Analysis and Generation of Trust Networks

Avinash V[1], G Manasa[2], G Sameer[3]

Abstract— **Trust is known to be the key component in human social relationships. It is the measure which can help us behave accordingly with someone. So,since the recent times these networks are being studied, thereby grabbing a lot of scope for research and depth. We, in this paper will analyze two well-known social networks called soc-bitcoin network and the wiki administrator network. Both of these help us with different scenarios to understand how trust networks can be combined with generative models. We try to generate graphs from these networks and compare them with the original graphs to see how trust is a good measure.**

I. INTRODUCTION

Trust networks are a special kind of social networks where there will be positive and negative edges which will represent the trust and distrust between two people. These kinds of networks as we know play an important role in the recent social networks which are very volatile in terms of human behaviour. Therefore, all the interest in social network analysis now-a-days in transiting towards these kinds of networks. The weights of the edges remarkably provide the best intuition of bond strength between two nodes/people in a simple manner.

Generative models in social networks are pieces of experimentation which was done to verify if the measures and the standard patterns of networks studied were right. In these models, edges are produced based on some probability measures and in the end the whole network graph is generated to see if it actually captured the characteristics of real world networks along with their structural dependencies.

We, in this paper on a whole try to generate the graph using the trust given as weights of edges in the mentioned datasets. Using parameters like number of triads along with their type are compared with those from the original graph to generate the accuracies and error measures. Above all, there is an algorithm which has been followed to generate the trust network which is based on a very simple intuition which is explained in the methodology section.

*This work was supported by BhaskarJyotiDas, PES University

[1] USN: 01FB15ECS061, Department of Computer Science PES University, Bangalore, India email at vavinash0497@gmail.com

[2] USN: 01FB15ECS099, Department of Computer Science PES University, Bangalore, India email at manasapriya17@gmail.com

[3] USN: 01FB15ECS101 Department of Computer Science PES University, Bangalore, India email at sameergadicherla56@gmail.com

II. RELATED WORK

In all the real world applications of social networks, it is seen that the measures used for measuring the correctness of generated graphs are centrality measures like betweenness and closeness centrality. But it has been harder to visualize graphs with these measures as key roles in any network now-a-days are at least more than one and also the decentralized scattered networks are more than before. To be pin point in this case degree distribution, the diameter between two nodes, clustering coefficients have been brought up. But as said, trust is another important factor to be worked on, it has also been used in few applications like betrayal detection, fraud detection etc.. In further sections, we will show how this trust changes over person to person.

III. DATA COLLECTION

We have used two datasets from the SNAP dataset:

1) Bitcoin OTC trust weighted signed network
This is who-trusts-whom network of people who trade using Bitcoin on a platform called Bitcoin OTC. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. Members of Bitcoin OTC rate other members in a scale of -10 (total distrust) to +10 (total trust) in steps of 1. This is the first explicit weighted signed directed network available for research. No. of nodes = 5,881 No. of edges = 35,592 Weights in the graph = -10 to +10

2) Wikipedia Requests for Adminship (with text)
For a Wikipedia editor to become an administrator, a request for adminship (RfA) must be submitted, either by the candidate or by another community member. Subsequently, any Wikipedia member may cast a supporting, neutral, or opposing vote. This induces a directed, signed network in which nodes represent Wikipedia members and edges represent votes. No. of nodes = 10,835 No. of edges = 159,388 Weights in the graph = 1,-1,0

IV. METHODOLOGY

For the bitcoin dataset, the weights are the ratings given to users by other users. This range of -10 to +10 was converted to 1 or 0 indicating positive and negative edges. Similarly for the wikipedia dataset was also converted. The data was preprocessed and extracted into a dataframe. We converted the data from csv and text files into a GML format graph.

The fraction of positive edges in the wikipedia dataset is 0.735660237585 and the fraction of positiv edges in the bitcoin dataset is 0.927111145015. We can see the difference between the datasets through this. We have taken 70 percent of the data as the training graph and tried to generate the rest of the graph. This division into training and testing data is done according to chronological order of edge creations.

The intuitive idea behind the model is that when a user i joins the network, he begins interacting with other users. If their interaction with a user j is positive, then over time i discovers the users that j knows. If i trusts j then i will trust anyone that j trusts. However, if i trusts j, he wont automatically start distrusting everyone that j distrust. However, the likelihood that i will trust user k will go down to 50 percent in our model. If j distrusts k.

On the other hand, if i does not think highly of j (i.e. if (i; j) is negative), then he will automatically form negative edges with everyone that j distrusts. However, i will disregard people that j trust because i doesnt have a high opinion of j to begin with. This model is based in spirit on the theory of status, although there is an asymmetry in creating edges to friends of a user that you dont trust versus the enemies of your friend. The probability rp determines the fraction of neighbors of j that i discovers given that (i; j) is positive whereas the probability rn determines the fraction of neighbors of j that i discovers given that (i; j) is negative. We impose the constraint on the model that rp rn which intuitively corresponds to the fact that i is much more likely to interact with js neighbors if (i; j) is positive.

The parameter 'p' is the probability of positive edges. The parameter 'rp' is fraction of neighbours of j, that i discovers given that (i,j) is + . The parameter 'rn' is the fraction of neighbours of j, that i discovers given that (i,j) is - . The parameter p along with the parameter rp controls the fraction of edges in the generated graph that get labeled +, whereas rp and rn control the number of triads and their relative frequencies.

The algorithm is as shown in fig.1

V. Visualization

For visualization, we will look at the training graph used and the one generated by the algorithm. As we can see the training graph used in fig.2 and the generated graph in fig.3, the number of edges in training graph is about 70 percent of the total edges of the original graph.

We will also plot the log-log graphs The log-log plot of predicted graph is as shown in fig.4 and the original graph in fig.5

VI. Results

The statistics for bitcoin dataset are as shown in fig.6. The triadic statistics for bitcoin dataset are as follows:

Fig. 1. The algorithm developed

The algorithm:

```
for i = 1 to n do
    Pick a node j uniformly at random from {1,2....i-1}
    Create edge (i; j)
    Label the edge + with prob. p, and - with prob. (1 - p)
    for all neighbors k of j do
        if (i; j) is + then
            Create edge (i; k) with prob. rp.
            if (j; k) is + then
                Label edge (i; k) +.
            else
                Label edge (i; k) + or - with prob. 1/2.
            end if
        else
            if (j; k) is - then
                Create edge (i; k) with prob. rn.
                Label edge (i; k) -.
            end if
        end if
    end for
end for
```

Fig. 2. The training graph used

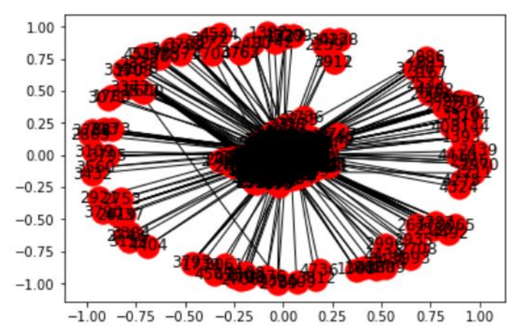
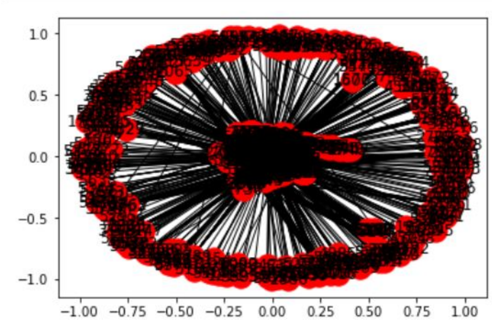


Fig. 3. The algorithm generated



Triad Type	Predicted	Original
003	0.9934164881	0.9963198425
102	0.006520572995	4.7100852670e-05
201	6.191982586e-05	4.7100852670e-05
300	1.0189866226e-06	9.884923990e-07

The accuracy of the prediction algorithm was calculated using Root Mean Square and R squared values. The bitcoin dataset was found to have an RMSE of 0.0010238758096072877 and r squared value of 0.999981967585. The wikipedia datset was found to have

Fig. 4. Log-log plot of predicted graph

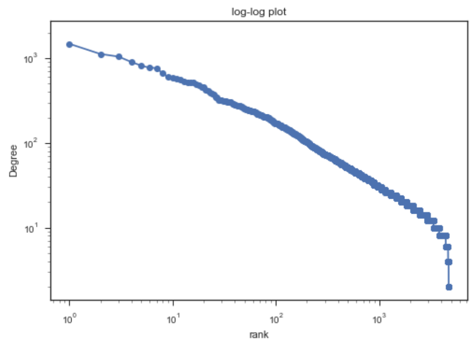
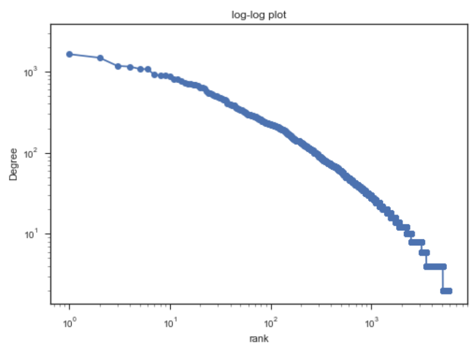


Fig. 5. Log-log plot of original graph



an RMSE of 0.0009504283652480449 and r squared value of 0.999984307064.

Fig. 6. Scores by triad types and no. of traids

VII. CONCLUSIONS

Looking at the analysis done in the previous sections, we can conclude that the algorithm has reached all the requirements and has potrayed similar characteristics of both the original graphs. We can see that in generative models, trust when used as the key component has provided the most accurate results over time. The dataset which has been split accordingly with time has given out its remarkable predictions over the model trained which resembled the same original network. We can see that in bitcoin network only one type of triad has been dominant which represents the similar nature of bonding via transactions by the peers. Where as in the wiki dataset, may be one type of triad is dominant but all the other three are somewhat near to its number, which says the possible types of relations in this is more varied.

REFERENCES

[1] T. Antal, P. L. Krapivsky, and S. Redner. Dynamics of social balance on networks, 2009.
[2] Wenjun Jiang a,b, Guojun Wanga,, Jie Wub Generating trusted graphs for trust evaluation in online social networks
[3] A. Jsang, R. Hayward, S. Pope, Trust network analysis with subjective logic, in: Proceedings of the 29th Australasian Computer Science Conference, ACSC2006, January 2006, pp. 8594.
[4] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. Science, 286:509512, 1999.
[5] Robert West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts: Exploiting Social Network Structure for Person-to-Person Sentiment Analysis. Transactions of the Association for Computational Linguistics, 2(Oct):297310, 2014.
[6] S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos. Edge Weight Prediction in Weighted Signed Networks. IEEE International Conference on Data Mining (ICDM), 2016.
[7] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In WWW 04: Proceedings of the 13th international conference on World Wide Web, pages 403412, New York, NY, USA, 2004. ACM.