Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 48 (1) 81-84 (2016)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Efficient Link Prediction Method in Dark Networks Analysis

A.W. MAHESAR[++], Z. BHATTI*, A. WAQAS**, M. Y. KOONDHAR**, M. M.RIND**, S. NIZAMANI***

Department of Computer Science, Institute of Information and Communication Technology, International Islamic University Malaysia

Received 05th April 2015 and Revised 26th January 2016

Abstract: - The prediction of future links is very important problem in the field of complex network analysis. Due to the intricate interplay between nodes in dark networks the links prediction problem become very hard. As the dark networks evolve slowly with few nodes and gradually formed as well organized networks due to scale-free topology and therefore need efficient method of links prediction. In this paper, we apply two useful metrics of links prediction namely Jaccard's coefficient and Adamic/Aadr on the 9/11 dark network dataset. The results of both these metrics has been compared and it has been observed that Adamic/Aadr metric is more efficient method of link prediction in case of dark networks as compared to Jaccard's coefficient. The performance of both these methods is evaluated by using standard evaluation metrics which is precision.

Keywords: -Dark networks; Link prediction; Network analysis metrics, Jaccard's coefficient, Adamic Adar

1. INTRODUCTION

There are many examples of various complex systems and these can be formally represented through complex networks. The network is composed of two things namely nodes and links. The node represents the units in the systems and links shows the connectivity pattern between these nodes. From the last decade in complex networks research the research community use networks to model a large variety of complex systems or phenomena, to gain deeper insights into their working mechanism than otherwise possible. Therefore, the applications of network analysis can be found in transportation, biology, security, sociology, technology and many more (Newman,2010) (Easlesly and Kleinburg, 2010). In connection with this, dark networks are extensively modeled and analyzed due to the availability of many datasets (Bisharat, 2012) (Sarwat, 2011) (Waheed, 2015) (Fellman, 2015). A dark network (Terrorist Network) is based on actors whose function is to plan and execute any sort of terrorism or criminal activity. In dark networks all the nodes and links are not equal in terms of importance and behavior (Xu and Chen, 2008). These types of networks are also dynamic as the new nodes and links appear and disappear as network grows. This dynamic feature of dark networks makes these systems very difficult for the network analysts to accurately predict the future behavior. Links in between vertices of networks are very fundamental, but all connections are not equal in importance. For example, for two nodes with equal number of links, the node with more powerful links should be more important in network as compared to the one with relatively weak links. Further, due to intrinsically dynamic nature of dark networks, both the structure and the behavior of a network change overtime. These, characteristics of dark networks make the link prediction very difficult. The Link Prediction Problem for social networks was formally introduced by Liben-Nowell and Kleinberg (Nowell and Kleinberg, 2007) as a basic computational problem to understand the mechanism underlying social-network evolution purely based on endogenous properties of the given network i.e., properties derivable from its existing link structure. Following that many link prediction methods have been introduced based on different structural and topological properties of social networks. Therefore, research community in security has realized the role of SNA, motivated by increasing trends in formation of terrorist networks. In this paper, we discuss and highlight the importance of link prediction in dark network analysis and try to apply two mostly used link prediction methods in social networks. We have used the 9/11 terrorist attack dataset (Krebs,2001) and "R-project" open source software for network analysis.

The rest of this paper is structured as follows. Section 2, discuss and reviews the two metrics based on node neighborhoods namely Jaccard's coefficient and Adamic/Adr. In Section 3, we highlight the features and analysis of 9/11 dark network based on these two methods and their comparison. Finally, Section 4 concludes the paper with future work.

[++]Corresponding author : abdul.waheed@live.iium.edu.my
* Institute of Information and Communication Technology, University of Sindh, Jamshoro, Pakistan
**Institute of Information and Communication Technology, International Islamic University, Kuala Lumpur, Malaysia
***Department Of Computer Science, Mirpurkhas Campus, University Of Sindh, Pakistan

2. LINK PREDICTION METHODS

The real world networks are not static as they evolve by addition of new nodes and links as well as shrink with the deletion of nodes and links. Therefore, two parameters are used when evaluating link predictions in dynamic networks. They are $k_{training}$, $k_{testing}$ and a set known as Core which consists all those nodes which are incidents to $k_{training}$ links in G[to,to'] as well as at least ktest links in G[t1,t1']. The two methods discussed in this section assign score (x,y) known as connection weight. The score is assigned to nodes (x,y), on the basis of topology of given network. After that, a ranked list of score is produced in decreasing order of all nodes in the network. Thus, it can be viewed as finding of similarity of two nodes based on the topology of network. Now we discuss two methods of link prediction based on nodes neighborhood. In both these methods let suppose for any node x, ┌(x) represents the set of neighbors of node x in a network. These two methods assume that two nodes in a network namely x and y have more chances to create a link in future if their set of neighbors ┌(x) and ┌(y) has large overlap as shown in **(Fig 1)**.

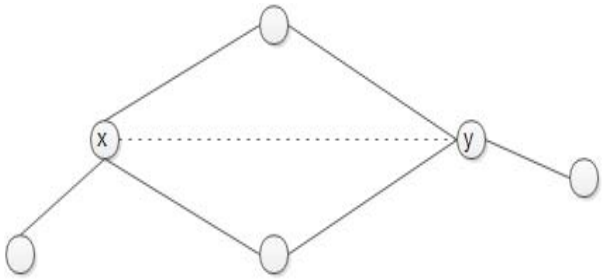

Fig.1. Overlapping nodes in network

Here, in **(Fig.1)** the two nodes x and y has many common neighbors as compared to other nodes in the network. Therefore, it follows the natural trend of creating new link in between two nodes x and y. For example, in case of routers network there are many chances of two routers to be connected in future if they both have many common nodes as neighbors. By using these intuition two methods namely Jaccard's coefficient and Adamic/Adr are explained below.

2 (a) Jaccard's coefficient

This method is based on common neighbors' concept which has been extensively used in information retrieval (Salton and McGill, 1984). This coefficient measures the probability of any two nodes x and y with assumption of any feature f, that can be in x or y nodes. Here, these features can be neighbors of nodes in a network then it can be formally defined as in **eq. (1)**

$$\text{Score}(x,y) = \left| \Gamma(x) \cap \Gamma(y) \right| / \left| \Gamma(x) \cup \Gamma(y) \right|$$

Here, in equation (1) the ∩ and ∪ are intersection and union of two sets respectively.

2 (b) Adamic and Adr

This metric was used for comparison between two web pages in World Wide Web network if they are very close to each other (Adamic and Adr, 2003). They defined the metric as in **eq. (2)**

$$\sum_{z:features\ sharedbyxandy} \frac{1}{log(frequency(z))}$$

Here, in this method the counting of rare features are used in weighting more heavily as compared to many. Therefore, it can be further defined as in **eq. (3)**

$$\text{Score}(x,y) = \sum z \in \Gamma(x) \cap \Gamma(y) \frac{1}{log\left|\Gamma(z)\right|}$$

In equation (3), the weighting scheme used is the inverse log frequency of their occurrence. For example, if only two people mention an item, then the weight of that item is 1/log (2) or 1.4. If five people mention the item, then its weight drops down to 1/log (5) or 0.62. This method evaluate the performance of the algorithm by computing the similarity score for all nods with respect to a single individual, and then ranking the nodes according to their similarity score.

3. A Look at 9/11 Network: (Features, Metrics, and Weights)

This network shows many features of dark/criminal networks. There are 62 nodes/vertices and 153 connections (Krebs, 2001). Although the density of this sample network is not high and there exists 8% possible conections/links. The reason behind that low density is quite obvious as dark networks try to focus on secrecy more as compared to efficiency (Krebs,2001). This is the reason of dark networks behaviour which shows the scale-free nature. As there are few nodes or actors with high number of links, and therefore random link failure does not effect on their functionality. In this scenario, the network has 4 clusters/groups and 19 major actors which are tightly linked with each other. Here, these 4 clusters/groups belongs to the persons who were present in the hijacking of US planes. This network has average degree distribution of 4.9 that means we can reach to each actor of the network from approximately five actors. Also, the metric diameter of the network is five. The names and the id of nodes are given in Appendix A.

3.1 Outcomes and Analysis

In order to test link prediction methods, we have used R-project open source software which is heavily used for network analysis. The Fig. 2 shows the R-project IDE. The visualization of dataset as a network is shown in **(Fig.2 and 3)** by using i-graph package of R-project.

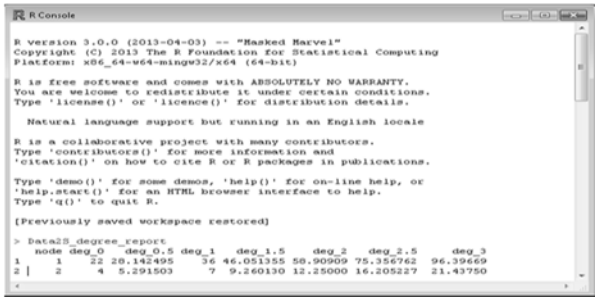

Fig.2. The R-project IDE used as simulation tool

In this network the total number of nodes are 62 with 153 links. The density of this network is very low which is 0.08. This shows that the network is sperse in terms of the number of links. In our experiment we removed 10% of links from the network randomly which became training set of the actual network. The training set has 138 links. The next step is to find testing set which comprises 14 missing links from the network. We applied both link prediction methods on the training set and the results of both methods are shown in (**Table.1 (a) and (b)**)

(a) All Nodes

Original Network	
Number of links	**153**
Density	**0.08**
Diameter	**5**

Largest connected components of network	
Number of links	**139**
Density	**0.07**
Diameter	**5**

(b) Precision values for both method applied namely Jaccard and Adamic Adr based on common neighbours

Method	**Precision**
Jaccard coefficient	**0.79**
Adamic Adr	**0.71**

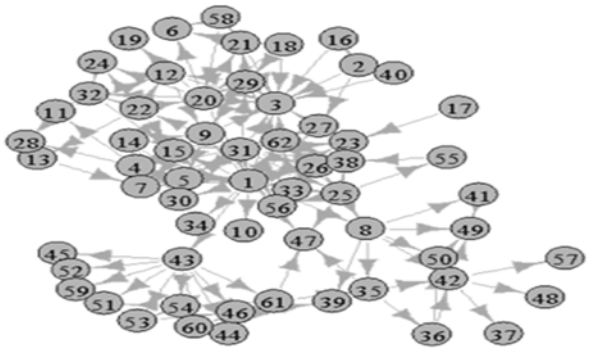

Fig.3 Network visualization of 9/11 network

The results of both these methods shows that Jaccard coefficient link prediction metric has outperform in case of dark network analysis. As this method consider the union and intersection ratio of neighbouring nodes as compared to the less number of neighbours.

Appendix:

The Table given below shows the IDs of actors/members and their respective names in 9/11 dark network [11].

NodeID	Name	NodeID	Name	NodeID	Name
1	Ziad Samir Jarrah****	26	Ramzi Omar	51	KamelDaoudi
2	Mohamed Atta*	27	Said Bahaji	52	Lased Ben Heni
3	RayedMohammed Abdullah	28	LotfiRaissi	53	MadjidSahoune
4	Hani Hanjour***	29	RaedHijazi	54	Mehdi Khammoun
5	Satam M. A. Al Suqami*	30	Salem Alhazmi***	55	Mohamed Bensakhria
6	Wail M. Alshehri*	31	ShaykhSaiid	56	Mohammed Belfas
7	AbdussattarShaikh	32	Marwan Al-Shehhi**	57	Mounir El Motassadeq
8	Fayez Rashid Banihammad**	33	SaeedAlghamdi****	58	NizarTrabelsi
9	HabibZacarias Moussaoui	34	ZakariyaEssabar	59	Osama Awadallah
10	AbdulazizAlomari*	35	AbdelghaniMzoudi	60	Samir Kishk
11	Ahmed Khalil Ibrahim Samir	36	Abu Qatada	61	Seifallah ben Hassine
12	Nabil al-Marabh	37	Abu Walid	62	TarekMaaroufi
13	Ahmed Alghamdi**	38	Abu Zubeida		
14	MohandAlshehri**	39	AgusBudiman		
15	Waleed M. Alshehri*	40	Ahmed Ressam		
16	Ahmed Ibrahim A. Al Haz****	41	Bandar Alhazmi		
17	Faisal Al Salmi	42	David Courtaillier		
18	Mamduh Mahmud Salim	43	DjamalBenghal		
19	MajedMoqed***	44	Essid Sami Ben Khemais		
20	Mohamed Abdi	45	EssoussiLaaroussi		
21	NawafAlhazmi***	46	Fahid al Shakri		
22	Khalid Almihdhar***	47	Haydar Abu Doha		
23	HamzaAlghamdi**	48	ImadEddinBarakat		
24	MamounDarkazanli	49	Jean-Marc Grandvisir		
25	Ahmed Alnami****	50	Jerome Courtaillier		

American Airline Flight 11(Crashed into WTC 1)** * American Airline Flight 77 (Crashed into the Pentagon)**
****** United Airline Flight 175 (Crashed into WTC 2)** ***** United Airline Flight 93 (Crashed in Pennsylvania)**

5. CONCLUSIONS

In case of dark networks analysis like homeland security, Drug trafficking gangs, human trafficking groups and terrorist organizations etc. the correct link prediction plays very important role for the network analysist. Now a dyas national security has become very important area of reaserch mainly after the terrorism activities around the globe. In this paper, we have analyzed the dark network by using two link prediction methods. The results satisfactorily shows that the nodes nearest neighbors are important to be focused for the more accurate link prediction in such type of networks. The limitation of this analysis is the lack of different dark networks dataset and their availability. We believe that if this network grows in this pattern then it may follow rich get richer phenomenon and can be best predicted by jaccard coefficient. There are many other link prediction in methods available like, sim rank, preferential attachment, common neighbour,hitting time and commute time etc., and these may constitute future work with few other data set of dark networks.

REFERENCES:

Easley, D., and J. Kleinberg, (2010). Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press.

Salton, G., and M. J. McGill, (1986). Introduction to modern information retrieval. McGraw-Hill

Mahesar, A. W., A. Waqas, N. Mehmood, A. Shah and M. R. Wahiddin, (2015). Analyzing the weighted dark networks using scale-free network approach. WSEAS Transactions on computers, 14(1):748-759.

Memon, B. R. (2012). Identifying important nodes in weighted covert networks using generalized centrality measures. In European Intelligence and Security Informatics Conference (EISIC), 131-140), Greece.

Newman, M. (2010). Networks: an introduction. Oxford University Press, Oxford..

Nizamani, S., and N. Memon, (2011). Evolution of Terrorist Network Using Clustered Approach: A Case Study. In 2011 European Intelligence and Security Informatics Conference 116-122, Denmark.

Krebs, V. E., (2002). Uncloaking terrorist networks, First Monday, Vol. (7) 4–11,

Adamic, L. A., and E. Adar, (2003). Friends and neighbors on the web. Social networks, Vol.25 (3), 211-230.

Liben-Nowell, D., and J. Kleinberg, (2007). "The link-prediction problem for social networks. "Journal of the American Society For Information Science And Technology Vol. 58 (7), 1019-1031.

Xu, J., and H. Chen, (2008). The topology of dark networks. Communications of the ACM, Vol. (10), 58-65.