Critical size of ego communication networks

Qing Wang[1,2], Jian Gao[3], Tao Zhou[3,4] (a), Zheng Hu[1,2] (b) and Hui Tian[1,2]

[1] *State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, PRC*
[2] *School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, PRC*
[3] *CompleX Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, PRC*
[4] *Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, PRC*

PACS 89.65.Ef – Social organizations; anthropology
PACS 87.23.Ge – Dynamics of social systems
PACS 89.75.Fb – Structures and organization in complex systems

Abstract – With the help of both communication and information technologies, studies on the overall social networks have been extensively reported recently. However, investigations on the structure and dynamic of the Ego Communication Networks (ECNs) remain insufficient, where an ECN stands for a sub network composed of a centralized individual and his/her direct neighbors. In this paper, the ECNs are built on the Call Detail Records (CDRs), which cover more than 7 million people of a provincial capital city in China for half a year. We investigate the network structural changes as the sizes of ECNs increase, focusing on three network metrics, namely, average link weight, balance index (the ratio of in-degree to out-degree), and self balance distance (the Jaccard distance between the sets of in-contacts and out-contacts). Results show that when the sizes of ECNs exceed a critical value at about 150, the average link weight drops dramatically, and the structural balance of ECN collapses as indicated by the remarkable decrease of both the balance index and the self balance distance. Our work highlights the significance of the network size in studying the structural organization of ECNs and provides a cross-culture supportive evidence to the well-known Dunbar's Social Brain Hypothesis (SBH).

Introduction. – The widespread of information technologies makes it convenient for people to communicate with each other. In the meantime, the social behaviors of users are much easier to be precisely recorded. The accumulation of such digital communication records opens a new way to study social networks, human dynamics and other related areas [1]. For example, the Call Detail Records (CDRs) can be used to study the human communication behaviors and human mobility [2–7]. Besides, the digital communication records are perfect for analysing the so-called Ego Networks (ENs), which examine the ties connecting the target individual (ego) and his/her direct contacts [8]. The studies on the ENs can benefit us in understanding how human organize their personal social networks [9]. Here, we focus on the Ego Communication Networks (ECNs) from CDRs, which is an important kind of social networks and usually represents closer relationships than online social networks, such as Facebook and Twitter.

Most literatures focus on the structure and function of the overall mobile communication networks [10–18]. For example, Onnela *et al.* [10, 11] uncovered the existence of the weak ties effects and further demonstrated its significance to the network's structural integrity by doing comprehensive analyses of the weighted mobile communication networks. Nanavati *et al.* [12, 13] revealed the Bow-Tie shape of the mobile communication networks and proposed a Treasure-Hunt model to picture the whole graph of telecom users according to the in-degrees and out-degrees. Eagle *et al.* [14] found that it is possible to accurately infer 95% of friendships only based on the mobile communication data in comparing with the self-

(a) E-mail: zhutou@ustc.edu
(b) E-mail: huzheng@bupt.edu.cn

report survey data. Wu *et al.* [15] revealed the bimodal distribution in human communication which is composed of a generic Poisson process in individual human behavior and a power-law-like bursts through the interaction with other individuals. Zhao *et al.* [16] showed that the short message communication system displays a strong bursty property yet very weak memory effect. Further, Jiang *et al.* [17] argued that the inter-call duration follows a power law distribution with an exponential cutoff at the population level and Weibull distribution at the individual level. Moreover, Lauri *et al.* [18] showed the existence of temporal homophily by studying a large dataset of mobile phone records. As illustrated above, much attentions have been paid to uncovering the overall features of mobile communications networks.

Apart from the overall analyses, some researchers also considered the mobile communication networks from ego perspective [19–21]. Miritello *et al.* [19, 20] uncovered the time constraints and communication capacity by studying the individual's communication strategies. Saramäki *et al.* [21] showed that individuals have robust and distinctive social signatures which persists overtime. Other well-known studies from ego network perspective are conducted on the online social networks. For example, Dunbar [22] raised the so-called Social Brain Hypothesis (SBH), which implies that the constraint on the group size of primates arises from the information-processing capacity, being highly correlated with the neocortex [23]. As to human society, the SBH suggests that the people can only maintain about 150 friends due to the cognitive constraints [24, 25]. The number 150 is thus well known as "Dunbar's Number".

Extensive empirical studies of online social networks have verified the "Dunbar's Number" [26–31], which can be seen as supportive evidences to the SBH theory. Arnaboldi *et al.* [26,27] argued that the new technologies could not alter the way people organize their social networks by analyzing the Facebook and Twitter networks. Zhao *et al.* [28] found that there is still a magic upper limit at $200 \sim 300$ for Facebook users. Moreover, Gonçalves *et al.* [29] argued that Dunbar's Number is $100 \sim 200$ in Twitter networks. Guo *et al.* [30] showed that one could only maintain relationships with 65 people in a single city based on the study of the New Orleans' Facebook data. Haerter *et al.* [31] proposed a generic model that reconciles social networks with finite capacity agents, and showed that both the model and empirical observations support the hypothesis of the limit $150 \sim 200$. These analyses can be treated as complementary results to off-line social networks.

Among all these diverse literatures, most studies paid attention to either the overall properties of the networks or just one single feature of ECNs. Results on both off-line and online social networks support the same organizing rule that human could not maintain a huge, stable and meaningful personal network due to the cognitive constraints. However, the impacts of the size of an ECN on its structural properties remain further investigation.

In this letter, we study how the network structure changes as the sizes of ECNs increase. Here, ECNs are built based on the CDRs, covering over 7 million people for half a year in a provincial capital city in China. We mainly focus on three network metrics, including average link weight, balance index and self balance distance. Empirical results show that when the sizes of ECNs go beyond a critical value at about 150, the average link weight drops significantly, and both the balance index and the self balance distance decreases remarkably, suggesting that the structure of ego network changes dramatically. Further experiments on the data of a larger period confirm the persistence of these findings. Interestingly, the critical size 150 of the ECN is equal to the well-known "Dunbar's Number". Indeed, the results provide a supportive evidence to the SBH from different cultural backgrounds.

Methods. – The mobile communication network can be represented by a directed graph $G(V, E)$ with the number of nodes and links being $|V| = N$ and $|E| = L$, respectively. Note that, there are no self-connections and multiple links in the network. The weight w_{ij} of a directed link l_{ij} is the number of calls that user i has made to user j. From ego perspective, we define two sets for each user i, namely, the direct in-contact sets C_i^{in} and the direct out-contact sets C_i^{out}. Clearly, the sizes of the two sets C_i^{in} and C_i^{out} are in-degree k_i^{in} and out-degree k_i^{out} of user i, respectively. Different from the directed mobile communication graph, human communications are intrinsically reciprocal. Besides, people usually call out to maintain their social relationships [32]. Thus in this paper we investigate the ego-contact link properties by studying the link between the ego and out-contacts instead of that between the ego and all the contacts. Further to investigate the properties of ECNs, three network metrics are introduced based on the above notations, namely, the average link weight, balance index and self balance distance. In the following, we will briefly describe the definitions of these metrics.

The average link weight measures the average link strength between an ego and his/her contacts in C_i^{out}. For ego i, the average link weight $\overline{w_i}$ is defined as

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

where w_{ij} is the weight of link l_{ij}, and k_i^{out} is the out-degree of ego i. The average link weight of ECN evaluates the relationships between the ego and the out-contacts, which is close related to the network function.

Structural balance is helpful for understanding the structure of signed social networks [33]. Here, the structural balance of ECN means the balance between in-contacts and out-contacts of an ego. In a straightforward way, the balance index is defined as the ratio of an ego's

Table 1: Basic statistics of the mobile communication networks. N_{total} is the *total* number of all users. N_{local} is the number of the *local* users. L is the number of all links.

Time	N_{total}	N_{local}	L
Jan.	6520121	751643	32521180
Feb.	6234877	742504	27600221
Mar.	6481767	783751	32720452
Apr.	6526250	777486	32383231
May	6561107	787614	34119390
Jun.	6531076	787156	33461297

in-degree and out-degree. Mathematically, it reads

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

The balance index $\eta = 1$ means that the number of contacts that a user calls is equal to the number of contacts who calls him/her, suggesting the balance of the ECN.

Another structural balance index is the self balance distance, which is defined as the ratio of the number of contacts in both in-contact and out-contact sets to the number of the contacts in either set. Mathematically, the self balance distance θ_i reads

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

In fact, θ_i is the Jaccard distance [34] between C_i^{in} and C_i^{out} thus $\theta \in [0, 1]$. Different from balance index η, self balance distance θ depicts the proportion of reciprocal contacts to all contacts. For example, $\theta = 1$ means all the ego's direct contacts have bi-directional links with the ego while $\theta = 0$ means that the ego has an extremely imbalanced network which even has no reciprocal contact. Most of the users can maintain a comparatively stable θ even thought it is not 1 which is acceptable for the reciprocity nature of social lifes [35, 36].

Data description. – The anonymous CDRs are collected by mobile operators for billing and network traffic monitoring. The basic information of the data sets contains the anonymous IDs of callers and callees, time stamps, call durations, and so on. In this study, we only use a subset of the original data furnished by one mobile operator. The used data covers 7 million people of a provincial capital city in China for half a year from January to June, 2014. According to the operators that one chooses, all the mobile users in the data set can be divided into two categories, namely, the *local* users (who are customers of the mobile operator) and the *alien* users (who get service from other mobile operators). The reason why doing such distinction is that we can only obtain a fraction of *alien* users' communication records. As a result, we will derive the degree and frequency distributions of all users based on the whole data set, but only focus on the *local* users for the rest analyses unless otherwise specified. Some

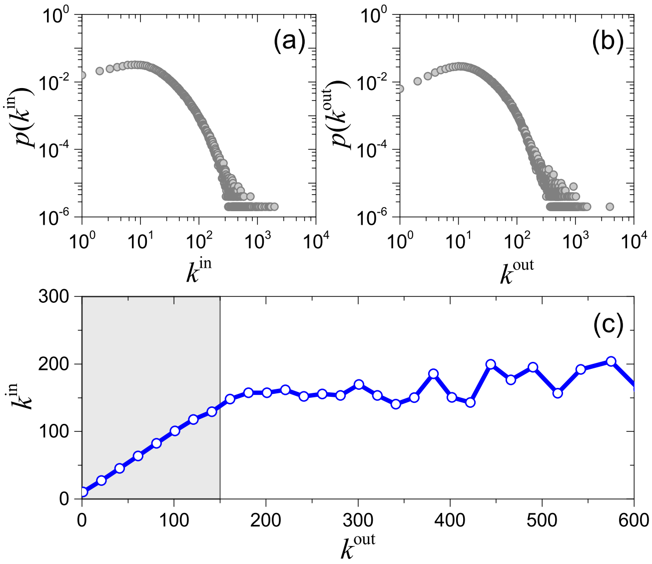

Fig. 1: (Color online) Degree distributions and degree correlation. Panel (a) and Panel (b) stand for in-degree k^{in} and out-degree k^{out} distributions of the ECNs, respectively. The distributions are broad with heavy tails. (c) The correlation of in-degree and out-degree. When $k^{out} < 150$ (marked by shaded areas), the Pearson correlation coefficient of k^{in} and k^{out} is about 0.89.

basic statistics of the mobile communication networks are summarized in Table 1.

Empirical results. – Based on the above mobile communication networks, we investigate the structural changes of the ECNs as the increase of their sizes. We mainly focus on the following three aspects of network properties: the in-degree and out-degree distributions and their correlation, the total link weight W and the average link weight $\overline{w}$, as well as the balance index η and the self balance distance θ. Without loss of generality, the data of January is selected and analyzed as an example, based on which the main results are obtained. And the data of the other five months are used to test the robustness of our findings.

Degree distribution is one of the most fundamental properties of the networks. Figures 1(a) and 1(b) respectively illustrate the in-degree and out-degree distributions, which are broad distributions with heavy tails. We mainly consider the correlation between the in-degree and out-degree of the egos. As shown in fig. 1(c), when $k^{out} < 150$ (i.e., the sizes of the ECN are less than 150), the in-degree is almost linearly correlated with the out-degree as indicated by a very high Pearson correlation 0.89. By contrast, once $k^{out} > 150$, the in-degree can not keep the same increasing speed with the out-degree. The result suggests that egos could only keep much less in-contacts than out-contacts when their ECN sizes exceed 150. Beyond the ECN sizes about 150, egos could only keep less in-contacts than out-contacts. The result is in accordance with the previous findings based on the Facebook data [30].

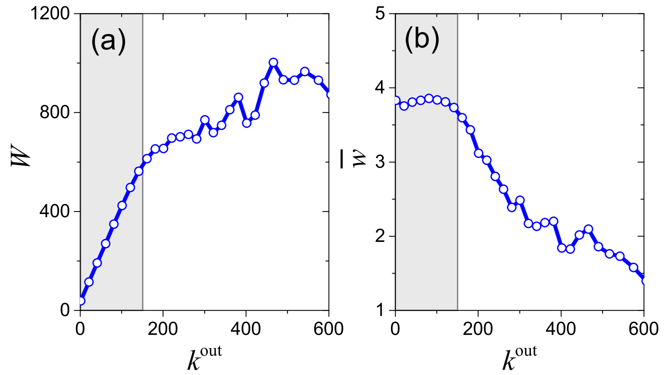

Fig. 2: (Color online) Relations between the weight of links and the ECN size, i.e., the out-degree k^{out}. (a) The total link weight W. (b) The average link weight $\overline{w}$. Shaded areas mark the area with ECN size below 150.

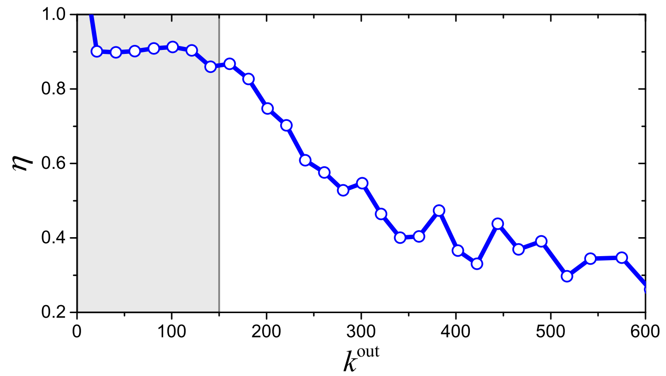

Fig. 3: (Color online) Balance index η with the increase of the ECN size k^{out}. The value of η is stable at about 0.90 when $k^{out} < 150$, and drops dramatically when $k^{out} > 150$. Shaded areas mark the area with ECN size below 150.

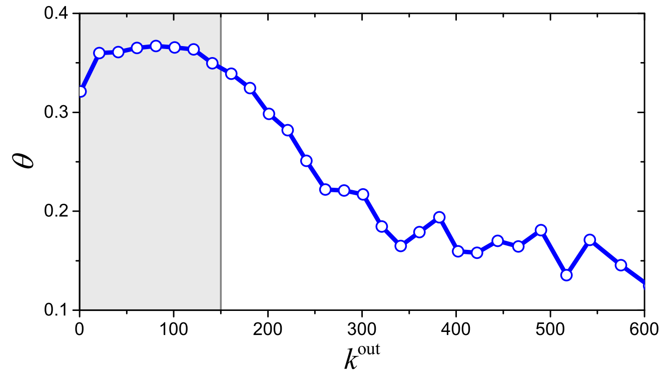

Fig. 4: (Color online) Self balance distance θ with the increase of the ECN size k^{out}. The value of θ is stable at about 0.35 when $k^{out} < 150$, and drops dramatically when $k^{out} > 150$. Shaded areas mark the area with ECN size below 150.

Link weight of the ECN also plays an important role in the network function as it reflects the communication strength. Here, we focus on the total link weight $W = \sum_{j \in C_i^{out}} w_{ij}$ and the average link weigh $\overline{w}$ with the expanding of ECN. As shown in fig. 2(a), the total link weight grows linearly with the out-degree for small ECNs. Value about 150 seems to be the critical point, above which the increase of the total link weight slows down. Further, it can also be observed from fig. 2(b) that the average link weight stays at almost a stable value about 3.8 when the sizes of the ECN are less than 150. The average link weight drops remarkably beyond this critical size. These observations suggest that most egos could only maintain close relationships with about 150 people by making averagely around 4 calls per month. Once the number of out-contacts exceeds 150, the average number of calls to each out-contact drops dramatically. That is to say, the increase of ECN sizes makes it hard to maintain the same level of communication strength between the ego and out-contacts.

Structural balance is important to the social networks. As shown in fig. 3, we can see that the value of the balance index is stable at about 0.90 if the sizes of ECNs are less than 150. The value of the balance index drops remarkably to 0.40 when the sizes of ECNs go to 350. This result suggests that most egos keep balance between in-contacts and out-contacts when their ECN sizes are smaller than the critical value of 150, above which their ECN structural balance collapses as indicated by the small value of η.

Furthermore, we consider the other structural balance index, i.e., the self balance distance θ, and study how it changes with the increase of the ECN size. It can be seen from the $\theta - k^{out}$ curve in fig. 4 that when the sizes of ECNs are smaller than 150, the value of θ can be kept at a relatively stable value around 0.35, meaning that over one third contacts are mutual contacts of an ego. Once again, the number 150 is confirmed to be the critical value, above which the structural balance of ECNs collapses. It is indicated by the remarkable decrease of the self balance distance, which drops to a value below 0.20 when the ECN sizes go beyond 150. The result indicates that an ego starts failing to maintain the same large ratio of mutual contacts to all his/her contacts once the out-contacts exceeds the critical value about 150.

In order to validate the robustness of the above findings, we test them on the data of the following five months from February to June, 2014. Varies of the three metrics (i.e., average link weight $\overline{w}$, balance index η and self balance distance θ) with the increase of the ECN sizes are shown in fig. 5. The curves of the same metric persist over months, demonstrating the robustness of our findings. The critical value of the ECN size shows consistency during the remarkable decrease of the three metrics in different months. The results of longer periods confirm the persistence of this value to be about 150. In addition, from these empirical results, one can conclude that the size of an ECN plays a crucial role in affecting its organization.

Conclusion and discussions. – In summary, we have investigated how the ECN sizes affect the structural properties of the mobile communication networks from ego

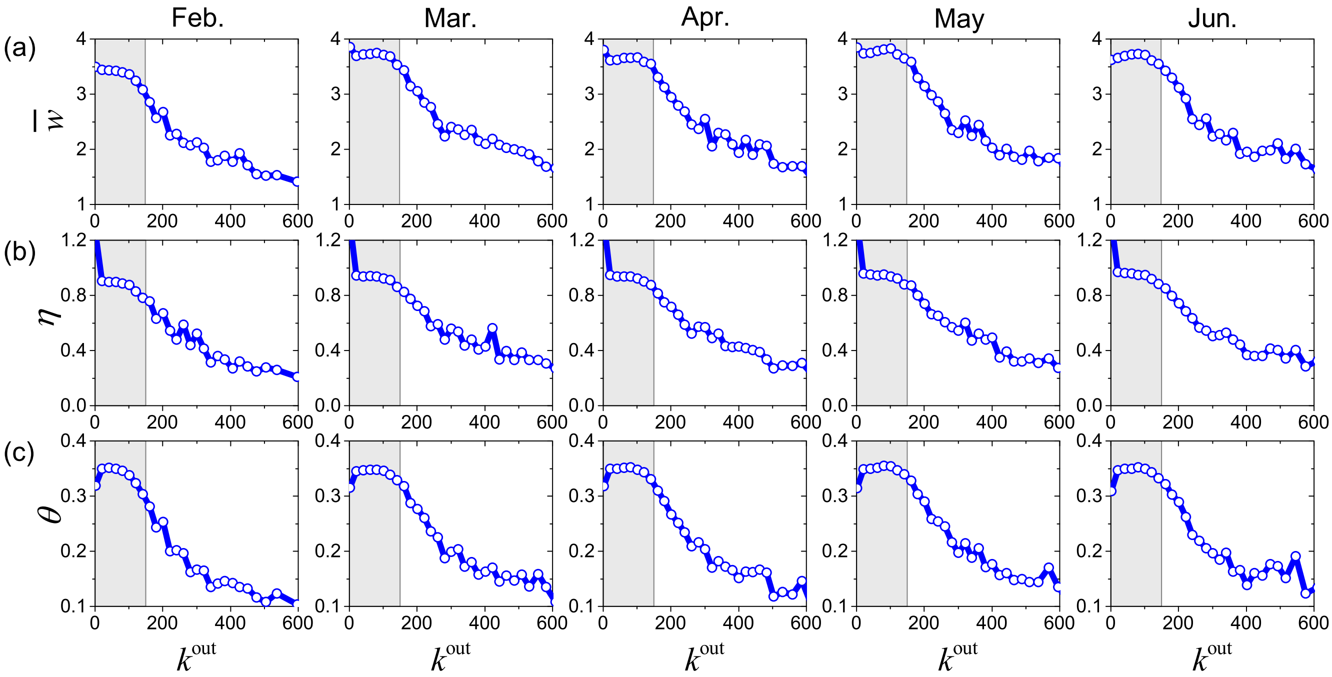

Fig. 5: (Color online) Values of the three structural metrics based on data of the following five months. (a) The average link weight $\overline{w}$. (b) The balance index η. (c) The self balance distance θ. Shaded areas mark the area with ECN size $k^{out} < 150$.

perspective rather than the overall perspective. By introducing three network metrics, we find a critical value of the ECN size at about 150, beyond which the average link weight drops dramatically and the structural balance collapses. These results suggest that the sizes of the ECNs have extraordinary influences on the structures and functions of mobile communication networks.

Interestingly, the critical size 150 of ECNs is just equal to the "Dunbar's Number". Indeed, our findings on ECNs can be viewed as a cross-culture supportive evidence for the SBH theory which is different to previous studies from survey data or online social networks like Twitter and Facebook. For a more intuitive understanding, egos have to spend more cognitive resources in maintaining their relationships as the ECN size increases. However, the total cognitive resources for a single ego are limited, resulting in the reduction of ego's reciprocal contacts once his/her ECN size exceeds some critical value. Mobile communication data is better in reflecting the users' off-line relationships than the online social network data, which makes this work a strong and distinguishable support for the SBH theory compared with the known results. Intuitively, the two introduced structural balance metrics are highly dependent on the ego's personal characteristics, which enables the application in user profiling and spammer detection.

In this work, we mainly focus on the links between the ego and his/her neighbors. In fact, the relationships among his/her neighbors also play a crucial role in understanding the structure of social networks, which is also worth further investigations. As future works, we can explore the network structures of the ECNs considering both link formations and node properties. Moreover, the development of information technologies makes it possible to record more dimensions of personal mobile communication data, and provide us better resources to achieve deeper understanding of both the structures of social networks and the patterns of human mobility.

* * *

The authors acknowledge Zhi-Hai Rong and Zhi-Dan Zhao 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). Q.W. acknowledges the joint cooperation project of BUPT and China Telecom Beijing Institute.

REFERENCES

[1] Borgatti S. P., Mehra A., Brass D. J. and Labianca G., *Science*, **323** (2009) 892.
[2] Fisher D., *IEEE Internet Computing*, **9** (2005) 20.
[3] Gonzalez M. C., Hidalgo C. A. and Barabási A.-L., *Nature*, **453** (2008) 779.
[4] Song C., Koren T., Wang P. and Barabási A.-L., *Nat. Phys.*, **6** (2010) 818.
[5] Song C., Qu Z., Blumm N. and Barabási A.-L., *Science*, **327** (2010) 1018.
[6] Miritello G., Moro E., Lara R., *Phys. Rev. E*, **83** (2011) 045102.

[7] Toole J. L., Herrera-Yaqüe C., Schneider C. M. and González, M. C., *J. R. Soc. Interface*, **105** (2015) 20141128.
[8] Roberts S. G. B. and Dunbar R. I. M., *Personal Relationships*, **18** (2011) 439.
[9] Arnaboldi V., Conti M., Passarella A. and Pezzoni F., in *IEEE International Conference on Social Computing (SocialCom)* (IEEE Press) 2012, pp. 31-40.
[10] Onnela J.-P., Saramäki J., Hyvönen J., Szabó G., M. A. D. Menezes, Kaski K., Barabási A.-L. and Kertész J., *New J. Phys.*, **9** (2007) 179.
[11] Onnela J.-P., Saramäki J., Hyvönen J., Szabó G., Lazer D., Kaski K., Kertész J. and Barabási A.-L., *Proc. Natl. Acad. Sci. USA*, **104** (2007) 7332.
[12] Nanavati A. A., Gurumurthy S., Das G., Chakraborty D., Dasgupta K., Mukherjea S. and Joshi A., in *Proceedings of the 15th ACM International Conference of Information and Knowedge Management (CIKM)* (ACM Press) 2006, pp. 435-444.
[13] Nanavati A. A., Singh R., Chakraborty D., Dasgupta K., Mukherjea S., Das G., Gurumurthy S. and Joshi A., *IEEE Transactions on Knowledge and Data Engineering*, **20** (2008) 703.
[14] Eagle N., Pentland A. S. and Lazer D., *Proc. Natl. Acad. Sci. USA*, **106** (2009) 15274.
[15] Wu Y., Zhou C., Xiao J., Kurths J. and Schellnhuber H. J., *Proc. Natl. Acad. Sci. USA*, **107** (2010) 18803.
[16] Zhao Z.-D., Xia H., Shang M.-S. and Zhao T., *Chin. Phys. Lett.*, **28** (2011) 068901.
[17] Jiang Z.-Q., Xie W.-J., Li M.-X., Podobnik B., Zhou W.-X. and Stanley H. E., *Proc. Natl. Acad. Sci. USA*, **110** (2013) 1600.
[18] Kovanen L., Kaski K., Kertész J. and Saramäki J., *Proc. Natl. Acad. Sci. USA*, **110** (2013) 18070.
[19] Miritello G., Moro E., Lara R., Martínez-López R., Belchamber J., Roberts S. G. B. and Dunbar R. I. M, *Soc. Networks*, **35** (2013) 89.
[20] Miritello G., Lara R., Cebrian M. and Moro E., *Sci. Rep.*, **3** (2013) 1950.
[21] Saramäki J., Leicht E. A., López E., Roberts S. G. B., Reed-Tsochas F. and Dunbar R. I. M., *Proc. Natl. Acad. Sci. USA*, **111** (2014) 942.
[22] Dunbar R. I. M., *Evolutionary Anthropology*, **6** (1998) 178.
[23] Dunbar R. I. M., *Journal of Human Evolution*, **22** (1992) 469.
[24] Dunbar R. I. M., *How many friends does one person need? Dunbar's number and other evolutionary quirks*, edited by Julian Loose (Faber & Faber, Cambridge) 2010, pp. 21-43.
[25] Ruiter J. D., Weston G. and Lyon S. M., *American Anthropologist*, **113** (2011) 557.
[26] Arnaboldi V., Conti M., Passarella A. and Pezzoni F., in *The 5th IEEE International Workshop on Network Science for Communication Networks (NetSciCom)* (IEEE Press) 2013, pp. 229-234.
[27] Dunbar R. I. M., Arnaboldi V., Conti M. and Passarella A., *Soc. Networks*, **43** (2015) 39.
[28] Zhao J.-C., Wu J.-J., Liu G.-N., Tao D.-C., Xu K. and Liu C.-Y., *Neurocomputing*, **142** (2014) 343.
[29] Gonçalves B., Perra N. and Vespignani A., *PLoS ONE*, **6** (2011) e22656.
[30] Guo Q., Shao F., Hu Z.-L. and Liu J.-G., *EPL*, **104** (2013) 28004.
[31] Haerter J. O., Jamtveit B. and Mathiesen J., *Phys. Rev. Lett.*, **109** (2012) 168701.
[32] Oswald D. L. and Clark E. M., *Personal Relationships*, **10** (2003) 187.
[33] Facchetti G., Iacono G. and Altafini C., *Proc. Natl. Acad. Sci. USA*, **108** (2011) 20953.
[34] Levandowsky M. and Winter D., *Nature*, **234** (1971) 34.
[35] Garlaschelli D. and Loffredo M.-I., *Phys. Rev. Lett.*, **93** (2004) 268701.
[36] Zhu Y.-X., Zhang X.-G., Sun G.-Q., Tang M., Zhou T. and Zhang Z.-K., *PLoS ONE*, **9** (2014) e103007.