egoPortray: Visual Exploration of Mobile Communication Signature from Egocentric Network Perspective

Qing Wang[1](✉), Jiansu Pu[2](✉), Yuanfang Guo[3], Zheng Hu[1], and Hui Tian[1]

[1] State Key Laboratory of Networking and Switching Technology,
School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{wangqingval,huzheng,tianhui}@bupt.edu.cn
[2] CompleX Lab, Web Sciences Center, Big Data Research Center,
University of Electronic Science and Technology of China, Chengdu 611731, China
jiansu.pu@uestc.edu.cn
[3] State Key Laboratory of Information Security, Institute of Information
Engineering, Chinese Academy of Sciences, Beijing 100093, China
eeandyguo@connect.ust.hk

Abstract. The coming big data era calls for new methodologies to process and analyze the huge volumes of data. Visual analytics is becoming increasingly crucial in data analysis, presentation, and exploring. Communication data is significant in studying human interactions and social relationships. In this paper, we propose a visual analytics system named egoPortray to interactively analyze the communication data based on directed weighted ego network model. Ego network (EN) is composed of a centered individual, its direct contacts (alters), and the interactions among them. Based on the EN model, egoPortray presents an overall statistical view to grasp the entire EN features distributions and correlations, and a glyph-based group view to illustrate the key EN features for comparing different egos. The proposed system and the idea of ego network can be generalized and applied in other fields where network structure exits.

Keywords: Communication network · Ego network · Visual analytics · Communication signature

1 Introduction

The booming of information and communication technologies nurtures the big data era [1]. Among all these large volumes of data, communication data records the behaviors of how people communicate with each other and how they organize their social networks. The accumulation of such digital records provides a new approach for studying the social networks, human dynamics, and other interesting topics [2,3]. For example, the Call Detail Records (CDRs) can be

L. Amsaleg et al. (Eds.): MMM 2017, Part I, LNCS 10132, pp. 649–661, 2017.
DOI: 10.1007/978-3-319-51811-4_53

used to study the human communication behaviors and human mobility [4–6]. Besides, the digital communication records are perfect for analysing Ego Networks (ENs), which examine the social relationships between a target individual (ego) and its direct contacts (alters) [7]. The key idea of ego network is paying more attention to individuals rather than the overall networks. In-depth insights can be obtained from studying the properties of ego communication networks (ECNs) [8].

With the coming of the forth paradigm [9], new methodologies are in urgent need. Visual analytics is an innovate approach, and it is becoming increasingly popular in data science [10]. Visual representations and interactive techniques take advantage of the human eye's broad bandwidth and pathway into the mind to allow users to see, explore, and understand huge amounts of information at once [11]. Sophisticated visual analytics system can highlight useful data thus convey large amounts of information in a more efficient way. This enables decision making with less cognitive efforts, thereby help the analysts adapt to the big data era.

In this paper, we propose egoPortray to study the ECNs based on ego network model. Specifically, we extract the ECNs from the communication data, and portray the ECNs with six network metrics. In order to visually explore the ego networks, we further design two views for interactive investigations: the first view is the macroscopic statistical view, which use the interactive scatter design to capture the holistic correlations and distributions of different ECN features for the entire data. The second view is the microscopic group view, which use glyph-based design to compare different ECNs from different groups. In summary, we build ECNs based on the communication data and further design a visual analytics system for interactively exploring from macroscopic and microscopic scales.

The rest of this paper is organized as follows. Section 2 presents the related research and makes comparisons. Section 3 describes the data and the methods applied in this paper. Section 4 presents the system overview and the detailed design of the proposed visual analytics system. The whole paper will be concluded in Sect. 5.

2 Related Work

The widespread of mobile communication accumulates the relevant data so that we are able to study the social networks at large scale [12,13]. Onnela *et al.* [14] uncovered the existence of the weak tie effect and further demonstrated its significance to the network's structural integrity by analyzing the weighted mobile communication networks. Eagle *et al.* [13] found it possible to infer 95% of friendships accurately based only on the mobile communication data. Miritello *et al.* [15] uncovered the time constraints and communication capacity by studying the individual's communication strategies. Saramäki *et al.* [16] showed that individuals have robust and distinctive social signatures that can persist overtime. Wang *et al.* [17] studied the communication network from ego perspective, and found that ECN size played a crucial role in affecting its structure properties. As illustrated above, much attentions have been paid in uncovering the overall features of the mobile communication networks while only limited studies on ECNs have been reported.

Ego network has also been a heated topic in the information visualization community recently. Shi *et al.* [18] proposed a new 1.5D visualization design to reduce the visual complexity of dynamic networks. Liu *et al.* [19] raised a constrained graph layout algorithm on dynamic networks to prune, compress, and filter the networks in order to reveal the salient part of the network. Wu *et al.* [20] presented a visual analytics system named egoSlider for exploring and comparing dynamic citation networks from 3 levels. Cao *et al.* [21] proposed TargetVue, which applied glyph-based design in detecting the anomalous users of online communication system via unsupervised learning. Liu *et al.* [22] introduced egoComp, the storyflow-like links design, to compare two ego networks. Among all these diverse literatures, most studies mainly focus on exploring the spatial and temporal features of communication behaviors. However, the directions of communications are also important in understanding social relationships [23], and the studies on directed ego network topological features are still insufficient.

In this paper, we propose a two-level visual analytics system egoPortray based on weighted directed EN model, which provides a macroscopic statistical view to display various ego network features and a microscopic multi-feature view to visually compare grouped users. Different from the 1.5D egocentric dynamic network visualization [18], egoPortray does not visualize the communication behaviors directly, but shows the EN properties. In order to display large networks, EgoNetCloud [19] proposed algorithms to compress the networks while egoPortray shows the statistical features of the ego networks. egoSlider explored the citation networks from 3 scales by applying node-link, time-line, and glyph-based designs, but such designs did not support visualizing large networks (with million nodes). TargetVue [21] utilized the glyph-based design to illustrate the top anomalous users whilst showed limited overall ECN information. egoComp [22] applied storyflow-like links into node-link graph to compare the alters from 2 ego networks and egoPortray supports comparing a group of users. Different from the above researches, egoPortray proposed the directed weighed ego network model and visualized ego network properties instead of displaying the communication behaviors directly. Visualizing ego network properties also enables egoPortray to present very large networks and compare more egos at the same time.

3 Data and Methods

The call detail records are collected by mobile operators for billing and network traffic monitoring. The basic information of such data contains the anonymous IDs of callers and callees, time stamps, call durations, and so on. In this study, the data set is provided by one of the largest mobile operators in China. It covers 7 million people of a Chinese provincial capital city for half a year spanning from Jan. to Jun. 2014. According to the operator the users choose, all the users can be divided into two categories, namely, the *local* users (customers of the mobile operator who provide this data set) and the *alien* users (customers from the other operators). The reason for such distinction is that the communication behaviors of *alien* users are not recorded completely by this dataset. As a result, we only focus on the *local* users whose entire calling behaviors are included within the dataset. The basic statistics of the mobile communication data are summarized in Table 1.

Table 1. Basic statistics of the mobile communication networks.

Time	N_t (*total* users)	N_l (*local* users)	L_t (*total* links)
Jan.	6520121	751643	32521180
Feb.	6234877	742504	27600221
Mar.	6481767	783751	32720452
Apr.	6526250	777486	32383231
May	6561107	787614	34119390
Jun.	6531076	787156	33461297

Mobile communication is important in maintaining social relationships nowadays [16,24]. Different from the reciprocity nature of communication in off-line life, mobile communication is intrinsically directed. The directions of communication are significant in understanding the relationships among people and the information diffusion process, especially for digital social networks [23]. Therefore, the mobile communication network can be modeled as a directed graph $G(V, E)$ with the number of nodes and links being $|V| = N$ and $|E| = L$, respectively. Link weight is defined as w_{ij} for a directed link l_{ij}, which is the number of calls that user i has made to user j. It is the link strength between two users. The directed weighted ego network model is built for all the egos within the communication graph, and it is composed of all the direct contacts of an centered ego as well as all the directed weighted links between them. The ego network model and the metrics applied are demonstrated in Fig. 1, the definitions of the metrics will be given in the following paragraphs.

The directions of communication divided the alters into two sets for ego i's ECN: the in-contact set C_i^{in} and the out-contact set C_i^{out}. The sizes of C_i^{in} and C_i^{out} are in-degree k_i^{in} and out-degree k_i^{out}, respectively. k_i^{out} represents the ECN size ego i maintains while k_i^{in} reflects the influence of ego i in the network. In this paper, we mainly focus on k^{out}, because it represents the number of alters an ego intends to spend cognitive resources to maintain. We further define the node weight of an ego as $W_i = \sum_{j \in C_i^{out}} w_{ij}$ to indicate the total amount of cognitive resources an ego spend on maintaining his/her social relationships. In fact, the call durations are also important in communication behaviors and the link weight in duration perspective can be defined as $Wd_i = \sum_{j \in C_i^{out}} wd_{ij}$, where wd_{ij} is the call duration from i to j. To further investigate the properties

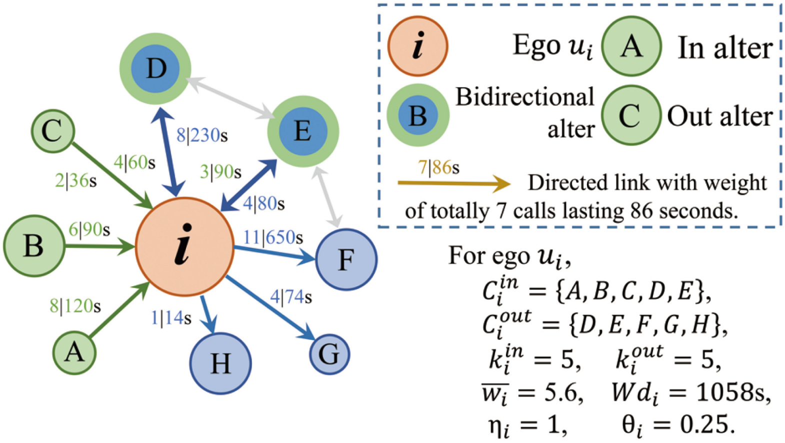

D
C
E
B
i
F
A
H
G
4|60s
2|36s
8|230s
3|90s
4|80s
6|90s
11|650s
8|120s
1|14s
4|74s
i
Ego u_i
A
In alter
B
Bidirectional alter
C
Out alter
7|86s
Directed link with weight of totally 7 calls lasting 86 seconds.
For ego u_i,
$C_i^{in} = \{A, B, C, D, E\}$,
$C_i^{out} = \{D, E, F, G, H\}$,
$k_i^{in} = 5$, $k_i^{out} = 5$,
$\overline{w_i} = 5.6$, $Wd_i = 1058s$,
$\eta_i = 1$, $\theta_i = 0.25$.

Fig. 1. The ego network structure and the network metrics.

of the ECNs, another three metrics are also introduced, namely, average node weight $\overline{w}$, attractiveness balance η, and tie balance θ.

For ego i, the average node weight $\overline{w_i}$ is defined as:

$$\overline{w_i} = \frac{1}{k_i^{out}} \sum_{j \in C_i^{out}} w_{ij}, \tag{1}$$

where w_{ij} is the weight of link l_{ij}, and k_i^{out} is the size of ECN. This metric indicates the average emotional closeness between an ego and the alters [16, 24].

Considering the communication directions, we need to pay attention to the structural balance between in-contacts and out-contacts. Large number of in-contacts indicates the attractiveness of this ego to the network while large number of out-contacts indicates the attractiveness of the network to this ego. Thus we introduce the attractiveness balance (AB) to measure such relationships. It is defined in a straight forward way:

$$\eta_i = \frac{k_i^{in}}{k_i^{out}}. \tag{2}$$

The attractiveness balance $\eta = 1$ means that the number of contacts a user calls is equal to the number of contacts who call him/her, suggesting the balance of the attractiveness. Large η implies strong attractiveness of an ego whilst small η refers to a weaker attractiveness.

Apart from the attractiveness balance, communication direction also distinguishes bidirectional alters (who appear in both C_i^{in} and C_i^{out}) from the unidirectional ones (who only appear in either C_i^{in} or C_i^{out}). Usually, the reciprocal relationships are stronger than the unidirectional relationships, thus they can be viewed as strong and weak ties [25]. Without lost of generosity, strong ties in this paper suggest reciprocal intentions of forming the relationships thus have larger chance to provide mutual support than the weak ties. In order to measure the balance between strong and weak ties within the ECN, we introduce another structural balance metric named tie balance (TB), which is defined as the Jaccard distance [26] between C_i^{in} and C_i^{out}. Mathematically, it reads:

$$\theta_i = \frac{|C_i^{in} \cap C_i^{out}|}{|C_i^{in} \cup C_i^{out}|}. \tag{3}$$

$\theta = 1$ means all of ego i's direct contacts have bidirectional links with ego i, while $\theta = 0$ means ego i has no reciprocal contacts. The above two kinds of ECNs are all extremely imbalance. Strong relationships can provide support while weak relationships can provide diverse information, and people tend to organize their ECNs with a balanced proportion of strong and weak relationships [27].

4 Visualization and Experiments

In this section, the visual analytics system egoPortray will be presented and demonstrated with experiments. The system overview, user interface, and the proposed two views will be illustrated and discussed consecutively.

4.1 System Overview

By analyzing the communication data interactively from macroscopic to microscopic perspective, egoPortray can be used to discover the overall data distributions and correlations as well as compare different ECNs. With these designs and functions, egoPortray can be applied in analyzing user behaviors such as anomalous user detection. Thus the most significant requirements of this system are: (1) extracting the ECN models from the raw communication data; (2) calculating the specified metrics for the ECNs; (3) conducting some statistical analysis and application algorithms like anomalous ranking; (4) visualizing the overall distributions of ECNs; (5) comparing different ECNs.

Figure 2 illustrates the system architecture and the data processing pipeline of egoPortray. It mainly consists of four modules: the data storage module, the processing module, the analysis module, and the visual representation module.

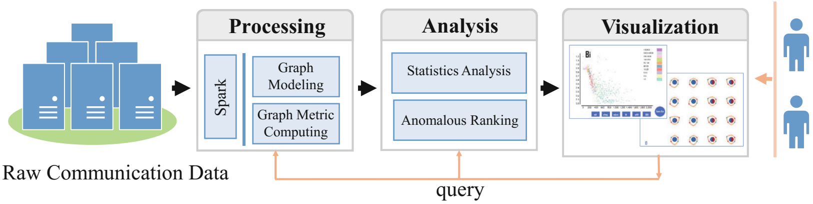


Fig. 2. The system overview and data processing pipeline.

Among them, the data storage module stores all the raw communication data (as described in Data section). These data are subsequently sent to the processing module which is built on Apache Spark [28] to get the ECNs by cleaning and processing. ECNs are stored as instances which contains the information of interactions between ego and alters. With such data, the analysis module can conduct the basic statistical analysis and specific computing tasks, *e.g.* similarity ranking, anomalous ranking, and filtering. After all these procedures, the results are visualized in the visualization module, where an user interface is designed to present the two views.

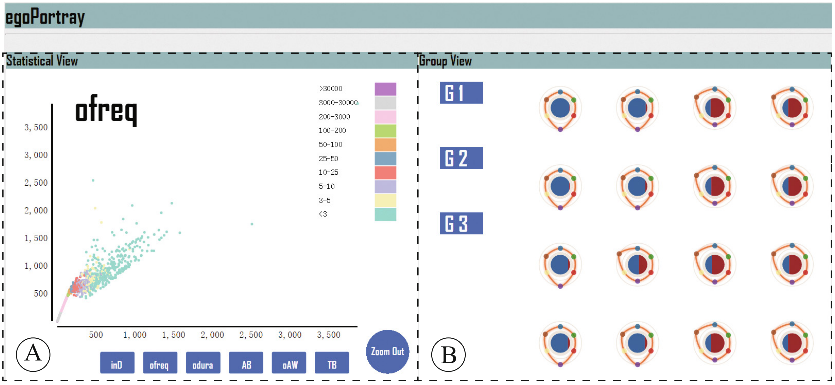
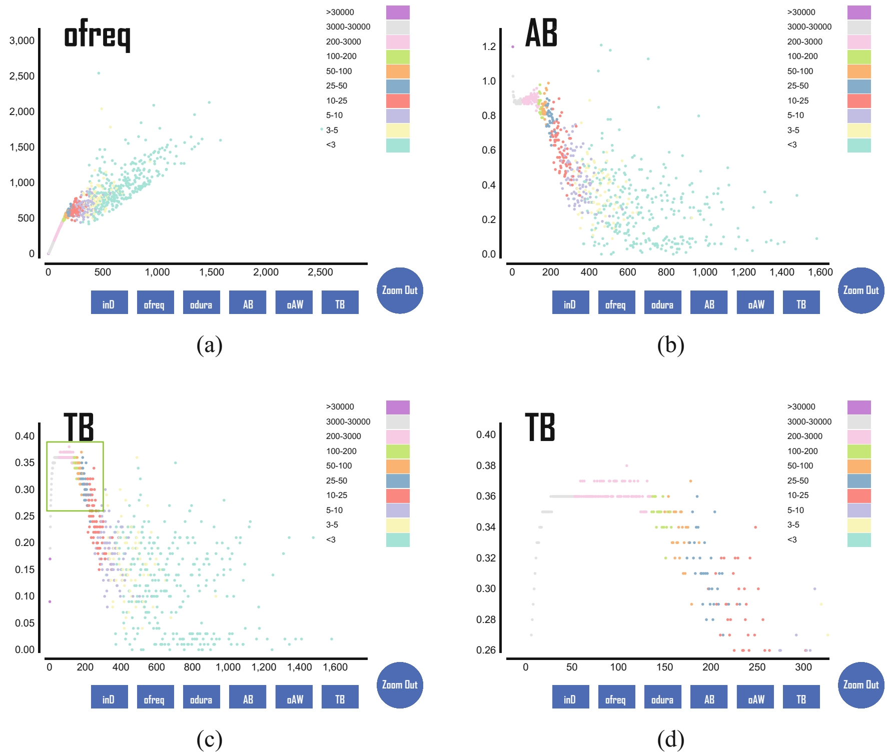


Fig. 3. The user interface of the egoPortray.

The user interface of egoPortray is illustrated in Fig. 3. In this figure, area A shows the statistical view which presents the distributions of the ECNs' features and their correlations with the ECN size. Analysts can zoom in and zoom out to interactively explore the correlation patterns in this view. Area B presents the group view, in which a few egos are taken out as groups for comparison. The design of each view will be presented and discussed in the following two sub-sections.

4.2 Statistical View

Due to the large quantity of users (more than millions), it is almost impossible to present all the users on the screen directly. Granted that it is possible, such large volume of information will overrun and distract the analysts. Statistical distributions are more efficient and practical than directly visualizing such communication data for grasping the overall information. According to Wang's research [17], the size of ECN plays a crucial role in affecting other ECN properties, thus it is better to show the correlation information rather than merely distributions. From this point, we combine the traditional distribution diagram with the correlation diagram in the statistical view.



(a)

(b)

(c)

(d)

Fig. 4. Statistical view of egoPortray. (a) Distribution of W and the correlation between k^{out} and W; (b) Distribution of η and the correlation between k^{out} and η; (c) Distribution of θ and the correlation between k^{out} and θ; (d) Zoom-in of (c) of $k^{out} \in (0, 300)$ and $\theta \in (0.26, 0.40)$.

The overall statistical view is shown in Fig. 4(a), and the buttons below are used for selecting different ECN properties: "inD" for k^{in}, "ofreq" for W, "odura" for Wd, "AB" for η, "oAW" for $\overline{w}$, and "TB" for θ. X-axis and y-axis corresponds to the ECN size (k^{out}) and the selected ECN property, respectively. Each point in the main view stands for a number of egos with the same ECN size, the x-coordinate is the ECN size and the y-coordinate is the average value of the egos' selected property. The color encodes the number of egos within one point, and the legend is placed on the top right corner. This view shows how the users are distributed according to the ECN size and the selected property, and the correlations between ECN size and the selected property.

4.3 Group View

With the help of the statistical view, we can figure out some specific user groups we are interested in (one or several points in the main view). In order to explore the egos within each group and compare different ECNs at the same time, we develop the multi-feature group view. Glyph-based design is intrinsically suitable for visualizing such multidimensional data [29]. The advantages of the glyph design are flexibility, elasticity, and easy for comparing. In this view, each ECN is visualized as a multi-feature glyph and the glyphs are densely packed for comparison in a matrix.

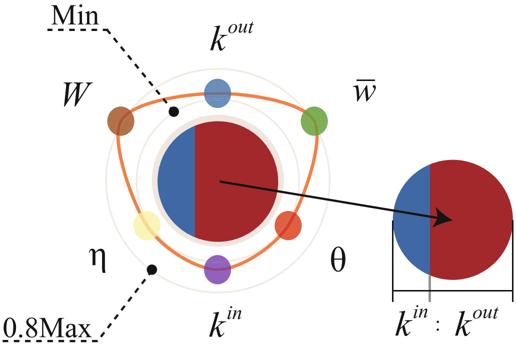


Fig. 5. Glyph design of the group view. (Color figure online)

The basic design of the feature glyph can be found in Fig. 5, where the six small rounds encode six normalized metrics of the ECN, and they are evenly allocated around the centered round. The six properties are: the ECN size (blue round at 0°), the average node strength (green round at 60°), the tie balance (red round at 120°), the in-degree (purple round at 180°), the attractiveness balance (yellow round at 240°), and the node weight on frequency perspective (brown round at 300°). The background rings represent the minimum and 80% of the maximum value for all the metrics. The large round in the center is split into red and blue parts, and the split position indicates the ratio of k^{in} to k^{out} ($i.e.$ η). All the small rounds will be connected by a smooth curve to form the main part of the glyph, thus different ECNs will be mapped to different glyphs. This design emphasize the attractiveness balance of ECN.

Based on the glyph design, the group view is presented in Fig. 6. As illustrated, different egos have different glyphs. There are totally 24 egos illustrated and they are placed according to the groups they belong. Different groups are separated by dashed lines, they are labeled as “G1”, “G2”, and “G3” (each for two columns). This view is useful in comparing different egos in the same and different groups.

4.4 Visual Results

In egoPortray, statistical view helps the analysts to explore the correlations between different ECN properties and the ECN size. In Fig. 4(a), *ofreq* cannot keep the same increasing speed with the increase of ECN size after some k^{out}.

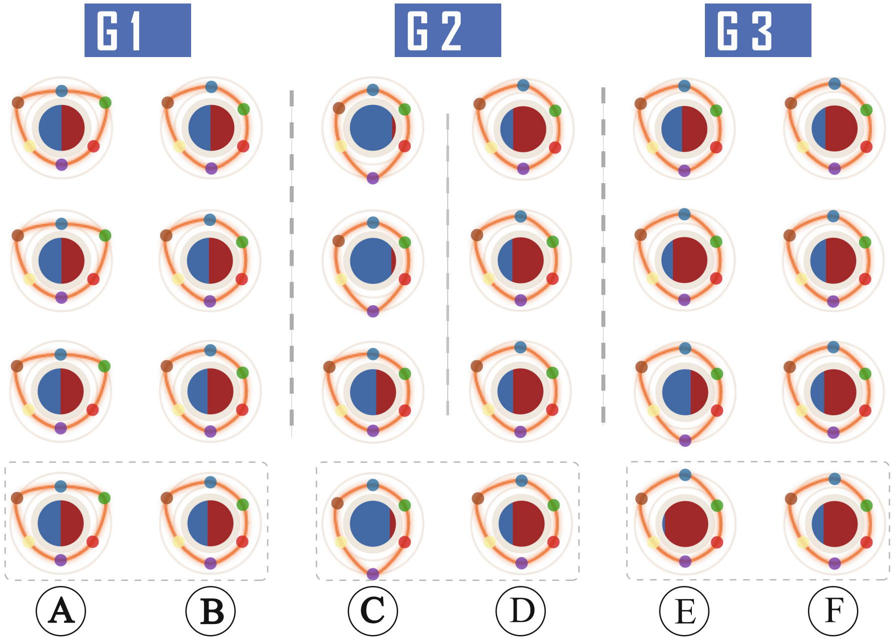


Fig. 6. Group view of egoPortray, users from different groups are visualized as feature glyphs.

In Fig. 4(b), the correlation shows different patterns for different ECN size intervals, but it is not easy to find the exact turning point visually. The similar patterns can be found in Fig. 4(c), which states the correlation between *TB* and ECN size varies with the increase of k^{out}. Thus in Fig. 4(d), we zoom in to examine the exact turning point of the scatter diagram. With the help of the above four steps, we can roughly divide the egos in three groups according to the correlations between the ECN properties and the ECN size. The groups are "G1": $0 < k^{out} \leq 50$ (different patterns for W, η and θ), "G2": $50 < k^{out} \leq 250$ (same pattern for η and θ but different from W), and "G3": $250 < k^{out}$ (same pattern for η and θ but different from W). Such results are agree with the results calculated by K-means algorithm on multi-ECN-property clustering [30].

To further investigate the ECN properties within each group, egoPortray proposes the group view. In Fig. 6, egos are listed in columns according to the groups they belong, and the egos of the last row (within the dashed box) are taken as examples for demonstrating, which are marked by A, B, C, D, E, F. Among them, A and B belong to group 1, C and D belong to group 2, E and F belong to group 3. A has large $\overline{w}$ and W (brown and green rounds are far away from the center), which means A frequently calls a small number of users. Different from A, B calls a small number of alters not so frequently (green round is close to the center). Both of them have balanced number of incoming alters and outgoing alters (the centered large round). In group 2, C has a large number of incoming alters (large proportion of blue area in the centered round), D has more outgoing alters (large proportion of red area in the centered round). In group 3, both egos have extremely large ECNs (the centered round is almost red), but F has a larger θ and larger $\overline{w}$ compared with E. By comparing the egos from different groups, we can see in group 1: With balanced ECN, A and B make lots of calls to the limited alters, this is because they only have small number of alters, thus have enough time and cognitive resources to keep all the social relationships strong enough. When it comes to group 2, users have diversified ECNs, some of them have very large number of incoming alters while others have more outgoing alters, and most of them make lots of calls (brown rounds on the outer background ring). As in group 3, they have similar features, like large number of calls made, larger ECN size (small blue round on the top), and small average number of calls to the alters (low green round). This means the average emotional closeness between ego and alters becomes weak in this group, *i.e.*, egos decrease their average social strength with alters when they have large ECNs. This agrees with the results in [17].

5 Conclusion

In summary, this paper brings about egoPortray, a visual analytics system for analyzing the communication data from ego network perspective. By proposing the directed weighted ego network model, this paper presents a macroscopic statistical view to illustrate the overall distributions and correlations for ECN properties, and develop a microscopic glyph-based group view for ECN comparison. Visual results show that egoPortray can present the data distributions and correlations, and further compare different ECNs from different groups. They also illustrate the potentials of this system in studying communication behaviors. Our design can be generalized to different scenarios where network model applies, and scale to different network sizes. As future works, the temporal information of the ego communication signatures can also be taken into EN models to better explore the communication behaviors. Another potential research direction is to present the geo-information of the ego and alters at the same time to improve mobility predictions.

Acknowledgments. The authors acknowledge Zhi-Yao Teng for useful discussions. This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 61302077, 61421061, 61433014, 11222543 and 61502083) and the Key Project of the National Research Program of China (Grant No. 2012BAH41F03). Qing Wang acknowledges the joint cooperation project of BUPT and China Telecom Beijing Institute.

References

1. Manyika, J., Chui, M., Brown, b., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: the next frontier for innovation, competition, and productivity. http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

2. Barabási, A.L.: The origin of bursts and heavy tails in human dynamics. Nature **435**, 207–211 (2005)
3. Borgatti, S.P., Mehra, A.M., Brass, D.J., Labianca, J.: Network analysis in the social sciences. Science **323**, 892–895 (2009)
4. Song, C., Qu, Z., Blumm, N., Barabási, A.-L.: Limits of predictability in human mobility. Science **27**, 1018–1021 (2010)
5. Miritello, G., Moro, E., Lara, R.: Dynamical strength of social ties in information spreading. Phys. Rev. E. **83**, 045102 (2011)
6. Toole, J.L., Herrera-Yaqüe, C., Schneider, C.M., González, M.C.: Coupling human mobility and social ties. J. R. Soc. Interface **12**, 20141128 (2015)
7. Roberts, S.G.B., Dunbar, R.I.M.: Communication in social networks: effects of kinship, network size, and emotional closeness. Pers. Relatsh. **18**, 439–452 (2011)
8. Fisher, D.: Using egocentric networks to understand communication. IEEE Internet Comput. **9**, 20–28 (2005)
9. Hey, T.: The fourth paradigm – data-intensive scientific discovery. In: Kurbanoğlu, S., Al, U., Erdoğan, P.L., Tonta, Y., Uçak, N. (eds.) Communications in Computer and Information Science, Berlin (2012)
10. Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko J.T., Fekete, J.D., North, C. (eds.) Information Visualization – Human-Centered Issues and Perspectives, Berlin (2008)
11. Mazza, R.: Introduction to Information Visualization. Springer, London (2009)
12. Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Menezes, M.A., Kaski, K., Barabási, A.L., Kertész, J.: Analysis of a large-scale weighted network of one-to-one human communication. New J. Phys. **9**, 179–206 (2007)
13. Eagle, N., Pentland, A.S., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proc. Natl. Acad. Sci. U.S.A. **106**, 15274–15278 (2009)
14. Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Kaski, K., Kertész, J., Barabási, A.L.: Structure and tie strengths in mobile communication networks. Proc. Natl. Acad. Sci. U.S.A. **104**, 7332–7336 (2007)
15. Miritello, G., Moro, E., Lara, R., Martínez-López, R., Belchamber, J., Roberts, S.G.B., Dunbar, R.I.M.: Time as a limited resource: communication strategy in mobile phone networks. Soc. Networks **35**, 89–95 (2013)
16. Saramäki, J., Leicht, E.A., López, E., Roberts, S.G.B., Reed-Tsochas, F., Dunbar, R.I.M.: Persistence of social signatures in human communication. Proc. Natl. Acad. Sci. U.S.A. **111**, 942–947 (2014)
17. Wang, Q., Gao, J., Zhou, T., Hu, Z., Tian, H.: Critical size of ego communication networks. EPL **114**, 58004 (2016)
18. Shi, L., Wang, C., Wen, Z., Qu, H., Liao, Q.: 1.5D egocentric dynamic network visualization. IEEE Trans. Vis. Comput. Graphics. **21**, 624–637 (2015)
19. Liu, Q., Hu, Y., Shi, L., Mu, X., Zhang, Y., Tang, J.: EgoNetCloud: event based egocentric dynamic network visualization. In: IEEE Conference on Visual Analytics Science and Technology (VAST 2015), pp. 65–72. IEEE Press, Chicago (2015)
20. Wu, Y., Pitipornvivat, N., Zhao, J., Yang, S., Huang, G., Qu, H.: EgoSlider: visual analysis of egocentric network evolution. IEEE Trans. Vis. Comput. Graphics. **22**, 260–269 (2016)
21. Cao, N., Shi, C., Lin, S., Lu, J., Lin, Y., Lin, C.: TargetVue: visual analysis of anomalous user behaviors in online communication systems. IEEE Trans. Vis. Comput. Graph. **22**, 280–289 (2016)

22. Liu, D., Guo, F., Deng, B., Wu, Y., Qu, H.: EgoComp: a node-link based technique for visual comparison of ego-network. http://vacommunity.org/egas2015/papers/IEEEEGAS2015-DongyuLiu.pdf
23. Brzozowski, M.J., Romero, D.M.: Who should i follow? recommending people in directed social networks. In: 5th International AAAI Conference on Weblogs and Social Media (ICWSM), pp. 458–461. AAAI Press, Barcelona (2011)
24. Zhou, W.X., Sornette, D., Hill, R.A., Dunbar, R.I.M.: Discrete hierarchical organization of social group sizes. Proc. R. Soc. B. **272**, 439–444 (2005)
25. Zhu, Y., Zhang, X., Sun, G., Tang, M., Zhou, T., Zhang, Z.: Influence of reciprocal links in social networks. PLoS One **9**, e103007 (2014)
26. Levandowsky, M., Winter, D.: Distance between sets. Nature **234**, 34–35 (1971)
27. Brown, J.J., Reingen, P.H.: Social ties and word-of-mouth referral behavior. J. Consum. Res. **14**, 350–362 (1987)
28. Spark. http://spark.apache.org/
29. Ward, M.O.: A taxonomy of glyph placement strategies for multidimensional data visualization. Inf. Vis. **1**, 194–210 (2002)
30. Jain, A.K.: Data clustering: 50 years beyond k-means. J. Pattern Recogn. **31**, 651–666 (2010)