Exploit of Online Social Networks with Community-Based Graph Semi-Supervised Learning

Mingzhen Mo and Irwin King

Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
{mzmo,king}@cse.cuhk.edu.hk

Abstract. With the rapid growth of the Internet, more and more people interact with their friends in online social networks like Facebook[1]. Currently, the privacy issue of online social networks becomes a hot and dynamic research topic. Though some privacy protecting strategies are implemented, they are not stringent enough. Recently, Semi-Supervised Learning (SSL), which has the advantage of utilizing the unlabeled data to achieve better performance, attracts much attention from the web research community. By utilizing a large number of unlabeled data from websites, SSL can effectively infer hidden or sensitive information on the Internet. Furthermore, graph-based SSL is much more suitable for modeling real-world objects with graph characteristics, like online social networks. Thus, we propose a novel Community-based Graph (CG) SSL model that can be applied to exploit security issues in online social networks, then provide two consistent algorithms satisfying distinct needs. In order to evaluate the effectiveness of this model, we conduct a series of experiments on a synthetic data and two real-world data from StudiVZ[2] and Facebook. Experimental results demonstrate that our approach can more accurately and confidently predict sensitive information of online users, comparing to previous models.

Keywords: privacy issue, social network, graph-based semi-supervised learning, community consistency.

1 Introduction

Currently, online social networks are becoming increasingly popular. For example, Facebook currently is utilized by more than 400 million active users and more than 500 billion minutes are spent on it everyday [1]. In these online social networks, people can form social links with others through making friends or joining groups with similar contents.

The security issue of online social networks turns into one of the hot topics, because it affects hundreds of millions users. Online social networks allow people to enable privacy restriction on their profiles. Nevertheless, the friendship and group membership are still visible to the public directly or indirectly. In other words, the public friendship or group information, which online social networks claim to be safe, becomes the

[1] http://www.facebook.com

[2] http://www.studivz.net

K.W. Wong et al. (Eds.): ICONIP 2010, Part I, LNCS 6443, pp. 669–678, 2010.

potential threat to users' privacy. [5,6,10] demonstrate that this information can leak a large quantity of sensitive information.

Recently, Semi-Supervised Learning (SSL) has become a useful technique to exploit unknown information. Compared to supervised learning, SSL has the advantage of avoiding high cost in labeling training data by utilizing large amount of unlabeled data. Thus, SSL can be applied on predicting or learning knowledge from the websites which contain massive unlabeled data, e.g., hidden or sensitive information.

As a technique to exploit hidden information, SSL suits well with the scenario that online social networks contain little public information and a large number of hidden ones [7]. In SSL learning model, the public information can be considered as labeled data and that hidden as unlabeled data. According to the statistics, on average 70% users in Facebook have incomplete profiles. It illustrates that labeled data are far fewer than the unlabeled data.

Especially, graph-based SSL further fits well the online social networks with graph structures. First, graph-based SSL is good at modeling objects with graph structures, in which relationship information is easily expressed by edges and their weights. Second, the learning procedure of graph-based SSL is spreading known information to unknown area to predict the result. That is very similar to the cases in the real world, e.g., online social networks: we expand our networks from existing friends to unacquainted persons and from familiar groups to strange communities. Hence, graph-based SSL is rather suitable for exploiting online social networks.

This paper proposes a novel graph-based SSL model with community consistency. There are several graph-based learning models were proposed before, e.g., basic graph learning with harmonic function [12], which mainly considers the local consistency, and Local and Global Consistency (LGC) graph leaning [11]. Now, we propose a novel graph learning model considering not only local consistency and global consistency but also community consistency. The relationship between this model and the previous ones is shown in Fig. 1.

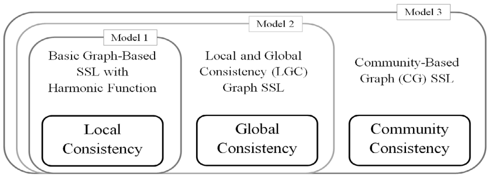

Model 1
Model 2
Model 3
Basic Graph-Based SSL with Harmonic Function
Local and Global Consistency (LGC) Graph SSL
Community-Based Graph (CG) SSL
Local Consistency
Global Consistency
Community Consistency

Fig. 1. The Relationship of Three Graph-based SSL Models

This novel SSL exploit model is evaluated on a synthetic dataset 'TwoMoons' and two real-world datasets from StudiVZ and Facebook, comparing with two previous graph-based SSL models and a Supervised Learning model. The evaluation criterion contains accuracy and weighted accuracy, which is defined to measure the confidence of predictions.

The contributions of this paper include the following:

- **A graph-based semi-supervised learning with community consistency is firstly proposed.** With the additional consistency in the objective, this learning model describes the real world more accurately and achieves better learning results.
- **This paper provides two algorithms for the Community-based Graph (CG) SSL exploit model: a closed form algorithm and an iterative algorithm.** The closed form algorithm has a very simple formula to obtain the prediction result, while the iterative algorithm could deal with large-scale datasets.

2 CG SSL Exploit Model and Algorithms

2.1 CG SSL Exploit Model

Preparation. Similar to [7], we define a social network as an undirected graph $G(V, E)$. In $G(V, E)$, every vertex (user) has feature vector $P_i = (p_i^1, p_i^2, \cdots, p_i^{n_f})$ and every edge (relationship) has weighted value $W_{i,j} = (w_{i,j}^{fd}, w_{i,j}^{gp}, w_{i,j}^{nk})$. n_f is the total number of features; $w_{i,j}^{fd}$ is a weight for friendship, $w_{i,j}^{gp}$ for group membership, $w_{i,j}^{nk}$ for network relationship and $0 \leq P_i, W_{i,j} \leq 1$. In the whole graph, there are l vertices labeled as $\bar{Y}_{label}$ and u vertices needed to predict their labels $\hat{Y}_{unlabel}$. So our objective is to let the prediction result agree with the true labels $\bar{Y}_{unlabel}$.

Definition 1 (Community). *We define a community as a group of users $V_c \in V$, who have strong connection with other users in one or more groups and networks in an online social network. They may not be friends, even have not similar profiles.*

According to the definition, we prepare community data. First, we construct all the communities according to the network and group information in online social networks (details in Section 2.3). Then, we can express all communities in a weight matrix. In a community $C_i, i \in N^+$, there are n_i^c members strongly connecting with each other, $v_{j_1}, v_{j_2}, \cdots, v_{j_{n_i^c}}$. We could express their relationship in a $(l+u) \times (l+u)$ matrix W_i^c. Then, for all communities, $C_1, C_2, \cdots, C_i, \cdots, C_{n_c}$, we have a separate matrix for each of them, $W_1^c, W_2^c, \cdots, W_i^c, \cdots, W_{n_c}^c$. The community weight matrix is $W^c = \sum_{i=1}^{n_c} W_i^c$, where n_c is the total number of communities in the data sample.

Model Building Up. In this part, we show the process of building up the new graph-based SSL exploit model step by step. First we construct the local data $W^g = (1 - \gamma)W^p + \gamma W^{fd}$, where W^p is a similarity matrix of personal information, W^{fd} for friendship and $0 < \gamma < 1$. Then, let $D_{ii}^g = \sum_{j=1}^{l+u} W^g(i,j), i \in \{1, \cdots, l, l+1, \cdots, l+u\}$ and D^g be the $(l+u) \times (l+u)$ diagonal matrix by placing D_{ii}^g on the diagonal. Now the unnormalized Laplacian matrix L^g is defined as $L^g = D^g - W^g$. Similarly, L^c for community information is constructed from W^c and D^c. Finally, based on the Local and Global Consistency (LGC) graph-based learning [11], we add the constrain of communities and formulate the problem as

$$\min_{\hat{Y} \in Y_{label}^{l+u}} tr\{\hat{Y}^\top L^g \hat{Y} + \mu_1 (\hat{Y} - \bar{Y})^\top (\hat{Y} - \bar{Y}) + \mu_2 \hat{Y}^\top L^c \hat{Y}\}, \qquad (1)$$

where the predicted result $\hat{Y} = (\bar{Y}_{label},\ \hat{Y}_{unlabel})^\top$, real label $\bar{Y} = (\bar{Y}_{label},\ \bar{Y}_{unlabel})^\top$ and $\mu_1, \mu_2 > 0$. $Y_{label} = \{0,1\}^{n_{label}}$, where n_{label} is the number of different labels. With this step, we have built up a complete CG SSL exploit model to solve the problem.

2.2 Algorithms

In this section, we propose two methods to solve the optimization problem we have formulated before. The first one is a closed form algorithm. Utilizing this method, the exact final result can be obtained directly. The other one is an iterative algorithm, by which we could compute an approximate result. This would be a time-consuming method, but it is able to deal with large-scale datasets.

To simplify the problem, we relax it and solve it. By the definition of this model in Eq. (1), we realize that this is an integer programming problem, which is hard to solve in the consideration of computational complexity. Thus, we relax the feasible region from discrete $\{0,1\}^{(l+u)\times n_{label}}$ to continuous $\{[0,1]\}^{(l+u)\times n_{label}}$.

Closed Form Algorithm. Here we first develop a regularization framework for the optimization problem formulated before. Rewriting the objective function associated with F replacing $\hat{Y}$ in Eq. (1), $F \in \{[0,1]\}^{(l+u)\times n_{label}}$, we have

$$\mathcal{Q}(F) = \frac{1}{2}\Big(\sum_{i,j=1}^{l+u} W^g_{ij}\Big\|\frac{1}{\sqrt{D^g_{ii}}}F_i - \frac{1}{\sqrt{D^g_{jj}}}F_j\Big\|^2 + \mu_1\sum_{i=1}^{l+u}\|F_i - y_i\| + \mu_2\sum_{i,j=1}^{l+u} W^c_{ij}\Big\|\frac{1}{\sqrt{D^c_{ii}}}F_i - \frac{1}{\sqrt{D^c_{jj}}}F_j\Big\|^2\Big), \quad (2)$$

where μ_1, μ_2 are regularization parameters and $\mu_1, \mu_2 \geq 0$. Here the first term (local consistency) and the third term (community consistency) is normalized with $\sqrt{D^g_{ii}}$ and $\sqrt{D^c_{ii}}$. $\frac{1}{2}$ is for the convenience of differentiation and does not affect the classification result. By mathematical deriving, the optimal solution is $F^* = (1-\alpha-\beta)(I - \alpha S - \beta C)^{-1}Y$, where $S = D^{g^{-1/2}} W^g D^{g^{-1/2}}$, $C = D^{c^{-1/2}} W^c D^{c^{-1/2}}$, $Y = (\bar{Y}_{label}, 0)$ and $\alpha = \frac{1}{1+\mu_1+\mu_2}$, $\beta = \frac{\mu_2}{1+\mu_1+\mu_2}$.

We need to design a strategy to make a final decision from F^*. Because we relax the problem before we solve it, the answer F^* is only the probability of unlabeled data belonging to labels, instead of the final result. $F^*(i,j)$ means the probability of the i-th vertex belonging to the j-th label. Thus, we may choose the label with the largest probability as the final label of a vertex, $\hat{y_i} = \arg\max_{1\leq j\leq n_{label}} F^*(i,j)$. According to this strategy, the closed form formular is clearly equivalent to

$$F^* = (I - \alpha S - \beta C)^{-1}Y, \quad (3)$$

where $0 < \alpha \leq 1, 0 \leq \beta < 1$ and $0 \leq \alpha + \beta \leq 1$.

Thus, we could develop a very simple closed form algorithm to solve the problem according to the Eq. (3).

Algorithm 1. Closed Form Algorithm for Community-Based Graph SSL

```
Input: Graph matrix $W^g$ and community matrix $W^c$.
1: Construct the matrices $S$ & $C$.
2: Predict the probability $F^*$ of every label by Eq. (3).
3: Decide the final labels $\hat{y_i} = \arg\max_{j\leq n_{label}} F^*(i,j)$.
Output: Predicting labels $\hat{Y}$.
```

Iterative Algorithm. Because of the need of processing large-scale dataset and the drawback of the closed form algorithm, we proposed an iterative algorithm.

Algorithm 2. Iterative Algorithm for Community-Based Graph SSL

```
Input: Graph matrix $W^g$ and community matrix $W^c$.
Initialize $F(0) = Y = (\bar{Y}_{label}, 0)$.
Construct the matrices $S$ & $C$.
repeat
    $F(t+1) = \alpha SF(t) + \beta CF(t) + (1 - \alpha - \beta)Y$.
until $|F(t) - F(t-1)| < \varepsilon$
Decide the final labels $\hat{y}_i = \arg\max_{j \leq n_{label}} F_{i,j}(t)$.
Output: Predicting labels $\hat{Y}$.
```

According to the mathematical deriving, we could obtain the limitation of $F(t)$ is equal to F^* in the Algorithm 1. Moreover, we easily found that the computation in every iteration only contains multiplication and addition of matrix, which have low computational complexity comparing to the computation of inverse matrix in the closed form algorithm.

2.3 Community Generation

In this section, we discuss the details of generating all the possible communities based on the groups and networks information in online social networks. First, we define the "distance" d between any two user v_i and v_j, $d_{i,j} = e^{-\|\langle w^{gp}_{i,j}, w^{nk}_{i,j} \rangle\|}$. According to this, we utilize a clustering method K-mean to generate communities $C_1, C_2, \cdots, C_{n_c}$.

In fact, many other methods can be utilized to generate communities, e.g., Gaussian Mixture Model (GMM) and Graph Cut. But no matter what method is applied to generate communities, the CG SSL exploit model is still in effect.

3 Experiments

In the experiments, we employ both novel SSL exploit model and other three learning models as comparison, including two graph-based SSL and a supervised learning model, to predict the labels on a synthetic dataset and exposing which universities users come from on two real-world datasets. The results are evaluated in terms of accuracy and weighted accuracy on these three datasets.

3.1 Dataset Description

We describe the details of three datasets in this part. Table 1 gives detail statistics of these three datasets.

TwoMoons Dataset. 'TwoMoons' is a simple dataset only with 2 classes and 200 vertices distributing in $2D$ space. The distribution of the original data is shown in Table 2. Based on this, friendship information (local similarity) and community information

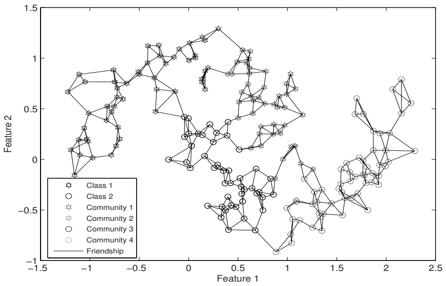


Fig. 2. The Synthetic Friendship and Networks Information of TwoMoons Dataset

Table 1. Statistics of TwoMoons, StudiVZ and Facebook Datasets

Dataset	Vertices	Edges	Groups	Networks	Classes
TwoMoons	200	381	0	4	2
StudiVZ	1, 423	7, 769	406	0	6
Facebook	10, 410	45, 842	61	78	3

Table 2. Statistics of Data Distribution on TwoMoons Dataset

Class	Class 1	Class 2
Size of Class	95	105

Table 3. Statistics of Data Distribution on StudiVZ Dataset

University	LMU Muenchen	Uni Wien	Uni Bayreuth
Size of Class	128	79	98
University	**Uni Frankfurt am Main**	**TU Wien**	**(Others)**
Size of Class	74	70	974

(community similarity) are artificially generated. These two kinds of synthetic information are shown in Fig. 2.

StudiVZ Dataset. The dataset has sufficient information of users' profiles and groups. Based on crawled data, we build a graph which contains $1,423$ vertices and $7,769$ edges. Data distribution is shown in Table 3.

Facebook Dataset. The dataset has sufficient number of vertices and all kinds of relational information, thus it is similar to the situation of the real world. Comparing with StudiVZ dataset, Facebook dataset has much more missing values in personal profile and more group information.

3.2 Data Preprocessing

For two real-world datasets, a series of data preprocessing such as feature selection, data cleaning and data translation are conducted before running algorithms.

Feature Selection. For users' profile information, we select top three features for which most people provide information. For relational information, a number of small groups and networks are removed. Apart from that, some networks whose names explicitly reveal universities' names, such as "LMU Muenchen", are removed manually.

Table 4. Statistics of Data Distribution on Facebook Dataset

University	CUHK	HKUST	(Others)
Size of Class	68	$1,583$	$8,759$

Data Translation. We need to translation some data into the proper forms. For example, we translate home town to its longitude and latitude values through Google maps API[3] to calculate the similarity. Moreover, missing data are filled with average value of existed data and noise data are treated as missing ones. Cosine similarity is applied between any two profile vectors. If both of the users fail to provide at least 50% information, we set the cosine similarity with mean value.

3.3 Experiment Process

Labeled Data Selection. Labeled data are selected randomly with two constrains: 1. each class must have labeled data; 2. the numbers of labeled data in all classes are similar. The second point suggests an assumption that we do not know the distribution of all classes when labeling data.

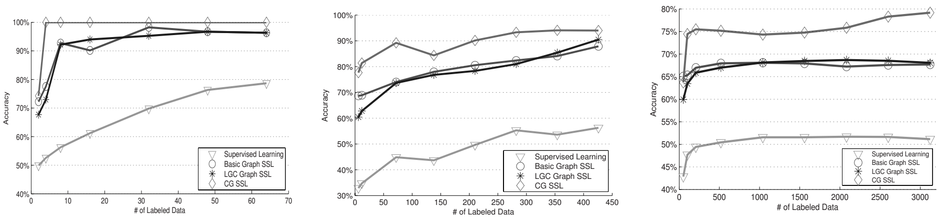
Evaluation Criterion. We mainly utilize the accuracy to measure the results of learning and a Weighted Accuracy (WA) measurement would assist us to analyst the confidence of the learning results. We define WA as $\frac{\sum_{i \in V_c} F^*(i,\hat{y}_i)}{\sum_{i \in V_c} F^*(i,\hat{y}_i) + \sum_{i \in V_{inc}} F^*(i,\hat{y}_i)}$, where V_c is a set containing all the vertices whose predictions are correct and V_{inc} contains all incorrect-prediction vertices.

3.4 Experiment Results

Table 5, 6 and 7 give the results of experiments, from which various algorithms' performance can be evaluated. Figure 3(a), 3(b) and 3(c) describe the accuracy of prediction with TwoMoons, StudiVZ and Facebook datasets respectively. What's more, the results of supervised learning are provided for comparison.

TwoMoons. Figure 3(a) shows the predicting results on the synthetic dataset 'TwoMoons'. First, the accuracy of graph-based SSL models is obviously better than that of supervised learning. Second, Consistencies make the learning models stabler. The global consistency makes the LGC SSL stabler - the learning accuracies would keep enhancing along with the increasing of the number of labeled data. Moreover, the community consistency keeps the CG SSL stably better than other graph-based SSL models. Third, the community information does help in prediction in term of accuracy. In Fig. 2, we observe that some vertices have strong local similarity (friendship) with each other, but actually they do not belong to the same class. Without the help from community information, basic graph SSL and LGC SSL always incorrectly predict the classes of 6 to 8 vertices (Table 5), even if the percentage of labeled data is pretty high. The experiment on this synthetic dataset illustrates that the CG SSL could really improve the learning result in some ideal conditions.

[3] http://code.google.com/apis/maps/



(a) **TwoMoons Dataset with** 200 **Vertices** (b) **StudiVZ Dataset with** 1, 423 **Users** (c) **Facebook Dataset with** 10, 410 **Users**

Fig. 3. Accuracy of Prediction on Three Datasets

Table 5. Accuracy & Weighted Accuracy of Learning on TwoMoons Data with 200 Vertices

# of Labeled Data	**Labeled %**	**Supervised Learning**	**Basic Graph SSL (weighted acc.)**	**LGC SSL (weighted acc.)**	**CG SSL (weighted acc.)**
2	1.00%	49.9%	72.22% (73.77%)	67.71% (98.24%)	**74.24% (100.00%)**
4	2.00%	52.55%	77.55% (83.34%)	72.96% (98.35%)	**100.00% (100.00%)**
8	4.00%	56.25%	92.71% (95.18%)	92.19% (98.95%)	**100.00% (100.00%)**
16	8.00%	61.20%	90.16% (93.05%)	93.99% (99.23%)	**100.00% (100.00%)**
32	16.00%	69.82%	98.22% (98.80%)	95.27% (99.38%)	**100.00% (100.00%)**
48	24.00%	76.32%	96.71% (97.78%)	96.71% (99.34%)	**100.00% (100.00%)**
64	32.00%	78.68%	96.32% (97.50%)	96.32% (99.17%)	**100.00% (100.00%)**

Table 6. Accuracy & Weighted Accuracy of Learning on StudiVZ Data with 1, 423 Users

# of Labeled Data	**Labeled %**	**Supervised Learning**	**Basic Graph SSL (weighted acc.)**	**LGC SSL (weighted acc.)**	**CG SSL (weighted acc.)**
6	0.42%	32.89%	68.67% (68.71%)	60.41% (59.89%)	**77.66% (70.91%)**
12	0.84%	34.73%	68.89% (68.95%)	62.79% (62.76%)	**81.40% (84.17%)**
72	5.06%	44.86%	74.02% (74.26%)	73.65% (73.88%)	**89.23% (92.06%)**
138	9.70%	43.66%	77.98% (78.37%)	76.81% (77.36%)	**84.35% (83.71%)**
210	14.76%	49.63%	80.54% (81.05%)	78.32% (78.98%)	**90.15% (91.40%)**
282	19.82%	55.30%	82.38% (82.91%)	80.98% (81.67%)	**93.30% (98.91%)**
354	24.88%	53.60%	84.10% (84.62%)	85.41% (85.94%)	**94.06% (95.06%)**
426	29.94%	56.27%	87.86% (88.29%)	90.47% (90.83%)	**94.00% (91.52%)**

StudiVZ. Figure 3(b) gives similar results. First, all graph-baesd SSL models outperform supervised learning. Second, the performance of CG SSL with 138 labeled data is worse than that with only 72 labeled data. We conjecture that it is due to the unstable of the clustering technique for generating communities. Although we could tend to the optional predicting result, the randomness of clustering still exists and affects the stability of the final learning results.

Facebook. Figure 3(c) illustrates various algorithms' performance on Facebook dataset. First of all, in most cases the results of SSL methods are still superior to supervised learning. Second, even if there are only a few labeled data, CG SSL method can still make good predictions. The last point is that there is little instability in CG SSL model. The accuracy of learning with 10.00% labeled data is a little worse than that with only 4.99% labeled data. This would be caused by the same reason as in the experiment on the StudiVZ dataset.

Comparing with StudiVZ dataset, the learning results of CG SSL on Facebook dataset are less accurate. This is probably due to the existing of many missing values in Facebook dataset. However, the difference between CG SSL and other two graph-based SSL models is more obvious on Facebook dataset. We conjecture the reason is that there

Table 7. Accuracy & Weighted Accuracy of Learning on Facebook Data with 10, 410 Users

# of Labeled Data	Labeled %	Supervised Learning	Basic Graph SSL (weighted acc.)	LGC SSL (weighted acc.)	CG SSL (weighted acc.)
51	0.49%	42.86%	**65.05%** (58.72%)	59.93% (53.29%)	63.71% (**62.05%**)
102	0.98%	47.69%	65.32% (59.68%)	63.45% (61.53%)	**74.44%** (**74.83%**)
507	1.99%	49.43%	66.97% (61.67%)	65.88% (69.43%)	**75.49%** (**76.52%**)
519	4.99%	50.45%	67.94% (65.07%)	67.00% (**82.27%**)	**75.17%** (72.51%)
1041	10.00%	51.56%	68.09% (66.19%)	68.14% (**88.85%**)	**74.32%** (75.69%)
1560	14.99%	51.57%	67.93% (66.35%)	68.47% (**91.29%**)	**74.76%** (78.14%)
2082	20.00%	51.69%	67.24% (65.89%)	68.68% (**92.43%**)	**75.84%** (79.94%)
2601	24.99%	51.65%	67.58% (65.31%)	68.51% (**92.81%**)	**78.30%** (84.96%)
3123	30.00%	51.16%	67.71% (65.31%)	68.05% (**92.59%**)	**79.17%** (80.32%)

is more effective community information on Facebook dataset and the distribution of community information is more relative and helpful to the predicted attribute.

Summary. In terms of both accuracy and confidence (or certainty), CG SSL exploit model performs better than other two graph-based learning models in most cases and its advantage is amplified gradually when the number of labeled data increases.

4 Related Work

Since the online social networks began to thrive, there has been a growing interest in the security of users' privacy under the current privacy protection. Among the previous work, the exposures using machine learning with public profile and relation information attract a large amount of attention and have great significance in the security of online social networks [7,10]. The exposures employing machine learning methods include supervised learning and unsupervised learning at the beginning.

Based on the characteristic of semi-supervised learning, it has attracted many researchers to study in the last decades. Semi-supervised learning can be divided into several typical kinds of models, including generative model [4], co-training method [3], graph-based methods [12], SVM [8,9], etc.

Graph-based semi-supervised learning methods model objects as weighted undirected graphs. Blum and Chawla [2] pose semi-supervised learning as a graph mincut problem. Another graph-based semi-supervised learning algorithm proposed in [12] is the harmonic function that is a function which has the same values as given labels on the labeled data and satisfied the weighted average property on the unlabeled data. Based on [12], [11] proposes the Local and Global Consistency graph-based method which improves harmonic function method.

5 Conclusions

Community-based Graph SSL model describes the real world exactly. With the help of community consistency, this model illustrates the further relationship among all users in the real world. Moreover, this paper provides two algorithms to solve the problem. In contrast with previous graph-based SSL models, CG SSL predicts the sensitive information of online social networks with higher accuracy and confidence. Thus, the privacy exposure problem in online social networks becomes more serious and the security of users' information is no longer secure.

References

1. http://www.facebook.com/press/info.php?statistics
2. Blum, A., Chawla, S.: Learning from labeled and unlabeled data using graph mincuts. In: Proc. of ICML, pp. 19–26. Citeseer (2001)
3. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual Conference on Computational Learning Theory, p. 100. ACM, New York (1998)
4. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(1), 1–38 (1977)
5. Getoor, L., Taskar, B.: Introduction to statistical relational learning. MIT Press, Cambridge (2007)
6. He, J., Chu, W., Liu, Z.: Inferring privacy information from social networks. In: Mehrotra, S., Zeng, D.D., Chen, H., Thuraisingham, B., Wang, F.-Y. (eds.) ISI 2006. LNCS, vol. 3975, pp. 154–165. Springer, Heidelberg (2006)
7. Mo, M., Wang, D., Li, B., Hong, D., King, I.: Exploit of Online Social Networks with Semi-Supervised Learning. In: Proc. of IJCNN, pp. 1893–1900. IEEE, Los Alamitos (2010)
8. Vapnik, V.: Structure of statistical learning theory. Computational Learning and Probabilistic Reasoning, 3 (1996)
9. Xu, Z., Jin, R., Zhu, J., King, I., Lyu, M., Yang, Z.: Adaptive Regularization for Transductive Support Vector Machine. In: Proc. of NIPS, pp. 2125–2133 (2009)
10. Zheleva, E., Getoor, L.: To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proc. of WWW, pp. 531–540. ACM, New York (2009)
11. Zhou, D., Bousquet, O., Lal, T., Weston, J., Scholkopf, B.: Learning with local and global consistency. In: Proc. of NIPS, pp. 595–602. MIT Press, Cambridge (2004)
12. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using Gaussian fields and harmonic functions. In: Proc. of ICML (2003)