Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 125 (2018) 609–616

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India

Anomaly Detection in Multiplex Networks

Ruchi Mittal*[a], M.P. S Bhatia[b]

[a]*Research scholar, Netaji Subhas Istitute of Technology, New Delhi-110075, India*
[b]*Professor, Netaji Subhas Istitute of Technology, New Delhi-110075, India*

Abstract

Detecting anomalies in social is a vital task, with numerous high impacted social networks such as WWW, Facebook, Twitter and so on. There are multiple of techniques have been developed for detecting outliers and anomalies in graph data. More recently, the area of multiplex networks has extended a considerable attention among researchers for more concrete results. A Multiplex network is a network, which contains multiple systems of the same set of nodes and there exists various types of the relationship among nodes. In this paper, we discover the anomalies across numerous multiplex networks. By anomalies or outliers means nodes, which behave abnormal or suspicious in the system. Compared to single layer networks, the outliers' nodes may found into many layers of the multiplex network and find anomalies in the multiplex network is still untouched. From this study, we propose a new metric called cross-layer anomaly detection (CAD). The CAD is a measure, which detects the anomalies in the multiplex network. For experiments, we make use of two real-world multiplex networks. We compare the results of our proposed metric with other similar methods, and we get encouraging and similar results.

Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications.

*Keywords:*Multiplex Network; Anomaly Detection; multiple Layers; Edges; Nodes;

* Corresponding Author.
Email: ruchi.mittal138@gmail.com

1877-0509
Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications
10.1016/j.procs.2017.12.078

1. Introduction

It is interesting to find out the nodes, which shows the least appearance in a given network. A network is nothing but a collection of nodes joined together by edges [1]. Many real-life applications consist of network type structure where individuals act as nodes and relationship among them serves as edges. Finding individuals, which holds suspicious behavior or have the least appearance of the network is an exciting task and called as anomaly detection. For example, detecting networks intrusion or network failure [2], credit card fraud [3], telecommunications fraud [4] and so on.

To all networks, there is an assumption that only one type of interaction exists between any two nodes. Sometimes this assumption may ignore the multiple interactions or relationship among entities of the network, which doesn't lead to refined results. A system formed by various interactions among entities is called a multiplex network [5, 6]. A multiplex network contains multiple layers; in which each segment represents one type of communication among entities. For example: in social networks, similar entities interact with others via Facebook, Twitter, LinkedIn and so on. Existing graph theory concepts cannot be directly applied to such networks, as there exist cross-layer interaction between entities. In Fig. 1, we show a sample multiplex network with two layers, where two type of communication between nodes exist.

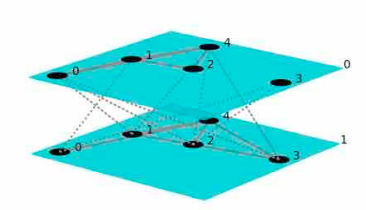

Fig. 1. A Sample multiplex network with two-layerarchitecture.

The combined concepts of data mining and graph theory are widely used to study the various features of the network. For example, using anomaly detection methods, one can find the suspicious nodes in the network. Anomaly detection is a branch of data mining, which is concerned with discovering rare occurrences in datasets [7]. In general, data objects are inter-related, and one can easily map this into graphs or networks. The multiple paths lying among two data object may efficiently capture the long-range relationship and helps to spot the suspicious object in the network.

There are various methods defined for anomaly detection in networks such as distance-based anomaly detection [8], density based anomaly detection [9], distribution-based anomaly detection [10], clustering-based anomaly detection [11] and so on. Depending on the application, one can apply the anomaly detection method to find out the outlier nodes. For example, Clustering based method first find out the cluster in the network and then find out the nodes, which have least interaction.

In this paper, we propose a methodology for anomaly detection in the multiplex network. Here, we used the famous Gaussian model [12] along with other defined features of a multiplex network to detect anomalous nodes in the network. The formulation for Gaussian model for the simple network is as follow:

$$p(x;\mu,\sigma^2) = {}^{1}\!/_{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

Here, $x = \{x^1, x^2, .., x^m\}$ is the training set, μ is the mean and σ^2 is the variance. In figure 2, we present a sample network, in which nodes inside the red circle are the anomalous nodes of the network.

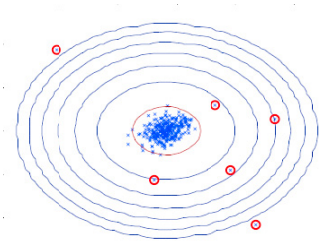

Fig. 2. Example of anomaly detection in simple network

In this paper, we propose an algorithm for detection anomalous nodes in multiplexed networks. We first find out the eigenvector centrality, page rank centrality and the degree centrality of a given multiplex network using the muxViz tool [13]. Next, we create a training set using the output of all the centrality measures. Next, we propose a methodology for anomaly detection. Here is a fundamental outlook of our proposed approach, which we discuss further in detail in section III. We validated our proposed algorithm on two multiplexed networks: Danio Rerio network and Florentine families network [14, 15]. To these datasets, our algorithm detects the anomalous nodes with comparably limited computational overhead.

Paper Outline: we present related work in the area of multiplex networks, centrality measures and anomaly detection in Section II. We discuss our proposed approach for detecting anomalies in the multiplex network in section III. In section IV, we present essential characteristics of datasets, which we use to validate our algorithm. In section V, we discuss our findings.

2. Related Work

Boccaletti et al.[16] discuss the various characteristics of the multiplex networks. These characteristics belong to the design and type of the network. Battiston et al. [5] analyze the structural properties of the multiplex network. These structural properties include clustering coefficient, vertex degree and so on of the networks.

Centrality is a measure to find out the importance of a node in a given network. Centrality measurement is one of the prominent areas of research in the multiplex network. Many researchers redefine centralities of the simple network to multiplex networks. Sola et al. [18] propose a new formulation for betweenness centrality. They also propose a new algorithm for finding the shortest path in the multiplex network. Tanmoy et al.[19] propose a fastest betweenness centrality algorithm for multiplex network compare to Sola betweenness centrality.

Sun et al. [20] proposed a graph-centric, i.e., community-based anomaly detection method. Gyongyi et al. [21] proposed an anomaly detection algorithm for web network. It finds out the spam and malware in a given web network. Li et al. [22] suggested the algorithm for the financial trading network. They use graph-based sub-structure to find out the outliers in the system. De Domnico et al. [13] create software called MuxViz for the multiplex network. This software has an interactive GUI, and numerous of algorithms of multiplex are well implemented in it. Which are used for analysis of the multiplex network.

3. Anomaly Detection in Multiplex Networks

There are numerous of algorithms exist for detecting anomalies in simple networks. Here, our work is directed on spotting anomalous nodes in the multiplex network. Our proposed algorithm starts with finding Eigenvector centrality, page-rank centrality and degree centrality defined for the multiplex network for each node. These centralities present the importance of a node in the given network. Next, we create a training set using the output of each centrality measures. To this training set, we fit a Gaussian distribution and then find the value that has very low probability. These low probability values nodes are considered as anomalous nodes. Here is an informal description of our proposed approach. We describe our proposed method in following steps.

3.1. Find Probability Set in Multiplex Network

a) Let's assume a graph G = (V_m, E_m, L_m), where $V_m$denotes the set of vertices, $E_m$denotes the set of edges and L_m

is the set of layers. A node in set V_m is represented as V_i^{α} here $V_i \varepsilon V_m$ and $\alpha \varepsilon L_m$.

b) Also, assume each layer includes all nodes belongs to V_m such that $|V_m| = N \times L$.
c) Compute eigenvector centrality, page rank centrality and degree centrality defied for multiplex network for each node.
d) Next, create a training set for each $x_L = \{x^1, x^2, .., x^m\}$ using centrality measures values.
e) Compute probability for each node using training set and Gaussain model as follow:

$$p(x_L; \mu, \sigma^2) = {}^{1}/_{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_L-\mu)^2}{2\sigma^2}}$$

Here, μ is the mean and σ^2 is the variance.

f) Create a probability set $P^i = P_L^i$, which contains probability of a node in each layer.
g) Next, we pick the highest probability form the probability set such as $P^i = \max(P_L^i)$.
h) Create a probability set for each node as follow $P = \mathrm{P}(x^i)$.

3.2. Compute Cross-layer Anomaly

The cross layer anomaly detection (CAD) for multiplex network is given as

$$P(x^i) < \varepsilon$$

Here, $P(x^i)$ is the probability of a node in multiplex network and ε is the threshold. If the probability value is less then the given threshold then the node is anomalous in the network. We check the same for multiple values of threshold. We validate our approach on two dataset, which we discussed in later sections.

.

4. Dataset Description

4.1. Danio-Rerio Dataset

The Danio Rerio network dataset is a genetic and protein multiplex network of compiled by C. Stark et al. [14]. This dataset consists of five layers; in which each segment represents the system of five components of genetic relations for organisms. Here, each organism act as the node of the network; genetic interaction between any two-organism acts as an edge of the network and different type of communication or association between organisms corresponds to each layer of the network. The basic statistics of each layer of the dataset is in table 1.

Table 1. Parameters of Danio-Rerio dataset

Layers	**# nodes**	**#edges**	**#Density**	**#Mean Path Length**
Association	4	2	0.5	1
Suppressive genetic interaction	33	25	0.8	1.4
Direct interaction	32	36	1.1	1.5
Additive genetic interaction	42	44	1.0	2.2
Physical association	69	81	1.2	2.6

4.2. Florentine Marriage Dataset

Padgett [7] compiled the Florentine Families dataset. This dataset represents the marriage ties or business ties among Florentine families. This dataset consists of two layers, one layer serves the marriage alliance, and one

layer represents the business alliance. In this dataset, each family member represents a node and relationship between two is represented by edges. The primary characteristic of the dataset is in table 2.

Table 2. Parameters of Florentine Marriage dataset

Layers	# nodes	# edges	#Density	#Mean Path Length
Marriage	15	20	1.3	2.5
Business	10	15	1.4	2.4

In this multiplex network, high number of node overlapping exists, which increase the computation overhead but helps us to find out comparable results of closeness centrality.

5. Results

In this section, we discuss our findings on anomaly detection in multiplex networks. We run our algorithm to two datasets, which we discussed in earlier section.

5.1. Danio-Rerio Dataset Results

In this dataset, there is least overlapping of nodes exist between two consecutive layers of the network but there is high connectivity within the layer. This results to get less anomalous nodes in the system. The results obtained from Danio Rerio network shows that there are 4% of the nodes are irregular nodes. In figure 5, figure 6 and figure 7 we have demonstrated the outcomes of our algorithm. In figure 3, we visualize the dataset and show the placement of nodes concerning latency and throughput.

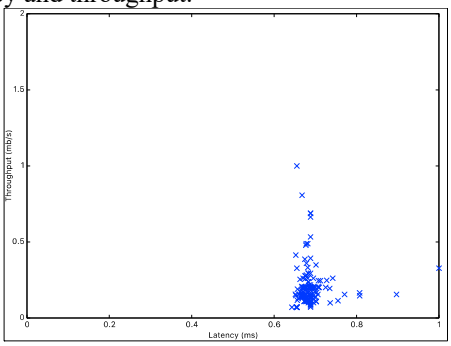

Fig. 3. Visualization of danio rerio dataset

In figure 4, we have shown the Gaussian distribution contours of the distribution fit to the dataset. The nodes fit inside the range of green line have the least possibility of being an anomalous node. In figure 5, nodes, which are irregular in the network, are encircled with red triangle.

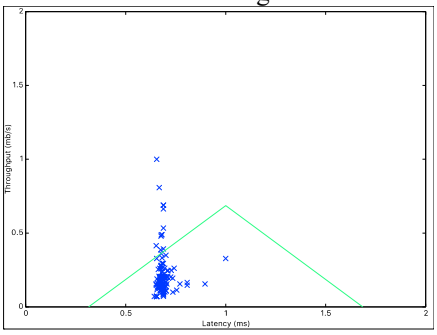

Fig. 4. gaussian distribution contours of the distribution fit to the dataset

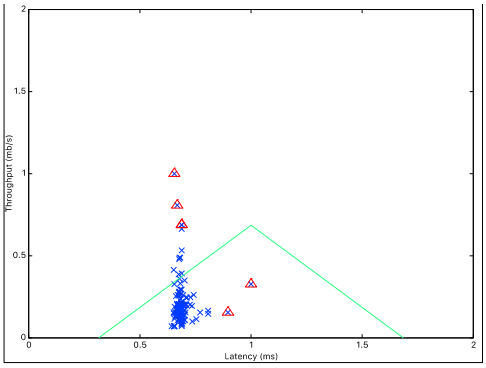

Fig. 5. Nodes highled with red traingle are anomolous nodes

5.2. Florentine Marriage Dataset Results

In this dataset, there is considerable overlapping of nodes exist between two consecutive layers of the network.The results obtained from Florentine families network shows that due to great marriage alliance and business alliance exist between individuals the percentage of anomalous nodes is very less. As the total number of nodes in the network is 15; there exist only two irregular nodes in the network. In figure 6, figure 7 and figure 8 we have shown the outcomes of our algorithm. Figure 8 shows the placement of nodes concerning latency and throughput.

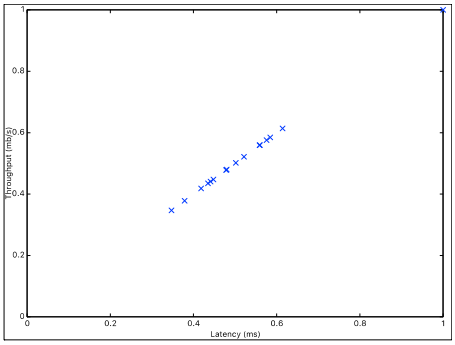

Fig. 6. Visualization of florentine marriage dataset

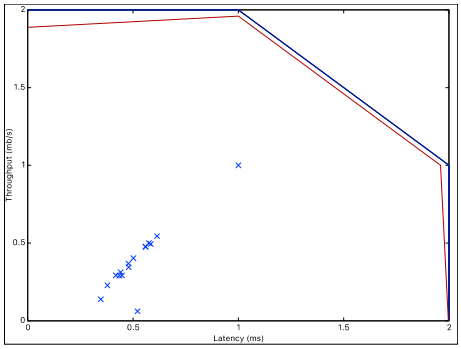

Fig. 7. gaussian distribution contours of the distribution fit to the dataset

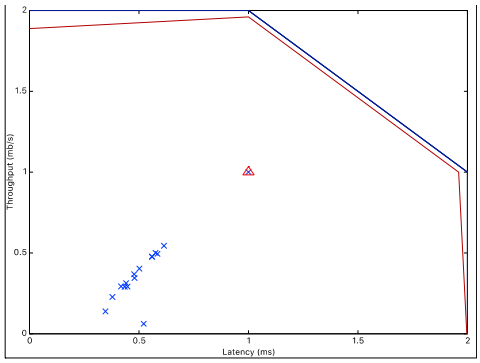

Fig. 8.Nodes highled with red traingle are anomolous nodes

4. Conclusion

In this paper, we present an approach for detecting anomalies in multiplex networks. From our research, we believe this is the initial effort to detect anomalies in the multiplexed network. These works encourage us to dig more knowledge from multiplex networks. There are number of ways to extend this work such as we may apply the proposed algorithm to directed or weighted multiplex networks.

References

[1] M. E. J. Newman, “Networks: An introduction”, Oxford University Press, Oxford, 2010.

[2] Qi Ding, Natallia Katenka, Paul Barford, Eric D. Kolaczyk, and Mark Crovella. Intrusion as (anti)social communication: characterization and detection. In Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Beijing, China, pages 886–894. ACM, 2012.

[3] Brigitte Boden, Stephan Gunnemann, Holger Hoffmann, and Thomas Seidl. Mining ¨ coherent subgraphs in multi-layer graphs with edge labels. In Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Beijing, China, pages 1258–1266. ACM, 2012a.

[4] Corinna Cortes and Daryl Pregibon. Signature-based methods for data streams. Data Mining and Knowledge Discovery, 5(3):167–182, 2001.

[5] V. L. F. Battiston, V. Nicosia, “Structural measures for multiplex networks,” *Physical Review E*, vol. 89, 032804, 2014.

[6] M. Kivela, A. Arenas, M. Barthelemy, J. Gleeson, Y. Moreno, and M. Porter, “Multilayer networks,” *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

[7] U Kang, Jay-Yoon Lee, Danai Koutra, and Christos Faloutsos. Net-Ray: Visualizing and mining web-scale graphs. In Proceedings of the 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Tainan, Taiwan, 2014.

[8] Heli Sun, Jianbin Huang, Jiawei Han, Hongbo Deng, Peixiang Zhao, and Boqin Feng. gskeletonclu: Density-based network clustering via structure-connected tree division or agglomeration. In Proceedings of the 10th IEEE International Conference on Data Mining (ICDM), Sydney, Australia, pages 481–490. IEEE Computer Society, 2010.

[9] Charu C. Aggarwal and Philip S. Yu. Outlier detection for high dimensional data. In Proceedings of the ACM International Conference on Management of Data (SIGMOD), Santa Barbara, CA, pages 37–46. ACM, 2001.

[10] Vydunas Saltenis. Outlier detection based on the distribution of distances between data points. Informatica (Lithuanian Academy of Sciences), 15(3):399–410, 2004.

[11] David J. Miller and John Browning. A mixture model and em-based algorithm for class discovery, robust classification, and outlier rejection in mixed labeled/unlabeled data sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(11):1468–1483, 2003.

[12] Charu Aggarwal and Karthik Subbian. Evolutionary network analysis: A survey. ACM Computing Surveys, 2014.

[13] A. A. M. De Domenico, M.A. Porter, “Muxviz: a tool for multilayer analysis and visualization of networks,” *Journal of Complex Networks*, pp. 1–18, 2014

[14] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. - "Biogrid: a general repository for interaction datasets" - Nucleic Acids Research 2006 34 (1) D535–D539

[15] JF Padgett, CK Ansell - "Robust Action and the Rise of the Medici, 1400-1434". American journal of sociology, 1259-1319 (1993)

[16] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Go´mez- Garden˜es, M. Romance, I. Sendin˜a-Nadal, Z. Wang and M. Zanin, ‘The structure and dynamics of multilayer networks’, Phys. Reps. 544 (2014) no. 1, 1–122.

[17] M. De Domenico, A. So´le-Ribalta, E. Cozzo, M. Kivela¨, Y. Moreno, M. A. Porter, S. Go´mez and A. Arenas, ‘Mathematical formulation of multilayer networks’, Phys. Rev. X 3 (2013) 041022.

[18] A. Sole´-Ribalta, M. De Domenico, S. Go´mez, and A. Arenas, “Central- ity rankings in multiplex networks,” in *Proceedings of the 2014 ACM Conference on Web Science*, ser. WebSci ’14, 2014, pp. 149–155.

[19] Tanmoy Chakraborty, Ramasuri Narayanam.” Cross-layer Betweenness Centrality in Multiplex Networks with Applications”, ICDE (2016)

[20] Jimeng Sun, Huiming Qu, Deepayan Chakrabarti, and Christos Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. In Proceedings of the 5th IEEE International Conference on Data Mining (ICDM), Houston, TX, pages 418–425. IEEE Computer Society, 2005

[21] Zoltan Gy ´ ongyi, Hector Garcia-Molina, and Jan Pedersen. Combating web spam ¨ with trustrank. In Proceedings of the 30th International Conference on Very Large Data Bases (VLDB), Toronto, Canada, pages 576–587, 2004.

[22] Lei Li, Chieh-Jan Mike Liang, Jie Liu, Suman Nath, Andreas Terzis, and Christos Faloutsos. Thermocast: A cyber-physical forecasting model for data centers. In Proceedings of the 17th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Diego, CA. ACM, 2011b.