Sindh Univ. Res. Jour. (Sci. Ser.) Vol.48 (4) 845-848 (2016)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Complex Network of Dengue Epidemic and Link Prediction

H. A. M. MALIK++, A. WAQAS*, F. ABID**, A. R. GILAL*, A. W. MAHESSAR, Y. KOONDAR

Department of Computer Science, KICT, International Islamic University Malaysia

Received 13th June 2015 and Revised 2nd November2016

Abstract: In complex network of epidemic diseases, link prediction can play a significant role especially for the policy makers so that they can take precautionary measures timely. Furthermore, it is important to determine which prediction method is more appropriate for some particular dataset and scenario. Dengue epidemic is a complex and dynamic phenomenon which has gained much attention due to its harmful effects. In this study, multifarious network of *Aedes Aegypti* (dengue vector) is addressed from the viewpoint of complex network and, three link prediction methods are applied on the dengue epidemic dataset of Selangor, Malaysia. Experimental results showed that Jaccard's metric is more efficient method for link prediction of such datasets as compared to other methods.

Keywords: Dengue epidemic network; Complex Networks; Link prediction; Clustering coefficient;

1. INTRODUCTION

In the last decades complex networks have been a main focus of research in the form of network structures As a matter of fact; different domains like technological, transportation, biological, networks have been modelled and analyzed as complex networks (Newman, 2009). It is very difficult to analyze the complex systems as these are huge in number of interacting elements and patterns between the elements. By converting these elements into complex network becomes convenient for efficient analysis. In such networks, nodes depict elements of the systems and links represent the intermingling patterns between these elements. Dengue epidemic is a vital health issue that has been exponentially increasing in the Malaysia. It is highly affecting the community with every passing year and has become a fatal disease (Schapira, *et. al,* 2012). Problem of dengue epidemic network have modelled and analyzed in Selangor, Malaysia by using two-mode network technique (Malik, 2014), (Malik, 2016). Dataset has been obtained by the Ministry of Health Malaysia (Oct 2013 to 2014) (MOH, 2014).The dataset contains the week-wise number of dengue cases from different localities in Selangor, Malaysia. Importance of link prediction in dengue epidemic network is discussed and three link prediction methods are used in this paper. With the help of link prediction, it can be projected that what new connections may form in future on the basis of network's current ceremonial.

1.1 Link Prediction in Socio-geographic Networks

Many recommendation tasks are based on link prediction, for examples friend recommendation, merchandise endorsement and commercial adds. Link prediction phenomenon is likewise identified with the issue of deducing absent links from an experiential complex network: in various spaces, someone builds links of interactions in view of noticeable data and later attempts to deduce the added links that are not straightforwardly obvious, are expected to exist **(Fig-1)** (Liben-nowell, 2003), (Liben-nowell, 2007). In fact, link prediction issue examines about "to what degree can the advancements of a network be demonstrated utilizing highlights natural for the network itself"? For example, ponder on the network of co-authorship among researchers, in which numerous reasons are in that link. Some researchers who have never contributed in a joint paper, why and how they will do as such in the following couple of years. They may come closer geologically if anyone of them quits the organization. Like these combined efforts can be difficult to anticipate. Two researchers who are close in links will have associates in like manner, and will go in comparative circles; this vicinity recommends that they themselves will probably work together sooner. This natural idea will probably exact, to comprehend the procedures contiguity in a system that move to the most precise link prediction (Sharara, 2010). Numerous vicinity measures guide and lead to predictions that outpace by elements, showing the network layout certainly having inactive data by that upcoming collaboration and associations can be persuaded.

1.2Dengue Epidemic

In most of the tropical andsub-tropical regions of

++Corresponding author: H. A. M. MALIK

*Department of Computer Science, Sukkur Institute of Business Administration, Pakistan

**Department of Computer Science, King Khalid University, Kingdom of Saudi Arabia

the world, dengue epidemic is a foremost ailment. According to an estimation, 3.6 billion human population is living in the dengue affected part of the world(Eduardo, 2012).World is in danger due to dengue ailment and Dengue Hemorrhagic Fever (DHF) (William, 2014). World Health Organization (WHO) reported that throughout the globe it may appear fifty to one hundred million dengue infections annually (Eduardo, 2012). Dengue ailment is an arbovirus that is caused by mosquito named as *Aedesaegypti* and *Aedes albopictus*. These are the super spreader vectors of this viral that are transmitted in human by its bite and symptoms appear after 3 to 14 days of the bite. Dengue Virus (DENV) 1, 2, 3 and 4 are different serotypes that have been found in this disease(William, 2014). These were detected from the infected person's blood samples. Any person who is infected by any one of these serotypes will never be infected again by the same serotype (Coutinho, *et. al,* 2006).According to the report, on 21 November 2015, there were 107,079 dengue cases with 293 deaths reported in Malaysia. Ministry of Health (MOH) Malaysia reported in July 2012, dengue incidence rate was 63.75% and Dengue Hemorrhagic Fever was 4.90%[4].Dengue viral is increasing in Malaysia; a trend is given from 1995 till 2014 in **(Fig-2)**. Purpose of this study is to identify which link prediction method is more appropriate for this dengue epidemic dataset.

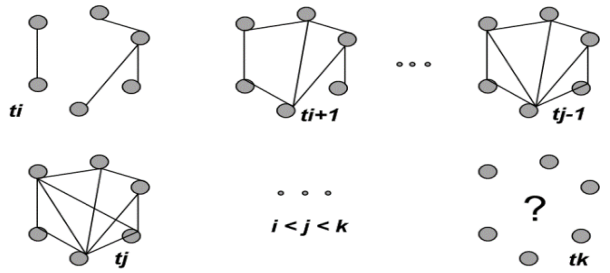


Fig.1: link prediction of missing links

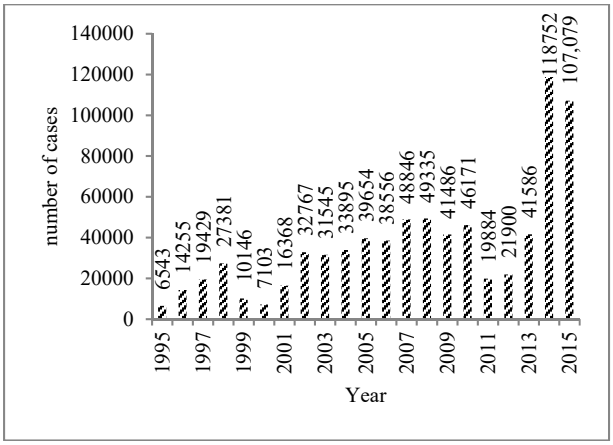


Fig-2: Dengue cases in Malaysia from 1995 till 2015(Source WHO)

2. RELATED WORK

There are various techniques that are being used for link prediction in networks. A great part of the work on this issue depends on structural proximity. The study has contemplated an arrangement of both local vicinity procedures and global vicinity procedures. The local vicinity procedures incorporates preferential attachment (Barabasi, 2003), Adamic-Adar, triadic closure looks at the area where neighbor of two nodes occur. The global proximity measures focuses on the indirect links between nodes of a network. For example, number of paths is counted between nodes in Katz score (Zhu,2014). Some current works additionally display the connection expectation issue as a regulated grouping errand, where every information guide relates toward a couple of vertices in the interpersonal organization chart. In existing study on link prediction, most of the characteristics are an arrangement of structural vicinity procedures that are removed from the graph layout. In the framework, links established not only on node similarity (e.g., homophile) but also because of the information flow and influence (Zhu, 2014). From the previous study it has been observed that how information dispersal affects follow link establishment. The elementary indication by this study is that information dispersal viaretweets sources new follow links that generate a short way to the information source. So, by the inspiration of these works, in this paper we utilize the two link prediction methods on our dengue epidemic dataset and compare the results with the (Mahesar, *et. al,* 2016)

3. EXPERIMENTAL SETUP AND PREDICTION METHODS:

The actual world complex systems are not fixed 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. Assume that in a complex network set $G = (V, E)$ where every edge $e = (u, v) \in E$ signifies an interface between u and $v$which appears on a fixed time t(e).It calculates various interfaces between u and v as equivalent edges, with hypothetically dissimilar time-stamps. For two times $t < t'$, say$G[t, t']$ represent the sub-graph of Gcontaining all ends with a time-stamp between t and t'. Now, it is a solid preparation of the link-prediction issue. Four times $t_0<t_0^{'}<t_1<t_1^{'}$are selected and provide an algorithm reaches to the system$[t_0,t_0^{'}]$; it should produce a list of edges not existing in $G[t_0,t_0^{'}]$ which are forecasted to appear in the network $G[t_1,t_1^{'}]$. It is refered to$[t_0,t_0^{'}]$ as the $k_{training}$ *interval* and $[t_1,t_1^{'}]$ as the k_{test} *interval*.(Sharara,2010).

The three 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. For both methods, let suppose for any node x, ┌(x) represents the set of neighboring nodes of x in a network. These two methods assume that two nodes in a network namely x and y have much chances to create a link in upcoming time if their set of neighbors ┌(x) and ┌(y) has huge overlap as shown in **(Fig. 3)** where, the nodes x and y has many common neighbors as compared to other nodes in the network (Mahesar, *et. al,* 2015). 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 (Mahesar, *et. al,* 2016). By using these intuition three methods namely common neighbors, Jaccard's coefficient and Adamic/Adr are explained below.

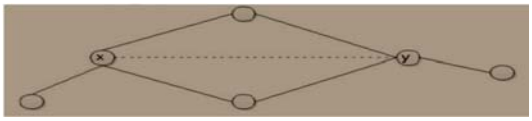


Fig.3: Overlapping nodes in network

3.1 Common neighbors

In (1), the numeral of neighbor that *x* and *y* have in common are the direct execution of this idea for link prediction. Newman (2003), has figured this measure in the perspective of network association, authenticating a relationship between the numeral of common neighbors of *x* and *y* at time *t* and the possibility that they will work together in the upcoming time.

$$score(x, y) = |\Gamma(x) \cap \Gamma(y)| \qquad (1)$$

It represents the possibility of scientists collaborating upsurges with the number of other common coworkers.

3.2Jaccard's coefficient

This method is based on common neighbors' concept which has been extensively used in information retrieval. 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 (Liben-nowell, 2007). Here, these features can be neighbors of nodes in a network then it can be formally defined as in eq. (2)

$$score(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \qquad (2)$$

Where,∩ and ∪ are intersection and union of two sets respectively.

3.3 Adamic and Adar

This metric was used for comparison between two web pages in World Wide Web network if they are very close to each other (Adamic, 2003). The neighbors not shared with many others, it gives more weight.

$$score(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(y)|} \qquad (3)$$

In this model the likelihood depends on the quantity of mutual fellows, mitigated by the log of the number of a neighbor. The degree *deg(z)* is the total of the 'in' and 'out' degree of the node (network followers).

$$Pf(u, v) = \sum_{z \in (friends(v) \cap followers(u))} \frac{1}{\log(\deg(z))} \qquad (4)$$

Inverse log frequency of their occurrence is the weight procedure. For instance, the weight of an item will be 1/log(2)=1.4 if only two people mention that item. Moreover, in case of five persons if they comment on an item, then it loses its weight to 1/log (5) = 0.62. By this the performance of the algorithm is evaluated by computing the resemblance score for all nods with respect to a single individual, and then ranking the nodes according to their resemblance score.

4. RESULTS AND ANALYSIS

In order to test link prediction methods on dengue epidemic dataset, we have used R-project open source software which is heavily used for network analysis. The visualization of dataset as a network is shown in **(Fig-4)** by using i-graph package of R-project. In this network the total numbers of nodes are 58with 574 links. Density value of this network is 0.521. It represents the network as dense, in terms of the number of links. In our experiment to get the training set of the real network we remove the 15% links randomly from the network. The training set has 488 links. The next step is to find testing set which comprises 34 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 and 2)**

Table-1

Real network	
Number of links	574
Diameter	2.5
Density	0.521

Table-2

Largest connected components of network	
Number of links	488
Diameter	2.5
Density	0.49

Table 3: Precision values for applied method

Methods	Precision
Common neighbor	0.56
Jaccard Coefficient	0.81
Adamic / Adr	0.76

The results of all three methods shows that Jaccard's coefficient link prediction metric has outperformed in case of dengue network analysis. As this method consider the union and intersection ratio of neighboring nodes as compared to the less number of neighbors.

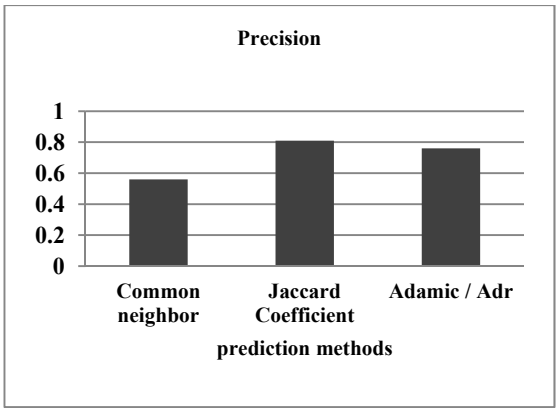


Fig.4: comparison of link prediction methods

4.1. Visualization

Fig-5is the graphical visualization of the dengue epidemic network that appeared in Selangor, Malaysia. It is generated by R-project tool.

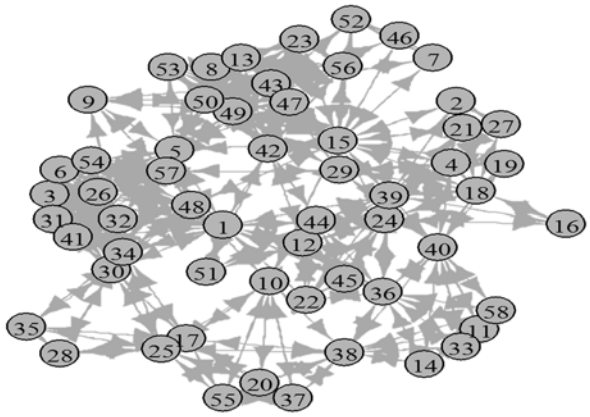


Fig. 5: Visualization Of The Network

We have observed this network from global and local structures perspectives. The analysis of this dengue network has shown many interesting results and has proved that dengue epidemic network follows a scale-free network. This network is robust under random attack, means when links are removed randomly. On the other sides, it shows the vulnerable behavior under targeted attack. From (**Fig. 5)** focal nodes can be seen which acts like super spreaders of the dengue virus.

5. CONCLUSION

In this paper, we have analyzed the dengue epidemic network where three methods of link prediction are utilized. After an analysis of comparison among the applied methods it has been observed that Jaccard's coefficient is producing the satisfactory results. The result by this shows that the node's nearest neighbors are important to be focused for the more accurate link prediction. It is believed that if this network grows in this pattern then it may follow rich get richer phenomenon and can be best predicted by jaccard's coefficient. Some other link prediction methods may also be applied in future such as, Katz, SimRank, rooted PageRank, hitting time and preferential attachment.

REFERENCES:

Adamic, L. A., E. Adar, (2003). "Friends and neighbors on the Web," vol. 25, no. 3, 211–230.

Barabasi, A. L., E. Bonabeau, (2003). "Scale free networks," Sci. Am., no. 50–59.

Coutinho, F. A B., L. F. Lopez, E. Massad, (2006). "Threshold conditions for a non-autonomous epidemic system describing the population dynamics of dengue.," Bull. Math. Biol., vol. 68, no. 8, 2263–82.

Eduardo, J., M. Pessanha, (2012). "Risk assessment and risk maps using a simple dengue fever model," Dengue Bull., vol. 36, 73–86.

Liben-Nowell, D., J. Kleinberg, (2003). "The Link Prediction Problem for Networks," Proc. Twelfth Annu. ACM Int. Conf. Inf. Knowl. Manag., 556–559.

Liben-nowell, D., (2007) "The Link-Prediction Problem for Networks," vol. 58, no. 7, 1019–1031.

Mahesar, A. W., A. Waqas, N. Mehmood, (2015). "Analyzing the Weighted Dark Networks using Scale-Free Network Approach," vol. 14, 748–759.

Mahesar, A. W., Z. Bhatti, A. Waqas, (2016). "Efficient Link Prediction Method in Dark Networks Analysis," Sindh Univ. Res. J. (Science Ser., vol. 48 (1), 81–84.

Malik, H. A. M., A. W. Mahser, (2014). "Two-Mode Complex Network Modeling of Dengue Epidemic In Selangor, Malaysia," in The 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M), 1 – 6.

Malik, H. A. M., A. W. Mahesar, F. Abid, A. Waqas, M. R. Wahiddin, (2016) Two-Mode Network Modelling and Analysis of Dengue Epidemic Behaviour in Gombak, Malaysia. Applied Mathematical Modelling. DOI: 10.1016/j. 10.060

MOH, (2014) "Official Portal for Ministry of Health, Malaysia,".

Newman, M. E. J., A. Arbor, (2009). I. Introduction, and A. Books, "Complex Systems: A Survey," no. I. 457-460.

Schapira, A., C. Garcia, M. Asor, I. Riquentino, S. Oguan, E. Olitan, A. Quitasol, S. Rabano, B. Tapit, (2012). "Baseline and key container survey for *Aedes* aegypti and Aedes albopictus in Albay Province, Philippines," Dengue Bull., vol. 36, 182–263.

William, T., J. Menon, (2014). "A review of malaria research in Malaysia," Med. J. Malaysia, vol. 69, 82–87.

Zhu, L., K. Lerman, (2014). "A Visibility-based Model for Link Prediction in Media," Proc. ASE/IEEE.