Wirtschaftsinformatik
und Maschinelles Lernen
Stiftung Universität Hildesheim
Marienburger Platz 22
31141 Hildesheim
Prof. Dr. Dr. Lars Schmidt-Thieme
Ahmed Rashed

Master – Seminar Data Analytics I

Exploiting behaviors of communities of twitter users for link prediction

Winter Semester 2018/2019

Raaghav Radhakrishnan

246097, radhakri@uni-hildesheim.de

Abstract

This summary paper explains one of the interesting research directions that has emerged in social network analysis, the link prediction. This paper discusses the proposal of additional information to the previous techniques (based on structural information) for link prediction. The additional information includes the interests or behaviors that the users or nodes have into their communities. To analyze the proposed technique, with an application on Twitter, a fast and less expensive approach, Label Propagation Algorithm (LPA), was used to partition the network into communities. After this, WIC (within and inter community) and W (within-community) form measures were used to extract the information based on the behaviors of users. From these experiments, this paper also reports the results and analysis with other local similarity measures. Finally, analysis and validation of the results on directed and asymmetric large-scale networks show that including the information about interests or behaviors of users into the communities provide better performance for link prediction.

Contents

1 Introduction

This summary paper will discuss the work "Exploiting behaviors of communities of twitter users for link prediction" by Jorge Valverde-Rebaza and Alneu de Andrade Lopes. It was first published online in 2013 in the 3rd volume of Social Network Analysis and Mining journal (SNAM) on Springer. SNAM is a multidisciplinary journal that serves researchers in the field of computer science, network science and many more. The main objective of this summary paper is to explore the proposed technique for link prediction, provide an idea of its methodology and analyze the experiments and results.

Large-scale social networks like Twitter have become highly influential and a part of our day-to-day life. Twitter, a microblogging service, provides a platform for millions of users [1] to share informative data. The enormous growth of online social networks has paved way for several research directions that sets a goal to understand the basis of large-scale networks' social structures. Link prediction problem, being one of these research directions, focuses on the prediction of future relation for the users which allows them to network with others of similar business and personal interests. To predict the future relation, deep understandings on the formation of existing relationships can help us know about the influence of an individual user on his/her friends [2]. Apart from social networks, link prediction has many applications in bioinformatics, ecommerce, security, information retrieval and information extraction, etc.

The goal of the authors was to analyze the performance of link prediction on the directed asymmetric networks by adding the information on interest or behavior of the users into their community. To attain this goal, they used LPA for community detection and WIC and W form hybrid measures to improve the performance of link prediction. The contributions of the authors are threefold:

1. Using the measures proposed by [3] and performing extensive analysis by using link

prediction measures in unsupervised and supervised strategies.

2. Analyzing the importance of community detection in Twitter and improve in link prediction accuracy.

3. Comparing the link prediction methods based on local structural information with hybrid measures based on community information.

The remainder of this summary paper is organized as follows. Section 2 presents the related work and state-of-the-art. Section 3 explains the methodology of link prediction using community information. Section 4 presents the experimental results obtained from Twitter using unsupervised and supervised link prediction strategies. Section 5 summarizes the main findings and conclusions of the authors' work.

2 Related Works

Although link prediction seems to be a common problem, lots of researches and surveys are being carried out to improve the performance of link prediction. The following are some of the researches and related works carried out for the link prediction problem.

In [4] and [5], Liben-Nowell and Kleinberg and Lü and Zhou respectively have compared the performance of unsupervised link prediction methods based on local structural information with global information.

Hasan et al. and Benchettara et al. considered the problem of link prediction as a classification problem with methods based on supervised strategy in their works [6] and [7] respectively.

Fortunato in his work [8] explains that understanding the property of community structure may convey appropriate information about the nodes with similar behaviors in the network.

In [9], Feng et al. conducted experiments and proved that as the community structure of the network grows, the accuracy of link prediction drastically improves.

In order to partition the network into communities, Raghavan UN, Albert R and Kumara S proposed a specific algorithm, LPA in their research work [10] which is computationally efficient and less expensive.

Though many state-of-art techniques were proposed to solve the link prediction problem, most of them were focused on the local or global structural information. However, the authors proposed a different method by including the behaviors of the users into their community. The research on link prediction in complex networks based on cluster information was carried out by the authors in their work [11].

The main influence of the authors was their previous work, structural link prediction using community information [3]. In [3], they proposed two new measures for link prediction: WIC and W form measures. Based on their previous work, they tried to improve the performance of link prediction using WIC and W form measures with community detection algorithm LPA.

3 Methodology

Most of the previous techniques for link prediction were based on structural (or topological) information. As structural information is not enough to achieve a good performance in the link prediction task on large-scale social networks, the use of additional information, such as interests or behaviours that nodes have into their communities, may improve the link prediction performance. So the authors used a set of simple and non-expensive techniques that combine structural with community information for predicting the existence of future links in a large-scale online social network.

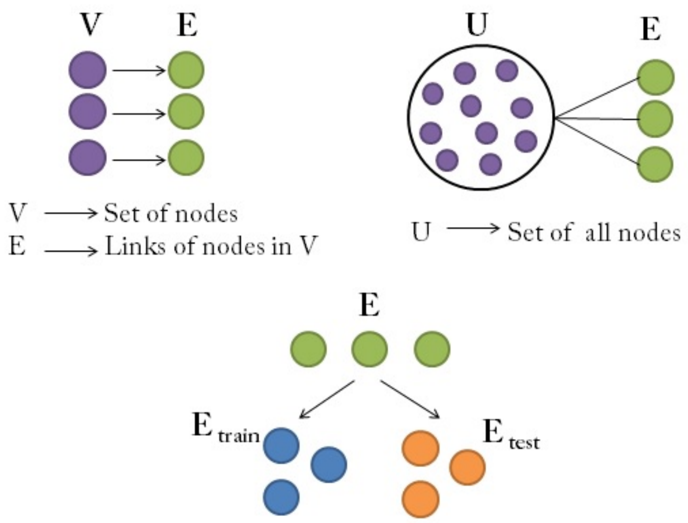


Figure 1: Link prediction in directed network

In Fig. 1), Given a directed network G(V,E), where V and E are sets of nodes and links. Consider the universal set, U, containing all potential directed links between every pair of vertices in V. Thus, the fundamental task is to predict the missing links in set U-E by assigning score for each link. The higher the score, the higher the connection probability is and vice versa. These missing links can be predicted using structural information and community information which are discussed as follows.

3.1 Link Prediction using structural measures

Structural measures use the similarity between nodes since similar nodes likely share same relations (links). Link prediction measures based on similarity can be classified into:

- Local structural information
- Global structural information

Global structural information	Local structural information
Pros: • Large amount of information • High accuracy	Pros: • Faster • Feasible for large-scale networks
Cons: • Time consuming • Infeasible for large-scale networks	Cons: • Only local information is available • Low accuracy

Table 1: Global vs. Local structural information

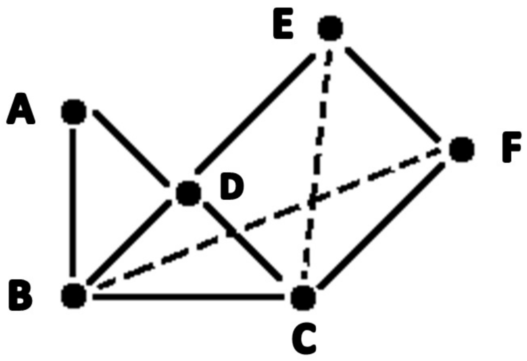


Figure 2: Social network with missing links

Considering the Fig. 2), in order to predict the missing links between (B,F) and (E,C) based on structural measures, the following local similarity measures are used: Common Neighbors, Adamic Adar, Jaccard Coefficient, Resource Allocation and Preferential Attachment

1. Common Neighbors (CN): This measure provides a score based on the set of all common friends of the links as defined in Eq. 1.

$$s_{B,F}^{CN} = |\Gamma(B) \cap \Gamma(F)| \tag{1}$$

$$|\Gamma(B) \cap \Gamma(F)| = \{C\} \qquad and \qquad |\Gamma(E) \cap \Gamma(C)| = \{D, F\}$$

$$s^{CN}_{B,F} = 1 \quad and \quad s^{CN}_{E,C} = 2$$

2. Adamic Adar (AA): This measure refines the score by taking log of the number of connections of the common neighbour as defined in Eq. 2

$$s^{AA}_{B,F} = \frac{1}{log|\Gamma(z)|} \tag{2}$$

$$s^{AA}_{B,F} = 2.09 \quad and \quad s^{AA}_{E,C} = 1.2$$

3. Jaccard Coefficient (Jac): This measures if two users have significant common neighbours with respect to their total neighbors as shown in Eq. 3

$$s^{Jac}_{B,F} = \frac{|\Gamma(B) \cap \Gamma(F)|}{|\Gamma(B) \cup \Gamma(F)|} \tag{3}$$

$$s^{Jac}_{B,F} = 0.2 \quad and \quad s^{Jac}_{E,C} = 0.4$$

4. Resource Allocation (RA): This measure punishes the high-degree common neighbours more than AA does which is shown in Eq. 4

$$s^{RA}_{B,F} = \frac{1}{|\Gamma(z)|} \tag{4}$$

$$s^{RA}_{B,F} = 0.33 \quad and \quad s^{RA}_{E,C} = 0.167$$

5. Preferential Attachment(PA): This measure provides a score which is a product of the neighbours of both the links as shown in Eq. 5

$$s^{PA}_{B,F} = |\Gamma(B)| \times |\Gamma(F)| \tag{5}$$

$$s^{RA}_{B,F} = 6 \quad and \quad s^{RA}_{E,C} = 6$$

In [5], the authors compared various local networks on many real networks and the results stated that Resource Allocation measure achieved the best performance followed by Adamic Adar and Common Neighbours. On the contrary, Preferential Attachment had the worst performance.

3.2 Link Prediction using community information

In order to apply the proposed approaches for link prediction: (1) WIC measure, and (2) W form measure, the authors previously applied a community detection algorithm, Label Propagation Algorithm, on the network. This is because usually the direct use of WIC and W form measures are not recommended on large-scale social networks as they are expensive. As a result, using LPA for community detection in large-scale networks incur less computational complexity.

3.2.1 Label Propagation Algorithm

Tha main goal of community detection algorithms is to find out groups with an inherent similarity among nodes within groups. The LPA is a fast, simple and less expensive method for community detection.

The step-by-step process involved in LPA method is as follows:

1. Assigns unique label to each node

2. For every iteration, a random pass over all nodes is performed such that each node takes label shared by majority of its neighbours

3. If there is no unique majority, a label is selected at random
4. Converges when each node has the majority label of its neighborhood, or a maximum number of iterations is reached.
5. If it is not converged, increase the value of t and the process continues from step 1 until convergence
6. After convergence, each node will have more neighbors in its community than in any other community

The community detection methods do not provide a unique solution as many relationships are encountered during the detection process. This leads to derive different partitions for the same initial condition. Also, the main advantage of using LPA for detecting communities is the fact that no information on the number and size of the communities is required.

The pseudo code of LPA is as shown in Alg. 1

Data: Node with unique label

Result: Identified communities

Initialize node label. For node $x, C_x(0) = x$;

Set $t = 1$;

while *Arrange nodes in random order and set it to X* **do**

  for $x \epsilon X$ **do**

    $C_x(t) = f(C_{x_{i1}}(t), ..., C_{x_{im}}, C_{x_{i(m+1)}}(t-1), C_{x_{ik}}(t-1))$;

    if *every node label is maximum neighbors label* **then**

      return final identified communities;

    else

      set $t = t + 1$;

      go back to the beginning of first section;

    end

  end

end

Algorithm 1: LPA - Pseudo code

3.2.2 WIC measure

As proposed in [9], it can be seen that the accuracy of structural similarity measures improve as the community structure of the network grows. So, in order to make use of the efficiency of link prediction measures based on local information, the WIC measure is used in large scale networks and it is based on Bayesian theorem (Eq. 6).

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{6}$$

Based on Bayesian theorem, given an undirected network with communities detected, the posterior probabilities that the same labels are assigned to (x,y) is given by Eq. 7

$$P(x^{C_\alpha}, y^{C_\alpha}|\wedge_{x,y}) = \frac{P(\wedge_{x,y}|x^{C_\alpha}, y^{C_\alpha})P(x^{C_\alpha}, y^{C_\alpha})}{P(\wedge_{x,y})} \tag{7}$$

Given an undirected network with communities detected, the posterior probabilities that the different labels are assigned to (x,y) is given by Eq. 8

$$P(x^{C_\alpha}, y^{C_\beta}|\wedge_{x,y}) = \frac{P(\wedge_{x,y}|x^{C_\alpha}, y^{C_\beta})P(x^{C_\alpha}, y^{C_\beta})}{P(\wedge_{x,y})} \tag{8}$$

They estimated the probability of set of common neighbours of $(x^{C_\alpha}, y^{C_\alpha})$, belonging to community with label C_α is as stated in Eq. 9

$$P(\wedge_{x,y}|x^{C_\alpha}, y^{C_\alpha}) = \frac{\wedge^W_{x,y}}{\wedge_{x,y}} \tag{9}$$

Similarly to estimate the probability for the nodes of inter community labels, they used the Eq. 10

$$P(\wedge_{x,y}|x^{C_\alpha}, y^{C_\beta}) = \frac{\wedge^I_{x,y}}{\wedge_{x,y}} \tag{10}$$

Substituting Eq. 9 and 10, in 7 and 8, we get the score, $s_{x,y}$ as:

$$s_{x,y} = \frac{\wedge^W_{x,y}}{l^W_{x,y}} \times \frac{P(x^{C_\alpha}, y^{C_\alpha})}{P(x^{C_\alpha}, y^{C_\beta})} \tag{11}$$

The WIC measure for incoming and outgoing neighborhood are computed according to the equations as shown below.

$$s^{WIC_{in}}_{x,y} = \begin{cases} |\wedge^{W_{in}}_{x,y}|, & if|\wedge^{W_{in}}_{x,y}| = |\wedge_{x,y}| \\ \\ \frac{|\wedge^{W_{in}}_{x,y}|}{|\wedge^{I_{in}}_{x,y}|}, & otherwise \end{cases}$$

$$s^{WIC_{out}}_{x,y} = \begin{cases} |\wedge^{W_{out}}_{x,y}|, & if|\wedge^{W_{out}}_{x,y}| = |\wedge_{x,y}| \\ \\ \frac{|\wedge^{W_{out}}_{x,y}|}{|\wedge^{I_{out}}_{x,y}|}, & otherwise \end{cases}$$

3.2.3 W form measure

It can be seen that except Preferential Attachment, other measures based on local structural information depend on the set of all common neighbors. It means that using a common neighbour gives same contribution to the connection likelihod. However, different connection probability are contributed by different common neighbors. Thus the authors proceeded to predict the links using within-community common neighbors as they contribute more to the connection likelihood than the inter-community common neighbors. This is mainly due to the similar behaviors of the within-community neighbors.

For a given directed or undirected network, the scores for the missing links using within-community common neighbors based on local structural information are as shown in Table 2.

Local measures	Undirected networks	Directed networks
Common Neighbours	$s_{x,y}^{\mathrm{CN-W}} = \left\|\Lambda_{x,y}^{W}\right\|$	$s_{x,y}^{\mathrm{CN-W_{in}}} = \left\|\Lambda_{x,y}^{W_{in}}\right\|$ $s_{x,y}^{\mathrm{CN-W_{out}}} = \left\|\Lambda_{x,y}^{W_{out}}\right\|$
Adamic Adar	$s_{x,y}^{\mathrm{AA-W}} = \sum_{z\in\Lambda_{x,y}^{W}} \frac{1}{\log\|\Gamma(z)\|}$	$s_{x,y}^{\mathrm{AA-W_{in}}} = \sum_{z\in\Lambda_{x,y}^{W_{in}}} \frac{1}{\log\|\Gamma_{in}(z)\|}$ $s_{x,y}^{\mathrm{AA-W_{out}}} = \sum_{z\in\Lambda_{x,y}^{W_{out}}} \frac{1}{\log\|\Gamma_{out}(z)\|}$
Jaccard Coefficient	$s_{x,y}^{\mathrm{Jac-W}} = \frac{\left\|\Lambda_{x,y}^{W}\right\|}{\|\Gamma(x)\cup\Gamma(y)\|}$	$s_{x,y}^{\mathrm{Jac-W_{in}}} = \frac{\left\|\Lambda_{x,y}^{W_{in}}\right\|}{\|\Gamma_{in}(x)\cup\Gamma_{in}(y)\|}$ $s_{x,y}^{\mathrm{Jac-W_{out}}} = \frac{\left\|\Lambda_{x,y}^{W_{out}}\right\|}{\|\Gamma_{out}(x)\cup\Gamma_{out}(y)\|}$
Resource Allocation	$s_{x,y}^{\mathrm{RAW}} = \sum_{z\in\Lambda_{x,y}^{W}} \frac{1}{\|\Gamma(z)\|}$	$s_{x,y}^{\mathrm{RA-W_{in}}} = \sum_{z\in\Lambda_{x,y}^{W_{in}}} \frac{1}{\|\Gamma_{in}(z)\|}$ $s_{x,y}^{\mathrm{RA-W_{out}}} = \sum_{z\in\Lambda_{x,y}^{W_{out}}} \frac{1}{\|\Gamma_{out}(z)\|}$

Table 2: W form for local measures

3.2.4 Evaluation measures

To validate the quality of link prediction, the authors used different measures for both supervised and unsupervised strategies. For a unsupervised strategy, Area Under ROC Curve (AOC) and Precision were used by them.

1. AOC: It measures the entire two-dimensional area underneath the entire ROC curve. AUC provides aggregate measure of the accuracy of prediction accross the list of scores of all non-observed links.

2. Precision: This measures the ration between relevant items selected L_T and items selected L.

$$Precision = \frac{L_T}{L} \tag{12}$$

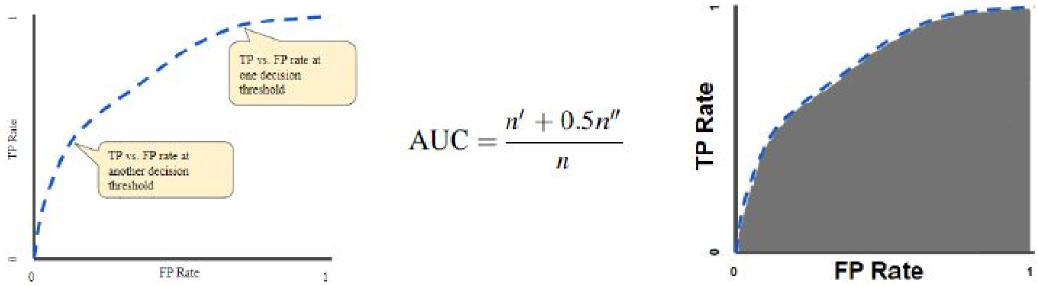
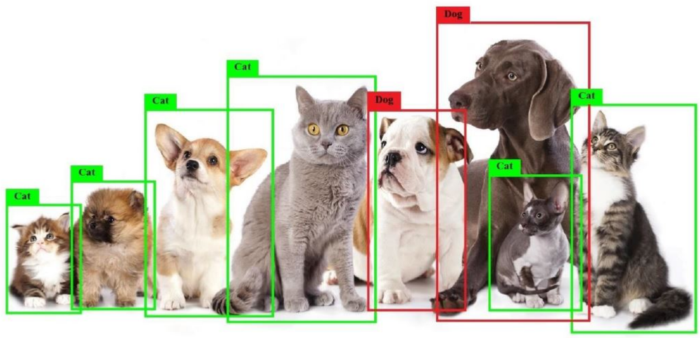


Figure 3: ROC Curve and AUC

For a supervised strategy, among the different standard measures, the authors used Accuracy and F-value to quantify the classifiers performance. An illustration of these measures is given below.



	Cat (True)		Dog (False)	
Actual	4		4	
Predicted	4 (TP)	2 (FP)	2 (TN)	0 (FN)

True Positive: Cat predicted as Cat
False Positive: Dog predicted as Cat
True Negative: Dog predicted as Dog
False Negative: Cat predicted as Dog

Figure 4: Sample prediction of Cat vs. Dog and Confusion Matrix

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \tag{13}$$

$$Precision = \frac{|TP|}{|TP| + |FP|} \tag{14}$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{15}$$

$$F - value = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

4 Experiments

In this paper, the authors considered a scenario of predicting new links of the Twitter network. Initially, for community detection, LPA was applied to assign a community label to each node. Following this, they used supervised and unsupervised strategies and compared the performance of WIC and W form measures to measures based on structural information. The experiments was split up into pre-processing the twitter network based on the requirements and the link prediction process. Finally, they validated and analysed the results with classical link prediction measures.

4.1 Twitter network

For the experiments, the authors chose Twitter network because of the directed relationship nature, i.e., when once a request is accepted, it is not required that the other user should follow back. In their experiments, they removed twitter users with more than 900 followers. Using map-reduce formalism with 55 node Hadoop cluster, they employed the LPA on the Twitter sample. The code available in [12] was used to detect communities. They performed two executions of LPA on the Twitter network, and converged when 7th and 15th iteration was achieved. The results of these executions are tabulated in Table 3.

	Twitter 7it	Twitter 15it
$\|V\|$	24617334	24617333
$\|E\|$	363565896	363565892
M	3415051	2250964
Max cluster size	1392411	10121242
Ratio of total links per user	14.77	14.77

Table 3: Features of the graphs built after 7th and 15th iterations of LPA

4.2 Dataset and Training

The experiments were performed in two phases: the network pre-processing and the link prediction process. In the former phase, the set was divided into training set and testing set. For the testing set, one-third of the links formed by users whose number of followers was two times greater than the ratio of total links per user was taken in random and the remaining links were made to the training set. The pre-processing phase was followed by the link prediction process. In unsupervised strategy, based on the link direction, the connection likelihood was calculated and in supervised strategy, they used decision tree (J48), Naive Bayes (NB), support vector machine (SMO) and multilayer perceptron with back propagation (MLP) classifiers. For supervised, data set were created with feature vectors as shown in Fig. 5)

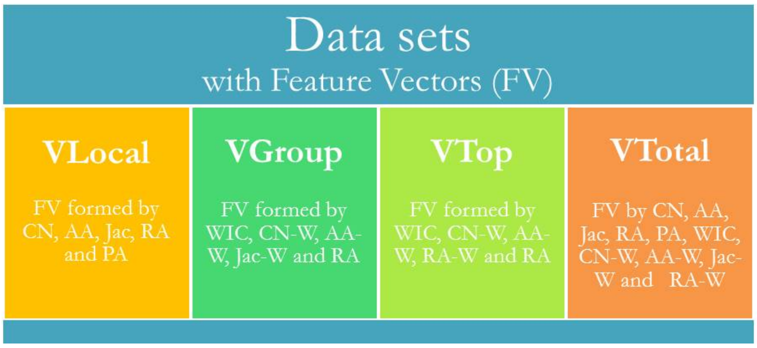


Figure 5: Data set with FV formed by differernt link prediction measures

The training and experiments were performed on a Linux machine with 99GB of RAM. The classification task was performed in Java and the rest of the processes were programmed in C++. In the unsupervised strategy, the execution time for each link prediction is shown in Fig. 6). It was observed that Jac-W, Jac and W form required more time to process the community information and PA was the fastest measure followed by WIC and CN.

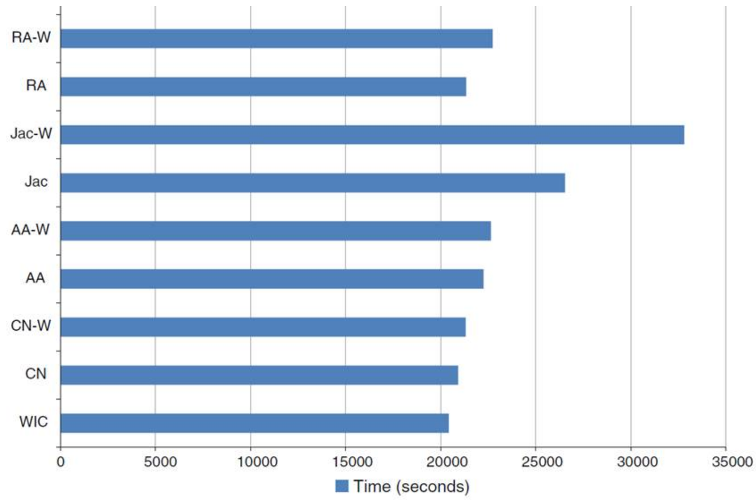


Figure 6: Execution time of all prediction measure
used in unsupervised strategy

4.3 Results and analysis

For validating the results, the authors used the evaluation measures that were presented before. For unsupervised strategy, the authors used AUC and Precision to evaluate the quality of link prediction measure. The AUC performance of each measure was same for both Twitter 7it and Twitter 15it subgraphs. It can be seen that that in the case of WIC and W form measures, the AUC seem to be same because of the similar behaviours of the nodes within the communities. From the AUC performance comparisons on all link prediction measures, WIC outperforms all of them which is followed by RA-W, RA, CN-W and AA-W. On the other hand, Jac has the worst performance. The results measured by AUC are tabulated in Table 4.

Graph	WIC	CN	CN-W	AA	AA-W	Jac	Jac-W	RA	RA-W	PA
Twitter 7it	0.62	0.56	0.59	0.53	0.58	0.45	0.56	0.6	0.61	0.51
Twitter 15it	0.62	0.56	0.59	0.53	0.58	0.45	0.56	0.6	0.61	0.51

Table 4: AUC measure results on two subgraphs of Twitter network

The results of the prediction quality measured by Precision on the two subgraphs are shown in Fig. 7). It can be seen that AA-W and RA-W has the best precision performance. Also, W form measures obtained good performance than their respective basic forms. PA has the worst performance compared to all the link prediction measures.

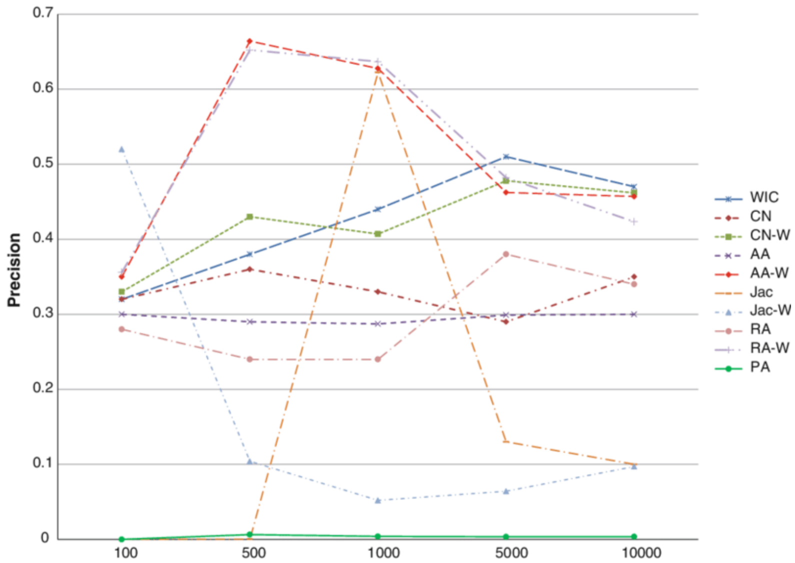


Figure 7: Precision results on the two graphs from Twitter network

For the supervised strategy on VLocal, VGroup, VTop and VTotal data sets, the authors used Accuracy and F-measure to validate the quality of the classifiers. From Table 5 (Left and Right), it can be seen that for J48, MLP and SMO classifiers, the best accuracy and F-measure were obtained on VTotal data set and for NB classifier, the best results were obtained on VTop data set in both the measures.

Data set	J48	NB	SMO	MLP
VLocal	83.74 (0.24)	71.63 (0.28)	81.35 (0.19)	82.73 (0.25)
VGroup	82.86 (0.25)	71.70 (0.28)	80.00 (0.20)	81.88 (0.26)
VTop	83.08 (0.25)	72.12 (0.28)	80.34 (0.20)	82.01 (0.26)
VTotal	83.80 (0.24)	72.05 (0.28)	81.71 (0.18)	82.84 (0.24)

Data set	J48	NB	SMO	MLP
VLocal	0.837	0.698	0.812	0.827
VGroup	0.829	0.699	0.798	0.819
VTop	0.831	0.703	0.801	0.820
VTotal	0.838	0.703	0.816	0.828

Table 5: Left: Accuracy results and Right: F-measure results

5 Conclusion

This paper of Valverde-Rebaza, J. and de Andrade Lopes, A. presents a new way of predicting links by including the community information and provides clear results and analysis for different strategies using different local prediction measures.

The authors show that the use of LPA for community detection and nodes' behavior into their communities improves the link prediction performance. When an unsupervised strategy was performed, the performance of WIC was better under AUC criterion and under precision, RA-W, AA-W and WIC outperformed other measures. When a supervised was performed, the results show that combining local information with the behavior information will improve the performance. Also, the experiments suggest that link prediction performance can be improved by WIC and W form measures as they use information of the behaviors of the nodes into their communities. These similar behaviors of the users into their communities may help in suggesting similar links, posts, tweets, pages and many more.

In the future work, this method can be improved to include users with more than 900 followers (large networks), one can try to employ similar method for undirected networks as this paper deals only with directed networks. Also, the prediction of suitable links is still in research and is a pretty popular research direction as it has huge impact on the future relationship of the users.

References

[1] Lunden, I.: Twitter passed 500M users in June 2012, 140M of them in US; Jakarta 'biggest tweeting' city. In: http://techcrunch.com/-2012/07/30/-analysttwitter-passed-500m-users-in-june-2012-140m-of-them-in-usjakarta-biggest-tweeting-city/,

(2012)

[2] Tang, J., Sun, J., Wang, C., Yang, Z.: Social influence analysis in large-scale networks. In: Proceedings of the 21st international conference companion on World Wide Web. In: Proceedings of WWW '12 Companion, pp. 607-608, (2009)

[3] Valverde-Rebaza, J., Andrade Lopes, A.: Structural link prediction using community information on twitter. In: Computational aspects of social networks (CASoN), 2012 Fourth International Conference on Nov 2012, pp. 132-137, (2012)

[4] Liben-Nowell, D., Kleinber, J.: The link-prediction problem for social networks. In: JASSIST 58(7): 1019-1031, (2007)

[5] Lü, L., Zhou, T.: Link prediction in complex networks: a survey. In: Phys A Stat Mech Appl 390(6): 1150-1170, (2011)

[6] Hasan M, A., Chaoji, V., Salem, Sj., Zaki, M.: Link prediction using supervised learning. In: Proceedings of SDM 06 workshop on link analysis, counterterrorism and security, (2006)

[7] Benchettara, N., Kanawati, R., Rouveirol, C.: A supervised machine learning link prediction approach for academic collaboration recommendation. In: Proceedings of RecSys Vol 10, pp 253-256, (2010)

[8] Fortunato, S.: Community detection in graphs. In: CoRR abs/0906. 0612v2, (2010)

[9] Feng, X., Zhao, J., Xu, K.: Link prediction in complex networks: a clustering prespective. In: Eur Phys J B 85(1): 3, (2012)

[10] Raghavan , UN., Albert. R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. In: Phys Rev E 76: 036106, (2007)

[11] Valverde-Rebeza, J., de Andrade Lopes, A., Link prediction in complex networks based on cluster information. In: Advances in artificial intelligence, SBIA 2012,

21th Brazilian symposium on artificial intelligence, ser, Vol 7589. Lecture Notes in Computer Science, Springer pp. 92-101 (2012)

[12] Bhat, AU.: Twitter community detection. Community detection for Twitter follower network. In: (Currently Not Available) https://github.com/AKSHAYUBHAT/TwitterCommunityDe-tection (2010)

[13] https://developers.google.com/machine-learning/crashcourse/classifica-tion/roc-and-auc