egoDetect: Visual Detection and Exploration of Anomaly in Social Communication Network

Jingwen Zhang[1], Jiansu Pu[1], Hui Shao[1], Yuwei Zhang[1], Tingting Zhang[1], Shaolun Ruan[1], Yunbo Rao[1], and Yadong Wu[2]

[1] University of Electronic Science and Technology of China.
[2]Sichuan University of Science & Engineering.

1 Introduction

Up to now, there are still many challenges in social anomaly detection. Supervised methods need tag data, but it is hard to gain. And unsupervised methods can't guarantee the accuracy of anomaly detection. Moreover, Dunbar and Zhou discovered that an ordinary person's social network is hierarchical [1, 3]. In addition, the egocentric network allows experts to learn about the topology and have an intuitive understanding of the ego's network. Besides, while the types of social data are diverse, it is hard to design a suitable visualization model to detect all of them.

Combining the above questions and thinks, we have designed a novel visualization system, egoDetect, which combine the time series and can explore anomalies from both global and local perspectives. We use the temporal LOF algorithm to filter the data. Compared to the existing work [2, 4], it can detect anomalies in the data of social networks without tags. Besides, inspired by the solar system and the social brain hypothesis, we have designed a novel glyph to explore an ego's topology and the relationship between egos and alters. It can help experts have an intuitive analysis on egos. We also add friendly and intuitive interactions to help experts quickly get the information they want.

2 Visualization System Implement

The goal of this system is to assist experts in identify and verify anomalies of users in social networks which are lack of tags, and provide multiple views to validate. Our system is designed as Figure 2. The interface consists of six major UI components: a) a view mapping user's multi-dimensional features into two-dimensional space and showing in the feature space; b) a list of whole users' detailed features sorted based on temporal LOF anomaly scores; c) some statistical information for each segment; d) a novel ego network glyph inspired by solar system for visualizing the structure of the ego's network and the relationship between alters and egos; e) a statistical view representing the active time and habit of the ego; f) a detail view to display the contact between ego and each alter and the anomaly of each alter. In summary, a), b), c) compose our group view, d) is our ego view, and e) and f) are our detail view.

The group view is aimed at displaying the whole picture of the network and helping us make a preliminary classification of data. Through the analysis of the group view, experts can pick out the points they want to continue analyzing. And in ego view, we need to display the relationship between the ego and all his/her alters, so that experts can explore through a sociological perspective. With the help of solar ego network, it is easier to analyze the topology and the relationship between egos and alters. It is often not enough to only provide the overall information of an ego. Sometimes anomalous egos may have no difference with others in a holistic perspective, or we want to dig deeper about egos' behavior with alters, and thus we design a detail view.

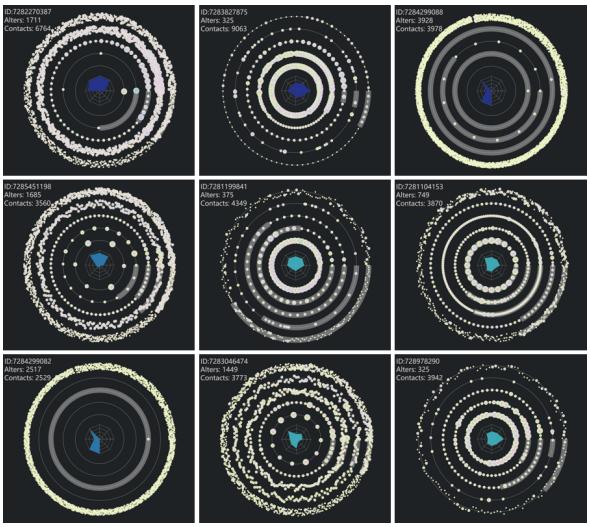

Figure 1: Top 9 users' ego views. Decrease from left to right and from top to bottom

3 Case Study

We apply our system in the task of anomaly detection with a call record data provided by an operator to demonstrate the effectiveness of our system. According to statistics, this is a data set of 32,521,180 call records data involving 6,520,121 users and 751,584 local users. First of all, we have made a preliminary exploration, shown in Figure 2 (a) & (b), we can find that most users are regarded as normal users, while a few users are regarded as outliers who are distributed at the edge.

From Figure 2 (b), we notice that nine egos score more than three, and from the Figure 2 (c), the data of each segment shows a downward trend. We also find that the number of alters with anomaly scores less than 1 does not exceed 150, but they make up only about 6%, while the percentage of egos whose anomaly scores are less than 1.5 is about 95%, and the alters' number of them is no more than 216, which is larger than the Dunbar's Number. We think the reason for this is that communication is bidirectional firstly, which means that you may receive calls you don't want to answer, leading to an increase in the number of contacts. Secondly, it may be affected by the algorithm, and there is the possibility of misclassification. We need a deep analysis to validate.

In order to further verify the effectiveness of the system, we

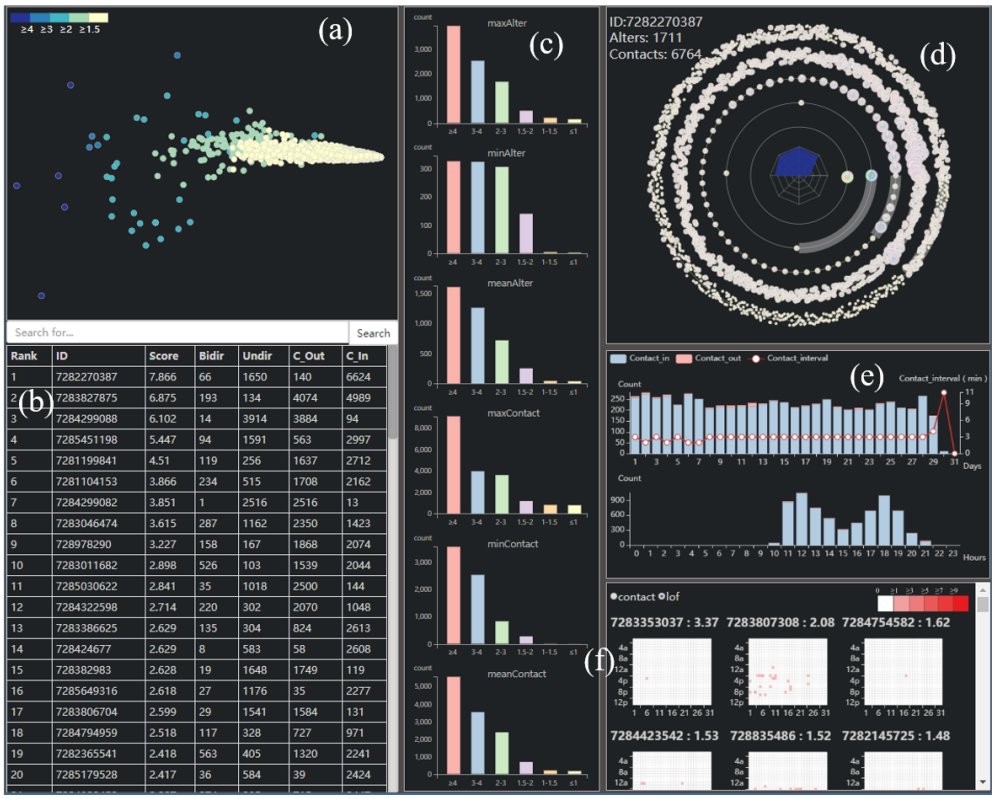

Figure 2: The overview of egoDetect based on the call record data. The user interface consists of six parts: (a) the distribution of users with their features, (b) a list sorted by users' anomaly scores, (c) statistical information for each segment, (d) the ego network glyph inspired from solar system, (e) the statistical view of ego's active time and behavior, (f) the detail view about the contact between the ego with each alter.

proceed from the actual case and demonstrate the system. We have drawn ego views of all users with ratings greater than 3. As shown in Figure 1, their alters and calls are particularly numerous. They mainly show two kinds of structure, either focusing on the outer layer, showing the characteristics of advertisement users, or mainly focusing on the inner layer, showing the characteristics of robots. From their central radar maps, we can see that they have distinct convex shapes.

4 CONCLUSION

We propose a novel visualization system, which has novel visual glyphs and uses multi-view to explore, detect and analyze the anomaly in the social network. The system analyses from macroscopic, mesoscopic and microscopic perspectives. We show the abnormal situation in the online social communication network after anomaly detection from a macroscopic point of view. In the mesoscopic view, we introduce galaxy maps, combined with the ego central network analysis method, to display the interested users in multi-dimensional, from the network structure, active time, alters intimacy and other aspects to judge the abnormal degree of users. And through the microscopic view, combined with timing, we can evaluate the abnormal degree of users from the point of view of alters. We also add friendly and intuitive interactions to help researchers quickly get the information they want. We use a call record data to demonstrate the system is beneficial for detecting abnormal behavior in online social communication.

But limited by time and energy, our work still has a lot to improve. Through the case study, we find that although the algorithm can help us to mine latent anomalous egos by combining time series, it also incorrectly classifies some normal errors. Restricted by data sets, it is difficult for us to analyze alien alters and egos, which is disadvantageous to our analysis.

In the future, we plan to design better anomaly detection algorithms. It can make our detection accuracy higher. Besides, the data set used in this experiment is only provided by a certain operator, so there are limitations in the analysis of specific contacts. In the follow-up experiments, we hope to deepen cooperation with other operators, obtain more and more communication data from the external network, and conduct more in-depth research.

5 ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61872066 and U19A2078). We would like to thank all the participants involved in the studies for their valuable feedback, the reviewers for their constructive comments. We are grateful for the informative discussions with Qing Wang, Salako Abdulhaq Adetunji, Yuewei Zhang, and Mei Han.

REFERENCES

[1] R. A. Hill and R. I. M. Dunbar. Social network size in humans. *Human Nature*, 14(1):53–72, mar 2003. doi: 10.1007/s12110-003-1016-y

[2] C. Nan, S. Conglei, L. Sabrina, L. Jie, L. Yu-Ru, and L. Ching-Yung. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization & Computer Graphics*, 22(1):1–1, 2015.

[3] L. Weixin, M. Vijay, and V. Nuno. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 36(1):18–32, 2013.

[4] R. Xiong and J. Donath. Peoplegarden: Creating data portraits for users. In *Acm Symposium on User Interface Software & Technology*, 1999.