x \in \Gamma(u)} \sum_{y \in \Gamma(v)} \delta(x, y) \tag{1.11}$$

Where $\delta(x, y)$ is defined as:

$$\delta(x, y) := \begin{cases} 1 & \text{if } x = y \text{ or } (x, y) \in E \text{ or } (y, x) \in E \\ 0 & \text{otherwise} \end{cases}$$

The friends measure represents the number of u's friends who also know v's friends. The higher the number of connections between u and v's friends, the greater the chance that u and v know each other.

1.3.2 Experimental Setup

To build community link prediction tools, we created an easy training set and a hard training set for each community. The easy training set for each community contained 50% positive and 50% negative links. As mentioned previously, the positive links are those that exist within the community, while the negative links are those that, to the best of our knowledge, do not exist.

In the easy training set, each of the negative links was created by randomly choosing two nodes in the community that did not have a link between them, while in the hard training set, negative links were created by choosing two nodes in the community that were at a distance of two from each other. Due to the small size of each community, the size of each training set would have been twice the size of the number of existing links in each community had all of the existing links in each community been included in our training sets.

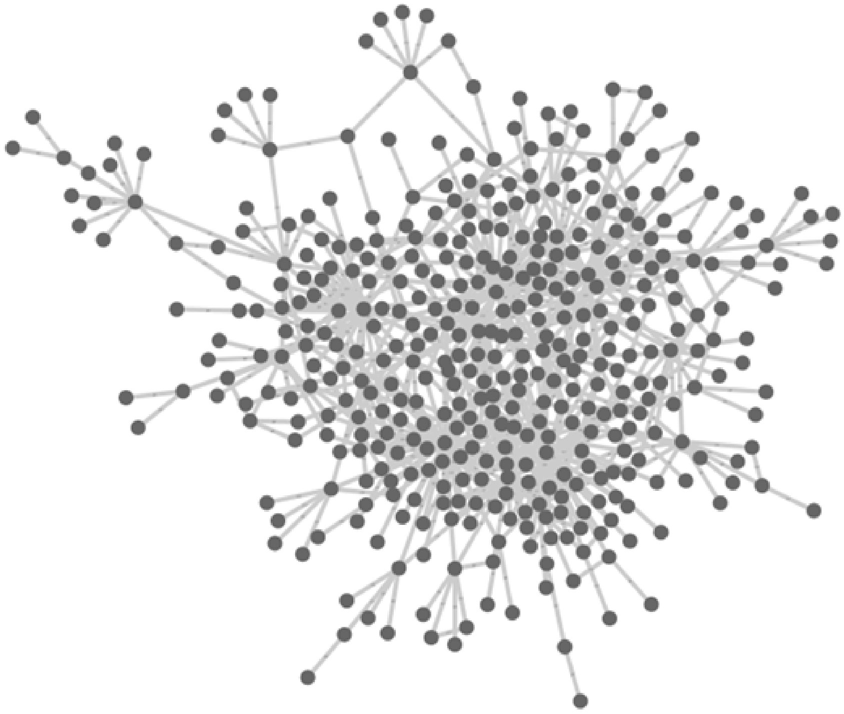
Once the training set links were selected, a Python code was developed using the Networkx package [37]. This code was used to extract the topological features mentioned above for each of the links (8 features for an undirected network and 14 features for a directed network). Our next step was to evaluate different link prediction methods created by different supervised learning algorithms. We used WEKA's C4.5 (J48), IBk, NaiveBayes, SMO, MultilayerPerceptron, Bagging, AdaBoostM1, RotationForest, and RandomForest implementations of the corresponding algorithms. For each of these algorithms, most of the configurable parameters were set to their default values, with the following exceptions: for C4.5, the minimum number of instances per leaf parameter was between the values of 2, 6, 8 and 10; for IBk, its k parameter was set to 10; for SMO, the NormalizedPolyKernel with its default parameters was used. The ensemble methods were configured as follows: The number of iterations for all ensemble methods was set to 100. The Bagging, AdaBoostM1, and RotationForest algorithms were evaluated using J48 as the base classifier with the number of instances per leaf set to 4, 6, 8, and 10.

1.4 Communities Datasets

We evaluate our community link prediction classifiers using four communities from different types of social networks datasets: Facebook [14], Academia.edu [38], Friends and Family study [4], and Students' Cooperation Social Network [2].

Facebook. Facebook is a website and social networking service that was launched in February 2004 [14]. As of January 2012, Facebook had more than 800 million registered users [5]. Facebook users may create a personal profile, add other users as friends, and interact with other members. Because the friendship link between two members must be reciprocal, the existence of a link between member A and member B induces a mutual connection. Therefore, we refer to Facebook's underlying friendship graph undirected. We evaluated our classifiers for a small community of co-workers that according to their Facebook profile pages worked for the same well-known high-tech company. These co-workers' community network graph contained 410 nodes and 635 links, it was obtained using a web crawler at the beginning of January 2012 (see Figure 1.4[2]).

Fig. 1.1 Facebook coworker community social network

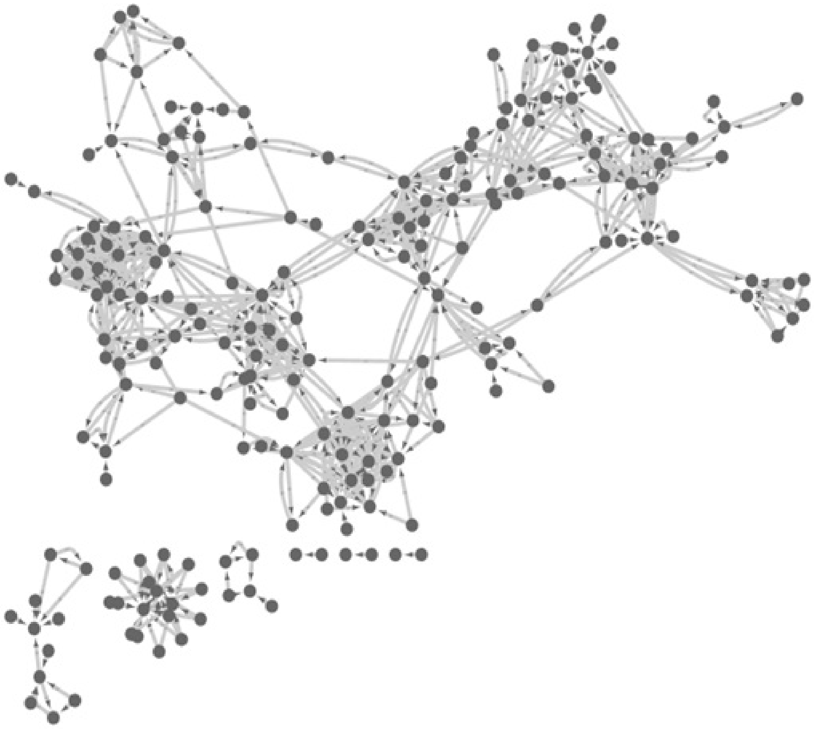


Academia.edu. Academia.edu is a platform for academics to share and follow research that is underway in a particular field or discipline [38]. Members upload and share their papers with other researchers in over 100,000

[2] All the social networks figures in this paper where created by Cytoscape software [39]

fields and categories. An Academia social network member may choose to follow any of the network's members; hence, the directed nature of the links within this network. We evaluated our classifiers for a small community of researchers who, according to their Academia.edu profiles, belonged to the same Ivy League University. The researchers' community network graph contained 207 nodes and 702 links (see Figure 1.4) and was obtained using a web crawler.

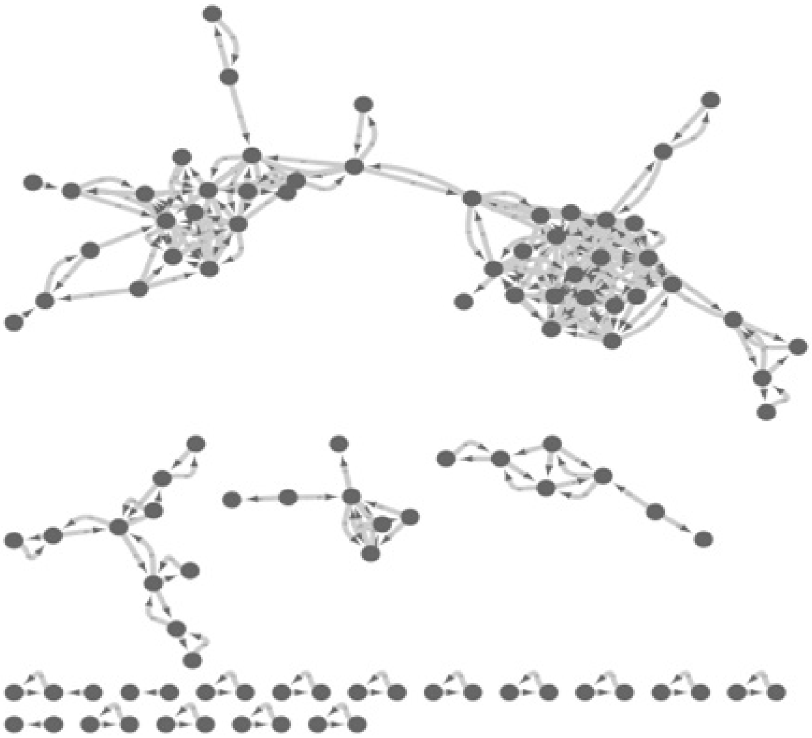
Fig. 1.2 Academia.edu researchers community social network



Friends and Family. The Friends and Family dataset contains rich data signals gathered from the smartphones of 140 adult members of a young-family residential community. The data were collected over the course of one year [4]. We evaluated our classifiers for members of a social network that was constructed based on SMS messages sent and received by the members. The SMS messages social network directed graph contained 103 nodes and 281 links (see Figure 1.4).

Students' Cooperation Social Network. The students' cooperation social network was constructed from the data collected during a "Computer

Fig. 1.3 Friends and Family SMS messages social network

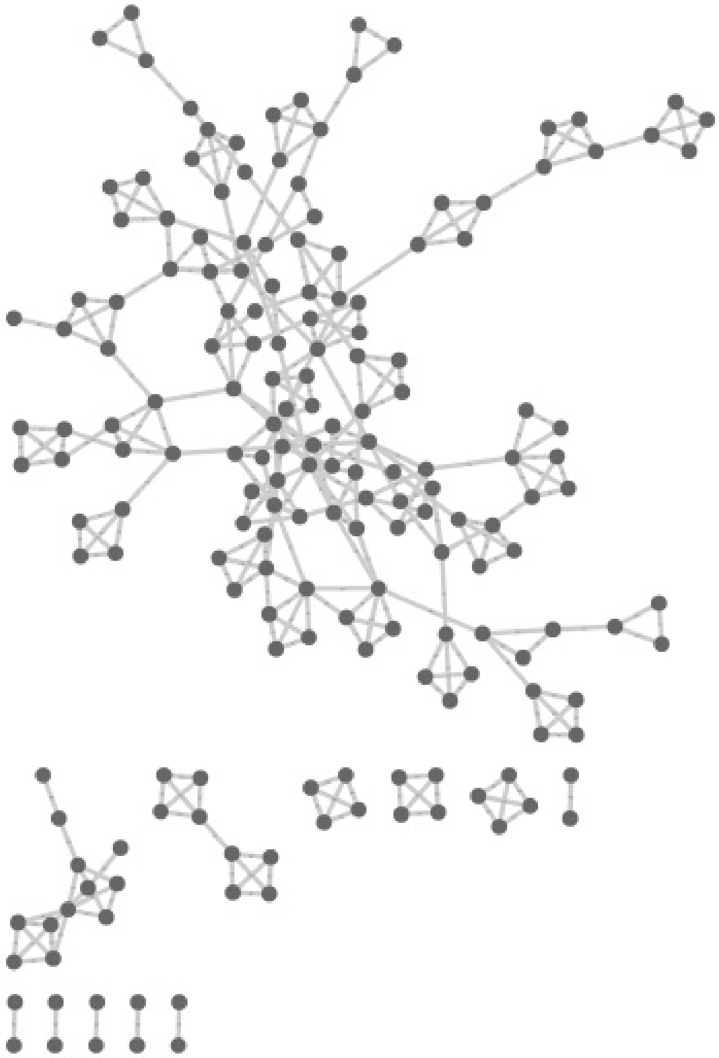


and Network Security" course; a mandatory course taught by two of this paper's authors at Ben-Gurion University [2]. The social network contains data collected from 185 participating students from two different departments. The course's social network was created by analyzing the implicit and explicit cooperation among the students while doing their homework assignments. The students' cooperation graph contained 185 nodes and 311 links (see Figure 1.4).

1.5 Results

For each community and for each easy and hard training set, we evaluated our list of different machine learning classifiers by using a 10-fold cross-validation approach. We used an area-under-curve (AUC) measure to evalu-

Fig. 1.4 Students' Cooperation Social Network



ate our results. In figures 1.5 and 1.5, we present the classifiers' performances for the easy dataset for the Facebook coworkers' community and for the Academia.edu researchers' community, respectively. The best results for each of our datasets are presented in Table 1.2 and Table 1.3.

As expected, the ensemble methods fared best, especially the Rotation Forest algorithm. In contrast to previous link prediction ensemble classifiers

Table 1.1 Communities Datasets

Community	Is Directed	Nodes Number	Links Number	Obtained by	Date
Facebook coworkers	No	410	635	Web crawler	2012
Academia.edu researchers community	Yes	207	702	Web crawler	2011
Friends and Family SMS network	Yes	103	281	Smartphone application	2010-11
Students' cooperation network	No	185	311	Web log and assignments analysis	2011

Fig. 1.5 AUC Results - Facebook coworkers' community

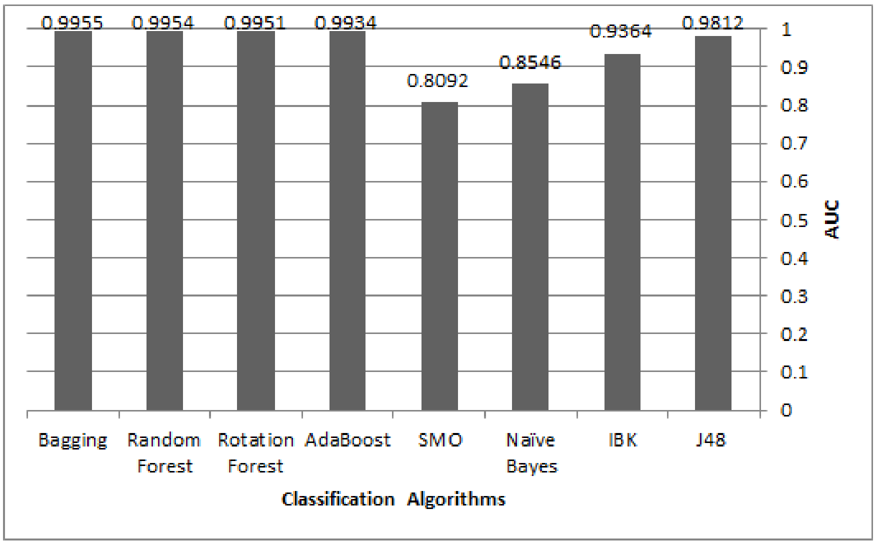
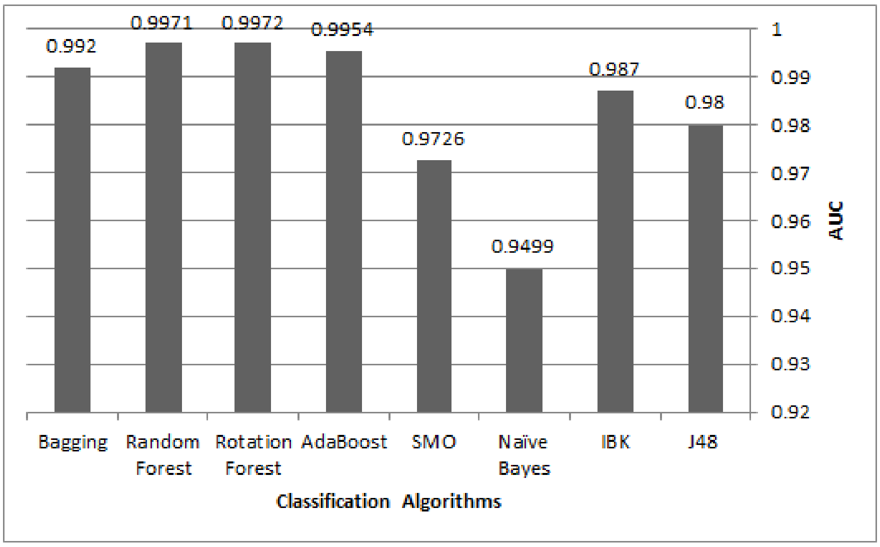


Table 1.2 Easy Training Set - Classifiers' Highest Results

Dataset	Classifier	Train set size	TP	F-Measure	AUC
Facebook	Rotation Forest	1,270	0.9983	0.9667	0.9951
Academia.edu	AdaBoost	1,445	0.969	0.9746	0.9954
Friends & Family	AdaBoost	563	0.9818	0.9745	0.9946
Student's Cooperation	Bagging	623	0.9643	0.9398	0.9775

that require large amounts of resources [13, 35], our community link prediction ensemble classifiers were quick to build and train due to the small size of the required training sets. For example, the average time for extracting link

Fig. 1.6 AUC Results - Academia.edu researchers' community



features in the Facebook community was 0.002 seconds, both for negative and positive links[3].

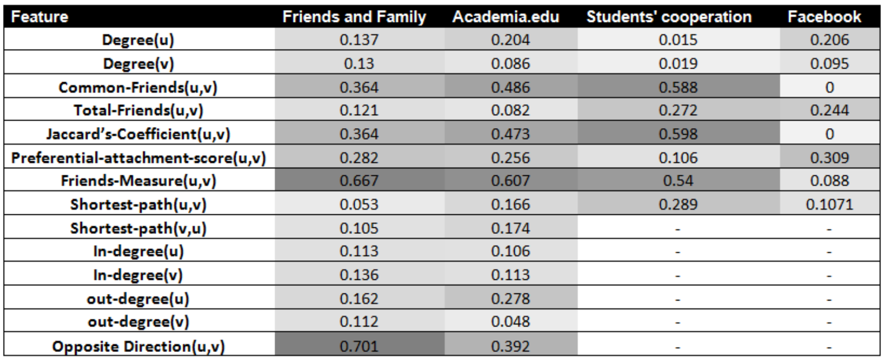
Table 1.3 Hard Training Set - Classifiers' Highest Results

Dataset	Classifier	Train set size	TP	F-Measure	AUC
Facebook	Rotation Forest	1,270	0.9835	0.9686	0.9981
Academia.edu	Rotation Forest	1,445	0.9127	0.9213	0.9756
Friends & Family	Rotation Forest	563	0.9537	0.9471	0.9831
Student's Cooperation	Rotation Forest	623	0.999	0.9883	0.9998

To obtain an indication of the usefulness of the various features in different communities and datasets, we analyzed their respective importance using Weka's information gain attribute selection algorithm. Our results are presented in Tables 1.4 and 1.5. Based on these results, it should be noted that a feature's importance varies among the different communities.

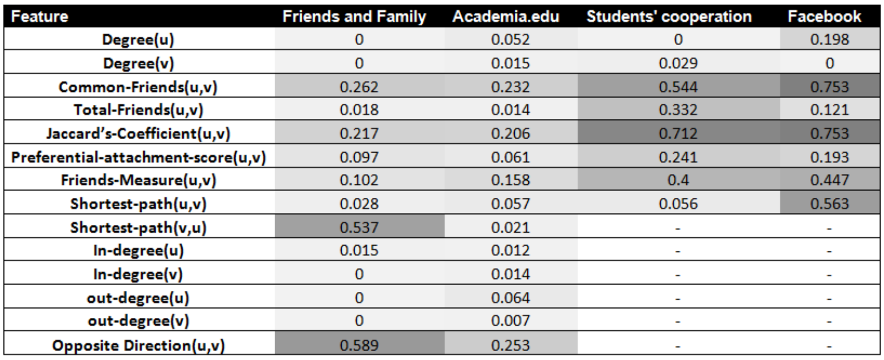
[3] We ran our algorithm using Python 2.7, on a regular Dell Latitude E6420 laptop with i7 core, and 8GB RAM

Table 1.4 Easy Training Set - Information Gain value of Different Features



Feature	Friends and Family	Academia.edu	Students' cooperation	Facebook
Degree(u)	0.137	0.204	0.015	0.206
Degree(v)	0.13	0.086	0.019	0.095
Common-Friends(u,v)	0.364	0.486	0.588	0
Total-Friends(u,v)	0.121	0.082	0.272	0.244
Jaccard's-Coefficient(u,v)	0.364	0.473	0.598	0
Preferential-attachment-score(u,v)	0.282	0.256	0.106	0.309
Friends-Measure(u,v)	0.667	0.607	0.54	0.088
Shortest-path(u,v)	0.053	0.166	0.289	0.1071
Shortest-path(v,u)	0.105	0.174	-	-
In-degree(u)	0.113	0.106	-	-
In-degree(v)	0.136	0.113	-	-
out-degree(u)	0.162	0.278	-	-
out-degree(v)	0.112	0.048	-	-
Opposite Direction(u,v)	0.701	0.392	-	-

Table 1.5 Hard Training Set - Information Gain value of Different Features



Feature	Friends and Family	Academia.edu	Students' cooperation	Facebook
Degree(u)	0	0.052	0	0.198
Degree(v)	0	0.015	0.029	0
Common-Friends(u,v)	0.262	0.232	0.544	0.753
Total-Friends(u,v)	0.018	0.014	0.332	0.121
Jaccard's-Coefficient(u,v)	0.217	0.206	0.712	0.753
Preferential-attachment-score(u,v)	0.097	0.061	0.241	0.193
Friends-Measure(u,v)	0.102	0.158	0.4	0.447
Shortest-path(u,v)	0.028	0.057	0.056	0.563
Shortest-path(v,u)	0.537	0.021	-	-
In-degree(u)	0.015	0.012	-	-
In-degree(v)	0	0.014	-	-
out-degree(u)	0	0.064	-	-
out-degree(v)	0	0.007	-	-
Opposite Direction(u,v)	0.589	0.253	-	-

1.6 Conclusions and Future Work

In today's world, many people use different types of social networks in order to communicate with each other and to share knowledge. One of the main problems with using and participating in social networks is that one's privacy can be easily become compromised. Even if the social network user does not expose information to other users in the network, and even hides all of his personal information, he may still be exposed to inference attacks due to his connections to other users [9, 12]. One can defend oneself against these types of inference attacks by hiding some of one's links in the network. In this study, we presented a method for reconstructing a user's hidden links to other users by creating a link prediction community classifier for different types of social networks.

The classifiers presented in this paper were created by using only a handful of graph topological features for each link and a small amount of training data for each data set. Despite these limitations, the tested classifiers succeeded in achieving high results both in terms of F-Measures and AUC measures (also referred to as ROC Areas) for all tested community data sets. While most of the tested classifiers produced positive results, the best results were obtained using ensemble supervised learning algorithms, with the Rotation Forest algorithm achieving the highest AUC rates. We also demonstrated that the presented classifiers could perform well even on the hard training set (See Table 1.3). These types of link prediction classifiers can assist attackers in reconstructing the hidden user links.

Several possibilities for future research are currently under consideration. The first direction is attempting to reconstruct cross-community hidden links. Another possibility is creating a method to accuratly predict and measure one's exposure to inference attacks. A third possible direction is creating a recommender system that would advise users to connect to other users in order to foil link reconstruction attacks.

References

1. Mislove A, Marcon M, Gummadi K. P., Druschel P., Bhattacharjee B. (2007), Measurement and Analysis of Online Social Networks, in Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC'07), San Diego, CA.
2. Fire, M., Katz, G., Elovici, Y., Shapira, B., Lior, R. (2011), Predicting Student Exam's Scores by Analyzing Social Network Data
3. Wang, N., Xu, H., Grossklags, J.(2011), Third-Party Apps on Facebook: Privacy and the Illusion of Control, Proceedings of the ACM Symposium on Computer-Human Interaction for Management of Information Technology (CHIMIT), Boston, MA.
4. Aharony, N., Pan, W., Ip, C., Khayal, I., Pentland, A. (2011), The Social fMRI: Measuring, Understanding and Designing Social Mechanisms in the Real World, in Proceedings of the 13th ACM international conference on Ubiquitous computing. ACM
5. Facebook statistics, https://www.facebook.com/press/info.php?statistics.
6. Acquisti, A., Gross, R. (2006), Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook, Privacy Enhancing Technologies
7. Boshmaf, Y., Muslukhov, I., Beznosov, K., Ripeanu, M.(2011),The Socialbot Network: When Bots Socialize for Fame and Money.
8. Nagle, F., Singh, L. (2009), Can Friends Be Trusted? Exploring Privacy in Online Social Networks, International Conference on Advances in Social Network Analysis and Mining, Athens.
9. Lindamood, J., Raymond, H., Kantarcioglu, M., Thuraisingham, B. (2009), Inferring private information using social network data, in Proceedings of the 18th international conference on World Wide Web
10. Altshuler, Y., Fire, M., Aharony, N., Elovici, Y., Pentland, A. (2012), How Many Makes a Crowd? On the Correlation between Groups' Size and the Accuracy of Modeling,Social Computing Behavioral Modeling and Prediction(SBP), University of Maryland
11. Altshuler, Y., Aharony, N., Pentland, A., Elovici, Y., Cebrian, M (2011), Stealing Reality: When Criminals Become Data Scientists (or Vice Versa) ,Intelligent Systems, IEEE, Vol. 26
12. Jianming, H, Wesley, W. C., Zhenyu, L. (2006), Inferring privacy information from social networks, IEEE international conference on intelligence and security informatics, San Diego, CA, USA
13. Fire, M., Tenenboim, L., Lesser ,O., Puzis R., Rokach, L, Elovici, Y. (2011), Link Prediction in Social Networks using Computationally Efficient Topological Features, SocialCom, MIT, Boston, USA
14. Facebook, http://www.facebook.com.
15. Twitter, http://www.twitter.com.

16. LinkedIn, http://www.linkedin.com/.
17. Flickr, http://www.flickr.com.
18. YouTube, http://www.youtube.com/.
19. LiveJournal, http://www.livejournal.com/.
20. Dwyer, C., Hiltz, S. R. Passerini, K. (2007), Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace, in Proceedings of the Americas Conference on Information Systems
21. Allan, A., Pete, Warden, P. (2011), Got an iphone or 3g ipad? apple is recording your moves, http://radar.oreilly.com/2011/04/apple-location-tracking.html
22. Eckhart, T., (2011), What is Carrier IQ?, http://androidsecuritytest.com/features/logs-and-services/loggers/carrieriq/
23. Airoldi, E. M. , Blei, D. M. , Fienberg, S. E.,Xing, E. P. (2006), Mixed membership stochastic block models for relational data with application to protein-protein interactions, Proceedings of Ineterational Biometric Society-ENAR Annual Meetings.
24. Huang, Z., Li, X., Chen, H. (2005), Link prediction approach to collaborative filtering, Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries.
25. Hasan M. A. , Chaoji, V., Salem, S.,Zaki M. (2006), Link prediction using supervised learning, SDM Workshop of Link Analysis, Counterterrorism and Security
26. Hasan, M. A.,Zaki, M. J. , (2011), A survey of link prediction in social networks, Social Network Data Analytics, C. C. Aggarwal, Ed. Springer.
27. Liben-Nowell, D., Kleinberg, J. (2007),The link-prediction problem for social networks, Journal of the American Society for Information Science and Technology, vol. 58, no. 7, 2007.
28. Doppa, J. R., Yu, J., Tadepalli, P. Getoor, L. (2009), Chance-constrained programs for link prediction, In Proceedings of Workshop on Analyzing Networks and Learning with Graphs at NIPS Conference.
29. Sa, H. R.,Prudencio, R. B. C. (2010), Supervised learning for link prediction in weighted networks, III International Workshop on Web and Text Intelligence.
30. Leskovec, J., Huttenlocher, D., Kleinberg, J. Predicting Positive and Negative Links in Online Social Networks (2010), WWW '10 Proceedings of the 19th international conference on World wide web.
31. Song, H. H. , Cho, T. W. , Dave, V., Zhang, Y., Qiu, L. (2009), Scalable proximity estimation and link prediction in online social networks, in Proceedings of the 9th ACM SIGCOMM conference on Internet measurement
32. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. (2009), The weka data mining software: an update, SIGKDD Explor. Newsl., vol. 11, pp. 10–18.
33. IJCNN social network challenge, http://www.kaggle.com/c/socialNetwork/Data.
34. Narayanan, A.,Shi, E. Rubinstein, B.I.P. (2011), Link prediction by de-anonymization: How We Won the Kaggle Social Network Challenge, The 2011 International Joint Conference on Neural Networks (IJCNN), San Jose, CA.
35. Cukierski, W. J. , Hamner, B., Yang, B. (2011), Graph-based features for supervised link prediction, International Joint Conference on Neural Networks (IJCNN), San Jose, CA.
36. Katz, L. (1953), A new status index derived from sociometric analysis, Psychometrika, vol. 18, no. 1, pp. 39–43, [Online]. Available: http://ideas.repec.org/a/spr/psycho/v18y1953i1p39-43.html
37. Hagberg, A. A. , Schult, D. A. , Swart, P. J. (2008), Exploring network structure, dynamics, and function using networkx," In Proceedings of the 7th Python in Science Conference (SciPy2008).
38. Academia.edu, http://academia.edu/.
39. Shannon, P. et. al. (2003),Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks, Genome research, Vol. 13, No. 11, pp. 2498-2504.