Are We Really Friends? Link Assessment in Social Networks Using Multiple Associated Interaction Networks

Mohammed Abufouda and Katharina A. Zweig
University of Kaiserslautern
Gottlieb-Daimler-Str. 48
67663 Kaiserslautern, Germany
{abufouda, zweig}@cs.uni-kl.de

ABSTRACT

Many complex network systems suffer from noise that disguises the structure of the network and hinders an accurate analysis of these systems. Link assessment is the process of identifying and eliminating the noise from network systems in order to better understand these systems. In this paper, we address the link assessment problem in social networks that may suffer from noisy relationships. We employed a machine learning classifier for assessing the links in the social network of interest using the data from the associated interaction networks around it. The method was tested with two different data sets: each contains the social network of interest, with ground truth, along with the associated interaction networks. The results showed that it is possible to effectively assess the links of a social network using only the structure of a single network of the associated interaction networks and also using the structure of the whole set of the associated interaction networks. The experiment also revealed that the assessment performance using only the structure of the social network of interest is relatively less accurate than using the associated interaction networks. This indicates that link formation in the social network of interest is not only driven by the internal structure of the social network, but also influenced by the external factors provided in the associated interaction networks.

Keywords

Link assessment, multiple networks, social network analysis

1. INTRODUCTION

Many complex networks suffer from noise, i.e., links that do not reflect a real relationship. These noisy links, especially false-positives, harm the quality of the network and, subsequently, hinder an accurate analysis of these networks. For example, researchers often base their analysis of protein-protein interaction networks on so-called high-throughput data. This process is highly erroneous (up to 50% false-positives and 50% false-negatives) [6] and thus results noise in the constructed protein-protein interaction networks. In response, a process for assessing the quality of the links in these networks is inevitable in order to get a high quality representation of the studied phenomenon and an accurate conclusion of the analysis results. Thus, many researchers have started assessing the quality of these biological networks [10, 5].

In social networks, the situation is similar as many online social networks experience such kind of noisy friendship relations. One reason for the noise relationship is the low cost of forming a link in the online social network platforms which results in a large number of connections for a member. Another reason for the existence of noisy relationships is that the automatic sent of the invitations when you firstly register in one of the social network platforms; these invitations may contribute in connecting with a person you really do not know in real life but you have contacted him/her once for any reason. An example of this situation is the Twitter social network where it is easy to be followed by those who are not real friends or those who have fake accounts. A member of an online social network can easily connect, as a friend, to another member based on strong motivation, like being a real-life friend or participating in the same political party, or weak motivation like being a friend of someone you know. This variation in the type of the friendship links in social networks led some researchers to quantify the relationship strength [9, 22, 24] within social networks. Pappalardo et. al [19] proposed a a multidimensional model to capture the strength of the ties in social networks of the same actors. Another related work was done by Xie et. al [23] where the authors investigated Twitter's users to identify real friends.

Most online social networks have an interesting feature which is the *unfriend* or to *unfollow* feature which, once used, removes the connection between the two members. A recent study has shown that at least 63% of Facebook users unfriended at least one friend for different reasons [21]. The reasons for unfriending, according to Sibona [21], are frequent or useless posts, political and religious polarization, inappropriate posts, and others. This means that social networks suffer from noisy relationships that need to be eliminated in order to keep only the desired friends. Accordingly, it is clear that the online social networks do contain many false-positive links that push the members to use the unfriend/unfollow feature or, less extremely, categorize unwanted connections as restricted members.

In this work, we aim at assessing the relationships within the social network (SN) based on the structure of networks related to the social network of interest SN. We call these networks $Associated\ Interaction\ Networks\ \mathcal{G}$. One contribution of this work is that we take a further step in under-

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05
http://dx.doi.org/10.1145/2740908.2742468.

standing the dynamics of social interactions by investigating multiple and *correlated* networks of the same actors [1]. That is because looking merely at an individual network, in this case the SN, is a rather simplistic abstraction of the social interaction, which is not sufficient to understand its dynamics [3]. Thus, utilizing the interaction networks that affects the structure of the social network is considered in this work.

Running Example: To better understand the concept of link assessment with associated interaction networks, let us consider a research center environment where the members can socialize online using the Facebook social network SN, which is chosen as the social network of interest. Along with the Facebook friendship network, the members of the research group have different interactions that affect the structure of their Facebook friendship network SN. These associated interaction networks $\mathcal{G}$ include:

- Work G_1: Where a link exists between two members if they work/ed in the same department.
- Co-author G_2: Where a link exists between two members if they co-authored publication.
- Lunch G_3: Where a link exists between two members if they had lunch together at least one time.
- Leisure G_4: Where a link exists between two members if they participated in the same leisure activity at least one time.

Obviously, these associated interaction networks affect the structure of the social friendship network SN as the link formation process within any social network is not only driven by its structure, i.e. internal homophily [17], but also extremely influenced by external factors (associated interaction networks $\mathcal{G}$) [1]. For example, it is highly probable that any two persons who have lunch and/or spent some leisure time together to be friends in the SN. However, if there is a friendship link between two members A and B in the SN and there is no link between A and B in any of the networks in $\mathcal{G}$, then this relationship might be a noisy one or it is a very low strength link that is not qualified to be a real friendship relation. In Figure 1, it is intuitive to deduce that the links that exist in the social network SN and also exist in any other network $G_i \in \mathcal{G}$ are presumably not a noise as the links in the associated interaction networks affects the link formation in the social network SN. On the other hand, there are 44 links that are not in any $G_i \in \mathcal{G}$, which pose the question: *Are these edges noise?*. In fact we cannot say that these 44 links are utterly noisy links because it is hard to capture all of the possible relations between the actors of this data set in real life. For example, one link of the 44 might be between two researchers who are living in the same building or they are members of the the same political party; data that we do not have or that is hard to be collected.

We think that utilizing exogenous information, like associated interaction networks $\mathcal{G}$ around the SN, with machine learning can help us to identify the noisy edges in the social network. More precisely, this work aims at answering empirically the following research questions:

- **RQ1**: Does every interaction network $G_i \in \mathcal{G}$ encompass sufficient information to assess the social network SN?
- **RQ2**: Does the whole set of the interaction networks provide a better link assessment than using a single interaction network?
- **RQ3**: Which is better for assessing the links of a social network SN, using *only* the structure of a social network SN or combining it with the structures of the associated interaction networks $\mathcal{G}$?

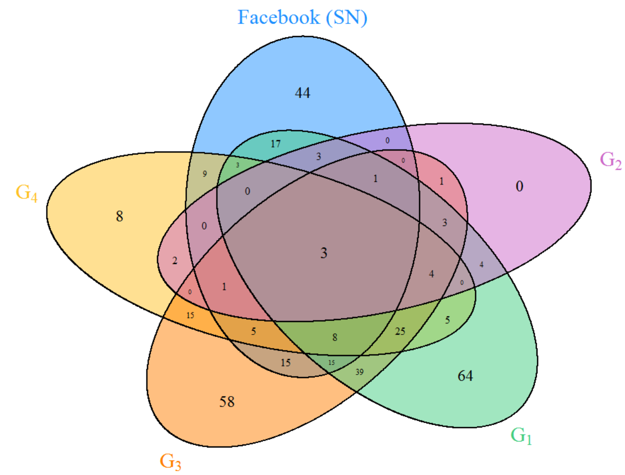

Figure 1: (Color online) The Venn diagram for edge overlapping between the Facebook social network SN and the other associated interaction networks $\mathcal{G}$ for the research group data set.

2. THE PROPOSED METHOD

This section presents the details of the proposed method. It starts with a general description of the framework followed by a detailed information.

2.1 Framework description

As described earlier, our aim is to assess the links in a social network SN using the associated interaction networks of the same actors. We employ machine learning classification techniques by building a feature data model (FDM) of the associated interaction networks and the social network of interest SN. The idea of the proposed framework is to convert the link assessment problem into a machine learning classification problem where the classifier should indicate whether a particular link is noise or a real link via learning from the associated interaction networks. In the following, a description of the FDM is provided.

2.2 Feature data model (FDM) specification

The FDM is a data model that represents a network structure using topological features for each pair of nodes of a network. More formally, $\forall\ v, w \in V(G_i)$ where $v \neq w$ and $G_i \in \mathcal{G}$, we find the feature value $F_i(v, w)$ such that $F_i \in \mathcal{F}$, and $\mathcal{F}$ is the set of features that will be described in the following section.

2.2.1 *Node pair-wise dependent features*

The node-dependent features include:

- *The number of Common Neighbors* ($\mathcal{CN}$): For any node z in a graph g, the neighbors of z, $\Gamma(z)$, is the set of nodes that are adjacent to z. For each pair of nodes v and w, the number of common neighbors these two nodes is the number of nodes that are adjacent to both nodes v and w.

$$\mathcal{CN}(v, w) = |\Gamma(v) \cap \Gamma(w)| \tag{1}$$

- *Resource Allocation* ($\mathcal{RA}$): This measure was proposed by Zhou et al. [25] for addressing link prediction and showed a slightly better performance than $\mathcal{CN}$. This measure assumes that each node has given some resources that will be distributed equally among its neighbors. This concept is adapted by incorporating two nodes v and w.

$$\mathcal{RA}(v,w) = \sum_{\substack{z \in \{\Gamma(v) \cap \Gamma(w)\} \\ z \neq v \neq w}} \frac{1}{|\Gamma(z)|} \tag{2}$$

- *Adamic-Adar coefficient* ($\mathcal{AAC}$): Since proposed by Adamic et al. [2], the Adamic-Adar coefficient has been used in different areas of social network analysis. The idea behind this measure is to count the common neighbors weighted by the inverse of the logarithm.

$$\mathcal{AAC}(v,w) = \sum_{\substack{z \in \{\Gamma(v) \cap \Gamma(w)\} \\ z \neq v \neq w}} \frac{1}{log|\Gamma(z)|} \tag{3}$$

- *Jaccard Index* ($\mathcal{JI}$): This measure was first proposed in information retrieval [20] as a method to quantify the similarity of the contents of two sets.

$$\mathcal{JI}(v,w) = \frac{|\Gamma(v) \cap \Gamma(w)|}{|\Gamma(v) \cup \Gamma(w)|} \tag{4}$$

- *Preferential Attachment* ($\mathcal{PA}$): In collaboration networks, Newman [18] showed that the probability of collaboration between any two nodes (authors) v and w is correlated to the product of $\Gamma(v)$ and $\Gamma(w)$.

$$\mathcal{PA}(v,w) = |\Gamma(v)| \cdot |\Gamma(w)| \tag{5}$$

The previous similarity measures are used for undirected networks. For directed networks, two versions of each measure are used by providing two versions of the neighborhood set Γ, the in-neighbors $\Gamma(v)_{in}$ and the out-neighbors $\Gamma(v)_{out}$. Based on this, an *in* and an *out* version of the above measures can be constructed. For example, the $\mathcal{CN}_{in}$ for two nodes v, w is:

$$\mathcal{CN}(v,w)_{in} = |\Gamma(v)_{in} \cap \Gamma(w)_{in}| \tag{6}$$

2.2.2 Network-dependent feature

To test the RQ2, we combined all of the FDM_{G_i} into a combined $\text{FDM}_{\mathcal{G}}$. That means, a pair of nodes v, w appears N times, where N is the number of associated interaction networks, in the combined $\text{FDM}_{\mathcal{G}}$. Therefore, it is crucial to label the instances of the $\text{FDM}_{\mathcal{G}}$ with network-dependent features. This feature helps the classification algorithm to discriminate different instances of v, w when their node-dependent features are close to each other. We used the network density as a network-dependent feature. *Network density* (η): Is a measure that reflects the degree of completeness of a network.

$$\eta(G_i) = \frac{2 \cdot |E(G_i)|}{|V(G_i)| \cdot (|V(G_i)| - 1)} \tag{7}$$

Based on the previous description, the $\text{FDM}_{\mathcal{G}}$ for a undirected network contains a number of instances that is equal to $\sum_{G_i \in \mathcal{G}} \frac{|V(G_i)| \cdot (|V(G_i)| - 1)}{2}$ such that for every pair of nodes $v, w \in V(G_i)$ and $G_i \in \mathcal{G}$, an instance $\mathcal{I}(v,w)$ is a tuple that contains:

- Pair-wise dependent features of v and w presented in Equations 1 to 5.
- Network-dependent features of $\mathcal{G}$ presented in Equation 7.
- A binary classification, $\{1, 0\}$, that indicates whether there is a link $e(v,w)$ in G_i or not [1].

[1]To test the RQ3, the FDM_{SN} is constructed in the same way.

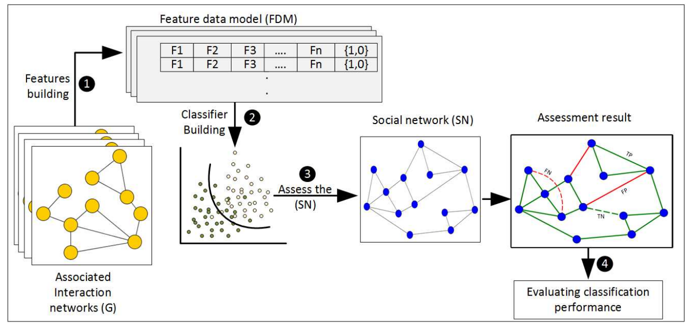

Figure 2: (Color online) The proposed framework for link assessment using associated interaction networks and machine learning classifier. (Adapted from [1])

Figure 2 depicts the process of assessing the links of a social network SN using associated interaction networks $\mathcal{G}$. In step 1, the FDM_{G_i} is constructed for each network $G_i \in \mathcal{G}$. In step 2, the constructed FDMs is used to train a machine learning classifier that is used, in step 3, to assess the SN by providing the binary classification value. In step 4, data with ground truth is used to evaluate the classification performance.

3. DATA SETS AND EVALUATION METRICS

In order to validate the proposed method, we test it using two different social networks with their associated interaction networks. The first data set (RG) is the research center data set described in Section 1. The social network for the research group [16], which is a Facebook social network, is considered as the ground truth online social networks for its members [2]. All the networks of the first data set are undirected. The second data set (LF) is a *law firm* data set [14] that contains an offline directed social network along with the following two associated interaction networks based on questionnaires:

- *Advice* G_1: If a member A seeks an advice from another B, then there is a directed link from A to B.
- *Cowork* G_2: If a member A considers another member B a co-worker, then there is a directed link from A to B.

The social network of the law firm data set in this article is considered as the ground truth social network.

Table 1 shows network statistics of the networks of the data sets used in this paper. These statistics include the number of nodes n, the number of links m, the average clustering coefficient $cc(G)$, and the network's density η.

In order to evaluate the prediction results, we present a set of classical classification evaluation metrics used for evaluating the classification results of our experiment. For any two nodes v and w, true-positive (TP) classification instance means that there is a link $e(v,w)$ between these two nodes in the test set and the classifier succeeds in predicting this link. true-negative (TN) instance means that there is no link $e(v,w)$ in the test set and the classifier predicts that this link does not exist. On the other hand, false-negative (FN) instance means that for a pair of nodes v and w there is

[2]This may sound contradicting to what we conjectured in the introduction concerning the noise in social networks. However, we contacted the owner of the data set and made sure that there is neither false-positives nor false-negatives in the Facebook network which is congruent with our experiment. Additionally, this Facebook network is a closed social network for certain known people where any member knows the other members.

Table 1: Data set statistics.

dataset	Networks	n	m	$cc(G_i)$	$\eta(G_i)$
RG	SN **Facebook**	32	248	0.48	0.5
	G_1 Work	60	338	0.34	0.19
	G_2 Co-author	25	42	0.43	0.14
	G_3 Lunch	60	386	0.57	0.21
	G_4 Leisure	47	176	0.34	0.16
LF	SN **Friends**	69	339	0.43	0.07
	G_1 Co-work	71	726	0.41	0.15
	G_2 Advice	71	717	0.42	0.14

a link $e(v,w)$ in the test set and the classifier tells that there is no link. Similarly, false-positive (FP) instance means that for a pair of nodes v and w there is no link $e(v,w)$ in the test set but the classifier predicts that there is a link.
Based on the previous basic metrics, we used the following additional evaluation metrics:
Precision ($\mathcal{P}$): is the number of true-positives to all positive classification.

$$\mathcal{P} = \frac{TP}{TP + FP} \tag{8}$$

Recall ($\mathcal{R}$): is also called True positive rate or *Sensitivity*.

$$\mathcal{R} = \frac{TP}{TP + FN} \tag{9}$$

Accuracy ($\mathcal{ACC}$): is the percentage of correctly classified instances.

$$\mathcal{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

F-measure ($\mathcal{F}$): is the harmonic mean of precision and recall.

$$\mathcal{F} = \frac{2.\mathcal{P}.\mathcal{R}}{\mathcal{P} + \mathcal{R}} \tag{11}$$

Area under Receiver Operating Characteristics curve ($AU-ROC$): ROC [12] curve plots the true positive rate against the false positive rate. The area under this curve reflects the goodness of the classifier and it is used to compare multiple classifiers' performance.

4. EXPERIMENT AND RESULTS

4.1 Ground truth

In order to validate the method and to provide a reliable evaluation, it is important to use data with ground truth. Let $SN = (V,E)$ be the observed social network of interest and let the ground truth $GT = (V,E')$ be a set of links on the same set of actors. Thus, $E' - E$ contains the false-negative links, i.e., links that exist in reality but were not observed in the SN. Similarly, $E - E'$ contains the false-positive links, i.e., those that do not exist in reality but were observed in the SN. The goal now is to get a classification result as close to GT as possible. That said, the more accurate the machine learning classifier is, the more efficient our link assessment method becomes. Thus, the classification problem is employed for the link assessment problem.
Having the FDM constructed, we utilize it in a machine learning problem *ψ(FDM)*, which means that we use the constructed FDM to train a machine learning classifier and, then, to classify the links in the SN using the FDM_{SN} test set. To test the effectiveness of this method, a social network with ground truth data will be assessed.

In order to validate the proposed solution, the following experiment steps were performed:

1. *Build the FDM*: In this step, we implemented a software that calculates the features described in Section 2.2 which provide the FDM_{G_i} for every network G_i of the associated interaction networks $\mathcal{G}$, and also for the social network of interest FDM_{SN}, where SN is the ground truth.
2. *Training and testing*: Considering the research questions described in Section 1, we train the classifier using different training sets depending on the goal of the experiment. For assessing the links of the social network of interest SN using a single network $G_i \in \mathcal{G}$, the training set was FDM_{G_i} and the test set was the FDM_{SN}. For assessing the links of the social network of interest SN using the whole set of the associated interaction networks, the training set was $FDM_{\mathcal{G}}$ and the test set was the FDM_{SN}. For assessing the links of the social network of interest SN using the SN itself, we train and test using the FDM_{SN} with cross validation technique.
3. *Evaluation*: The evaluation metrics described in Section 3 were used to evaluate the classification results of the training sets.

The Weka [11] data mining and machine learning tool was used for applying the classification. The implemented Logistic Regression($\mathcal{LR}$) classifier with ridge estimator [15] is used with its default configurations. The $\mathcal{LR}$ classifier was chosen because it shows satisfactory and stable prediction performance for the data sets and later we will compare the performance of different classifiers. In the following section, the results and the discussion is presented to answer the research question described in Section 1 by mapping each research question to an experimental hypothesis.

4.2 Results

4.2.1 Assessing the SN using single interaction network $\psi(FDM_{G_i})$

In order to answer RQ1 described in section 1, the following hypothesis is being tested:
H1: *Any single interaction network $G_i \in \mathcal{G}$ encompasses sufficient structure to assess the links of the social network SN.*
To reject the null hypothesis of hypothesis H1, 100 random graphs [7] with the same number of nodes, number of links, and degree sequence as in the G_i were built and used to assess the links in SN. The results are shown in Table 2. It is clear from the table that any random network has no sufficient structure to assess the links in the SN. On the other hand, when the method was applied on the real data, the assessment result were significantly high which supports H1. In all cases, assessing the SN using single network provides a satisfactory performance in terms of the prediction metrics. The only network that provided a relatively low performance was the network G_2, co-author network, of the research group data set. The low f-measure compared to f-measures of the other networks is due to the very low number of edges in this network which contributes in an extremely imbalanced FDM model to learn from. These results reveal also a strong correlation between the interaction networks and the social network. Thus, the answer to RQ1 is: Yes, every interaction network $G_i \in \mathcal{G}$ encompasses sufficient information to assess the social network SN.

4.2.2 Assessing the SN using the set of interaction networks $\psi(FDM_{\mathcal{G}})$

Table 2: Link assessment of the SN using single interaction networks $\mathcal{G}$ for both the real data and the corresponding random networks.

Data Set	Real Data			Random Networks		
	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
RG	$\psi(\mathrm{FDM}_{G_1})$ =0.83	0.83	0.83	0.01	0.024	0.02
	$\psi(\mathrm{FDM}_{G_2})$ =0.78	0.44	0.43	≈ 0	≈ 0	≈ 0
	$\psi(\mathrm{FDM}_{G_3})$ =0.83	0.84	0.84	0.02	0.036	0.03
	$\psi(\mathrm{FDM}_{G_4})$ =0.83	0.81	0.82	0.035	0.027	0.03
LF	$\psi(\mathrm{FDM}_{G_1})$ =0.89	0.91	0.89	0.098	0.167	0.13
	$\psi(\mathrm{FDM}_{G_2})$ =0.84	0.91	0.87	0.08	0.14	0.1

Here, the results of assessing the SN from the set of **all** associated interaction networks $\mathcal{G}$ is provided. To answer RQ2, the following hypothesis is being tested:
H2: *The links of the SN can be effectively assessed using the FDM of the associated interaction networks.*
Table 3 shows that the assessment results using the FDM generated from all interaction networks is not only satisfactory, but also better than the assessment results of the single interaction networks. Thus, the answer to the RQ2 is: Yes, using the whole set of the interaction networks provides a better link assessment than using a single interaction network.

Table 3: Link assessment of the SN using only the $\mathrm{FDM}_{\mathcal{G}}$ and using only the FDM_{SN} itself.

Data set	$\mathrm{FDM}_{\mathcal{G}}$			FDM_{SN}		
	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
$\psi(\mathrm{FDM}_{RG})$	0.84	0.84	0.84	0.82	0.83	0.82
$\psi(\mathrm{FDM}_{LF})$	0.89	0.90	0.90	0.84	0.87	0.85

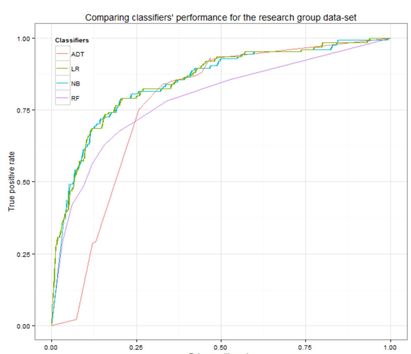

Figure 3: (Color online) Different classifiers' performance for the research group data set link assessment.

4.2.3 *Link assessment using the structure of the social network only*

It is interesting to test whether the assessment of a social network SN can be performed using the SN itself or not. Also, it is important to compare the assessment results with the results obtained from assessing the SN using the interaction networks. To that end, we have the following hypothesis to address RQ3:
H3: *Assessing the links in the SN using the interaction networks is better than assessing them using only the SN itself.*
In order to validate this hypothesis, only the social network SN was used for both training and testing. We used k-fold-cross-validation, with *k=10*, to avoid overfitting. The results in Table 3 show that the assessment using multiple interaction networks is slightly better than using the social network only, which answers RQ3.

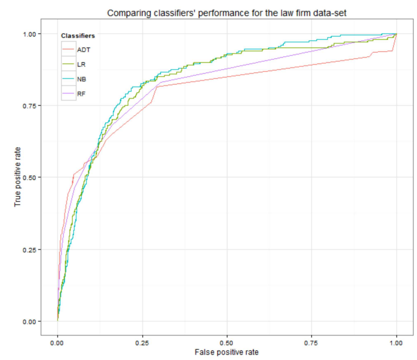

Figure 4: (Color online) Different classifiers' performance for the law firm data set link assessment.

4.2.4 *Comparing different classifiers performance*

It is important to show that the results of our approach for link assessment is classifier independent. Traditionally, AU-ROC is used to compare classifiers' performance. Table 4 shows a comparison of different classifiers performance of the link assessment using the two data sets. It is clear from Table 4 that Logistic Regression($\mathcal{LR}$) [15] provided a slightly better performance compared to Random Forest($\mathcal{RF}$) [4], Naive Bayes($\mathcal{NB}$) [13], and Alternating Decision Trees ($\mathcal{ADT}$) [8] in terms of AU-ROC curve. Nevertheless, the Naive Bayes and the Random forest provide comparable results with the logistic regression. Figure 3 and Figure 4 show the classifiers' performance for the research group data set and the law firm data set, respectively. From those figures, it is not

Table 4: Classifiers benchmarking for link assessment using the whole set of associated interaction networks

Data set	Classifier	Prediction performance				
		$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{ACC}$	AU-ROC
RG	$\mathcal{LR}$	0.834	0.86	0.834	83.47	0.85
	$\mathcal{NB}$	0.83	0.78	0.78	77.62	0.85
	$\mathcal{RF}$	0.79	0.79	0.79	79.032	0.79
	$\mathcal{ADT}$	0.8	0.798	0.8	79.84	0.77
LF	$\mathcal{LR}$	0.84	0.915	0.874	91.5	0.84
	$\mathcal{NB}$	0.89	0.90	0.82	90.54	0.85
	$\mathcal{RF}$	0.89	0.91	0.89	91.7	0.83
	$\mathcal{ADT}$	0.89	0.92	0.89	91.67	0.79

easy to identify which classifier algorithm is better as the learning curves are hard to distinguish. This indicates that the proposed method provides a successful way for assessing the links in the social networks regardless of the used classifier.

4.3 Closing thoughts

Based on the results described and discussed above, it is obvious that the proposed method is in position for assessing the links of the social network of interest. The performance when using a single network $G_i \in \mathcal{G}$ for assessing the links of the SN indicates that there is a correlation between this network and the social network. We think that this correlation contributes in the link formation process of the social network SN. Additionally, more links were assessed correctly when using the whole set of the associated interaction networks than using only the social network SN. This also suggests that the external factors might contribute in forming the structure of a social network. The used data sets have a relatively medium number of nodes and links. Real life social networks contain a larger number of nodes and links which require more computational power for constructing the FDM. This challenge exists for all huge data sets. Also, acquiring a social network with additional interaction networks as well as a ground truth is not easy. Therefore, only two data sets were studied in this paper, as no others were available.

5. CONCLUSION

In this paper, we have addressed the link assessment problem in social networks to identify and eliminate the noisy links in these networks using a set of associated interaction networks that come with the social network. These associated networks were used to build a feature data model to represent each network. Then, the feature data model is used in machine learning classifier to assess the links in the social network. Two data sets were used to test the proposed method, and the results showed that it is possible to assess the links in a social network using any network in the associated interaction networks. Also, the results revealed that the assessment using the interaction networks is slightly better than the assessment using the social network itself, which suggest the existence of a correlation between the associated interaction networks and the formation process within the social network.

6. REFERENCES

[1] M. Abufouda and K. Zweig. Interactions around social networks matter: Predicting the social network from associated interaction networks. In *Advances in Social Networks Analysis and Mining, IEEE/ACM International Conference on*, pages 142–145, 2014.

[2] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.

[3] S. Boccaletti and et. al. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1 – 122, 2014.

[4] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[5] J. Chen and et. al. Systematic assessment of high-throughput experimental data for reliable protein interactions using network topology. In *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on*, pages 368–372. IEEE.

[6] C. M. Deane and et. al. Protein interactions two methods for assessment of the reliability of high throughput observations. *Molecular & Cellular Proteomics*, 1(5):349–356, 2002.

[7] P. Erdős and A. Rényi. On the evolution of random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[8] Y. Freund and L. Mason. The alternating decision tree learning algorithm. In *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, pages 124–133, 1999.

[9] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211–220. ACM, 2009.

[10] D. S. Goldberg and F. P. Roth. Assessing experimentally derived interactions in a small world. *Proceedings of the National Academy of Sciences*, 100(8):4372–4376, 2003.

[11] M. Hall and et al. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[12] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[13] G. H. John and P. Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the eleventh conference on uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.

[14] E. Lazega. *The Collegial Phenomenon: The Social Mechanisms of Cooperation among Peers in a Corporate Law Partnership.* Oxford: Oxford University Press, 2012.

[15] S. Le Cessie and J. Van Houwelingen. Ridge estimators in logistic regression. *Applied statistics*, pages 191–201, 1992.

[16] M. Magnani and L. Rossi. Formation of multiple networks. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 7812 of *Lecture Notes in Computer Science*, pages 257–264. Springer Berlin Heidelberg, 2013.

[17] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, pages 415–444, 2001.

[18] M. E. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):025102, 2001.

[19] L. Pappalardo and et. al. "how well do we know each other?" detecting tie strength in multidimensional social networks. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 1040–1045. IEEE Computer Society.

[20] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill, Inc., 1986.

[21] C. Sibona. Unfriending on facebook: Context collapse and unfriending behaviors. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 1676–1685, Jan 2014.

[22] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 981–990, New York, NY, USA, 2010. ACM.

[23] W. Xie and et. al. When a friend in twitter is a friend in life. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 344–347. ACM, 2012.

[24] X. Zhao and et. al. Relationship strength estimation for online social networks with the study on facebook. *Neurocomputing*, 95:89–97, 2012.

[25] T. Zhou and et. al. Predicting missing links via local information. *The European Physical Journal B*, 71(4):623–630, 2009.