

Characterizing Large-scale Population's Indoor Spatio-temporal Interactive Behaviors

Y-Q Zhang
Fudan University,
Shanghai, China
10210720048@fudan.edu.cn

Xiang Li^{*}
Fudan University,
Shanghai, China
lix@fudan.edu.cn

ABSTRACT

Human activity behaviors in urban areas mostly occur in interior places, such as department stores, office buildings, and museums. Understanding and characterizing human spatio-temporal interactive behaviors in these indoor areas can help us evaluate the efficiency of social contacts, monitor the frequently asymptomatic diseases transmissions, and design better internal structures of buildings. In this paper, we propose a new temporal quantity: ‘Participation Activity Potential’ (PP_A) to feature the critical roles of individuals in the populations instead of their degrees in the corresponding complex networks. Especially for the people with high degrees (hubs in the network), *Participation Activity Potential* which is directly from the statistics of their daily interactions, can easily feature the rank of their degree centrality and achieve as high as 100% accuracy rating without building the corresponding networks by high-complexity algorithms. The effectiveness and efficiency of our new defined quantity is validated in all three empirical data sets collected from a Chinese university campus by the WiFi technology, a small conference and an exhibitions by the RFID technology.

Categories and Subject Descriptors

J.4 [Social And Behavioral Sciences]: Sociology; G.3 [Probability And Statistics]: Statistical computing; C.2.m [Computer Systems Organization]: Computer-communication networks—*Miscellaneous*

General Terms

Measurement, Experiment, Performance

Keywords

Participation activity potential, complex networks, close proximity interactions, degree centrality, human dynamics, temporal networks

^{*}Correspondence and requests for materials should be addressed to X L

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UrbComp'12 August 12, 2012, Beijing, China

Copyright 2012 ACM 978-1-4503-1542-5/08/2012 ...\$15.00.

1. INTRODUCTION

With the rapidly urbanization, the populations are explosive increasing in the urban areas. More and more people's daily work, study and leisure are dependent on the public places of the urban area: subway stations, restaurants, café, museums, department stores, office buildings and etc. In these interior places, human daily interactions, e.g., face-to-face talks, hand touches, are named as human close proximity interactions (CPIs)[28], which impact our understanding of the respiratory disease transmissions, the efficiency of social contacts and the convenience of internal structures.

Since 1930's, social scientists have collected and studied human interactions and relationship data. The traditional technologies of data collections are inviting volunteers to complete questionnaires or hiring an observer to record data, and because of its time-consuming nature, the collected data is constrained to a small number of people. Nowadays, with advanced development in digital technologies, GPS, WiFi, RFID and mobile phones are widely used as sensors to locate people and build a huge volumes of spatio-temporal data, in the form of traveling trajectories[2, 15, 10, 24, 30, 17] and interactive traces[20, 4, 13, 14, 26].

Barabási, et al.[10, 24, 20] apply mobile phones to approximate human traces to mine the pattern of human mobility. GPS [2, 15, 30, 17] as a typical located technology, is widely used in tracing passengers' traveling trajectories. Barrat and their colleagues use the RFID technology to trace human CPIs in distinct rendezvous [4, 13, 14, 26] and provide a 'reinforcement dynamics' to model human face-to-face interactions[25, 31]. By using wireless sensors, Salathé, et al. [23] record an American high school students' daily interactions and evaluate their respiratory diseases infectious risks.

With the concept of 'intelligent city' becoming reality, WiFi hotspots are distributed in every corner of the urban areas. The popularity of wireless devices, such as smartphones, laptop computers and tablets, further spurs the WiFi technology (or known as 802.11b) to assist in tracing human mobilities and interactions. Therefore, we use the WiFi technology to trace a Chinese university students' interactions over 3 months. These digital 'sensors' automatically record the human CPIs without additional observers or interviewers and avoid build-in errors caused by human memory[6, 7].

The data of human CPIs were traditionally very easy to map into a static network(contact network, or named aggregated contact network with individuals as nodes and interactions between them as links) to study the embedded social phenomena. In the past more than 10 years, researchers

have witnessed the power of complex network theory to influence the spreading processes which help understand the biological and computer virus transmission. For instances, the small-world properties[29] and scale invariant power-law features of degree consequence distributions[1] can decrease the threshold of spreading processes[21, 22]. Moreover, in the scale-free physical and/or social networks, several ‘super-connectors’ (hubs in the networks) play the role of self-sustained sources to spread the infection to the rest of the system[3], which leads that there are many methods[5, 16] to find and vaccinate these ‘super-connectors’ in order to control the whole disease transmission.

However, with the development of complex network theory, the static networks as the body to study the spreading processes have faced with many new challenges[27, 18, 12, 8]. Empirically, human close proximity interactions are not stationary. A ‘super-connector’ can not simultaneously contact such huge number of neighbors. Very recently, the temporal network has become to attract researchers’ attention [11]. In this study, we use three empirical data sets, two from the RFID technology (SocioPatterns) and one from the WiFi technology to uncover the distinct features between the temporal networks and static networks. Especially, we define a new temporal quantity, ‘Participation Activity Potential’ to measure the critical roles of individuals in the corresponding networks, by which, ‘super-connectors’ in the temporal and static network can both be characterized effectively and efficiently.

The rest of the paper is organized as follows. In Section 2 we firstly introduce a series of transformed temporal networks and the corresponding aggregated static networks from three empirical data sets, and moreover, we define some preliminary concepts and notations used in this paper. In Section 3 we analyze the difference between the transmission graphs and the contact networks to investigate the typical features of ‘super-connectors’. In Section 4 we define a new temporal quantity to characterize the rank of degree centrality of individuals in two benchmark networks, and conclude the whole paper in Section 5 with future steps of work.

2. FROM HUMAN CLOSE PROXIMITY INTERACTIONS TO NETWORKS

We firstly transform human close proximity interactions (CPIs) to the corresponding networks as shown in Figure 1, a schematic representation of human CPIs of an exemplified case with the detailed descriptions see Example 1.

Example 1. There are three individuals A, B, C which have the corresponding CPIs AB, AC, BC (Figure 1(A)). Generally, these CPIs are the origin collected data, which records when and how long a pair of individuals have an interaction. Traditionally, the data is directly transformed to a contact network which is illustrated in Figure 1(D), where individuals are considered as nodes and individuals’ interactions as links. The algorithms to build a contact network are classic and we show it in Appendix (*Algorithm 1*). In Figure 1(A), it is possible to find there exists an interval, $[t_2, t_3]$, three individuals contact with each other building a clique. However, the clique is not very clearly observable in the data of human CPIs directly. Therefore, in Figure 1(B), we define ‘event interaction’ (EI) to focus on the clique. Each EI represents the interactions among all individuals in a given interval. The CPIs can be directly transformed to EIs (see

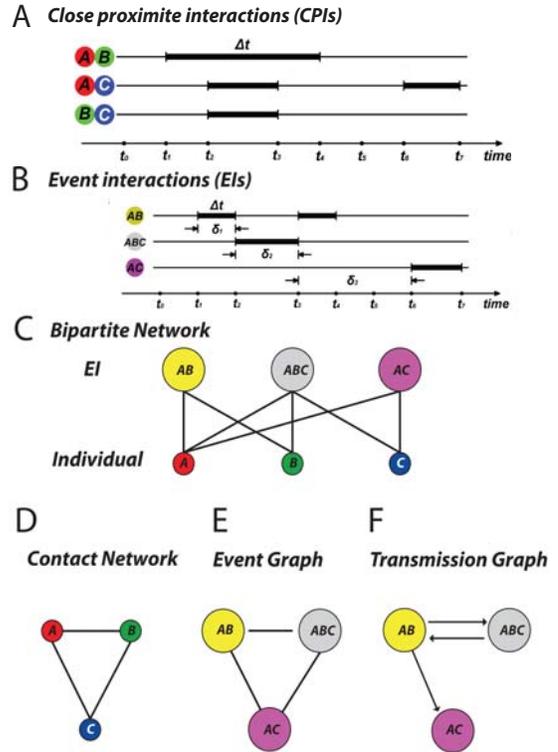


Figure 1: Schematic Representation of Human close proximity interactions (CPIs) and the corresponding (static/temporal) transformed networks.

Appendix, *Algorithm 2*). Therefore, it is easy to build a bipartite network (Figure 1(C)) exhibiting the affiliating relations between individuals and EIs. In a bipartite network, the links between individuals are based on the same EI two individuals affiliate with. As shown in Figure 1(E), an event graph is defined with the EIs: the links between the EIs are based on the same individuals two EIs both contain. Both contact networks and event graphs are static networks without temporal information when the link connects and how long the link persists to. In order to fuse temporal information, we further build a transmission graph based on the EIs. The links between the EIs consider not only their active interval, but also the order sequence of distinct EIs which is caused by the mobility of individuals contained in. The algorithm to build a transmission graph refers to *Algorithm 3* in Appendix.

To characterize the detailed human interactive behaviors, we define the following definitions.

Definition 1. The quantities of individuals:

1. *Self-activity duration:* If an individual j is active during the interval $[t^{begin}, t^{end}]$, the self-activity duration of individual j is $\Delta t^{SA}(j) = t^{end} - t^{begin}$.¹

¹As shown in Figure 1(A), during the interval of $[t_1, t_4]$, A has two records of CPIs with the overlapped duration, and A 's self-activity duration is the maximal duration of the CPIs.

2. *Self-activity frequency*: Given an observational interval T (without specially statement, T is the maximum observation interval), an individual j is active n times. Denote the self-activity frequency of individual j is $n^{SA}(j) = n$.² Therefore, the total self-activity duration of individual j is defined as

$$\Delta t_{sum}^{SA}(j) = \sum_{m=1}^{n^{SA}(j)} \Delta t_m^{SA}(j)$$

3. *Self-activity potential*: Given an observational interval T , the self-activity potential of individual j is defined as

$$P_{SA}(j) = \left[\frac{\Delta t_{sum}^{SA}(j)}{n^{SA}(j)} \right]^p n^{SA}(j) \quad (1)$$

where p is an independent variable belongs in $[0, 1]$.

Definition 2. The quantities of event interactions(EIs):

Event interaction duration: If an EI e_i is active during the interval $[t^{begin}, t^{end}]$, the event interaction duration of e_i is defined as $\Delta t^{EI}(e_i) = t^{end} - t^{begin}$

Definition 3. The quantities of a bipartite network:

1. *The rate of participation*: If an individual j participates in r event interactions, the rate of participation of individual j is defined as $r(j) = r$.
2. *Event interaction size*: If an event interaction e_i contains s individuals, the size of event interaction e_i is $s(e_i) = s$.

Definition 4. The quantities of a contact network:

Degree of individuals in the contact network: If an individual j has k distinct neighbors in the contact network, the degree of individual j is defined as $k^{CN}(j) = k$.

Definition 5. The quantities of a transmission graph and the corresponding aggregated transmission graph:

1. *Event interaction frequency*: Given an observation interval T , an EI e_i is active n times in the transmission graph. Denote the frequency of EI e_i $n^{TG}(e_i) = n$. Therefore, the total self-activity duration of EI e_i is defined as

$$\Delta t_{sum}^{EI}(e_i) = \sum_{m=1}^{n^{TG}(e_i)} \Delta t_m^{EI}(j)$$

2. *Self-activity potential*: Given an observational interval T . Define the self-activity potential of EI e_i as

$$P_{SA}^{EI}(e_i) = \left[\frac{\Delta t_{sum}^{EI}(e_i)}{n^{TG}(e_i)} \right]^a n^{TG}(e_i) \quad (2)$$

where a is an independent variable belongs in $[0, 1]$.

3. *Degree of EIs in the aggregated transmission graph*: If an event interaction (EI) e_i has been directed by k neighbors in the aggregated transmission graph, the

²The corresponding self-activity frequency is not the number of the given individual's CPIs, but the number of all unoverlapped self-activity duration.

in-degree of EI e_i is $k_{in}^{ATG}(e_i) = k$. Similarly, if the EI e_i directs to k neighbors, the out-degree of EI e_i is $k_{out}^{ATG}(e_i) = k$. Without considering the direction in the transmission path (or known as 'weakly connected'[19]), the degree of EI e_i is $k^{ATG}(e_i)$, which is equal to the number of all neighbors.

In this paper, there are three empirical data sets of human CPIs investigated. By the WiFi technology, we have collected the data of human CPIs in a Chinese campus during the 2009-2010 fall semester (3 complete months). Each student, teacher, and visiting scholar has a unique account to access the Campus WiFi system, which will automatically record their devices' MAC addresses to build a log with the MAC address of the WiFi access points(APs) they access to, and the connecting/disconnecting time as well. The logs indicate that when and where the person use the WiFi network. In the public classrooms, the spatial distance between any two individuals accessing to the same AP is as close as less than 8 meters³. From the logs of each individual, we can build human CPIs which is named as 'FudanWiFi09'.

The other two empirical data sets are both collected by the RFID technology freely achieved in the website of 'SocioPatterns' (<http://www.sociopatterns.org>), and a related research report refers to [14]. One of the data set was collected during the ACM Hypertext 2009 conference, where the 'SocioPatterns' project deployed the Live Social Semantics application. The conference attendees volunteered to wear radio badges which monitored their face-to-face interactions named as 'HT09'. The other data set contains the daily dynamic contact networks collected during the artsience exhibition 'INFECTIOUS: STAY AWAY' which took place at the Science Gallery in Dublin, Ireland, and we name it as 'SGInfectious'. In Table 1 we summarize the basic properties of these three empirical data sets.

Notice that these three data sets represent three different types of human proximity close interactions. The individuals in 'FudanWiFi09' have stable social ties: they can be classmates, teacher-student, lovers and so on. During 3 months, the interactions between the individuals may repeat in an acquaintance community. Individuals in 'HT09' are the attendees of the conference. Most of the people did not know each other. During the conference period of 3 days, they began to know each other, and the data represents the repeatable interactions in a stranger community. The individuals in 'SGInfectious' also have stable social ties, they know each other and then go with each other to visit the exhibition 'Infectious: stay away'. However, most of the involved visitors will not visit the exhibition again, which indicates that the data exhibits the unrepeatable interactions in an acquaintance community.

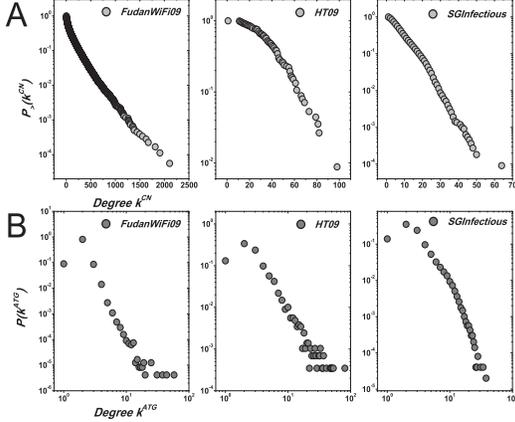
3. CONTACT NETWORKS AND TRANSMISSION GRAPHS

3.1 The differences between contact networks and transmission graphs

³In the paper 'Two Categories of Interaction Dynamics of a Large-scale Human Population in a WiFi covered university campus' (unpublished), we give the details about the spatial distance between any two individuals accessing to the same AP.

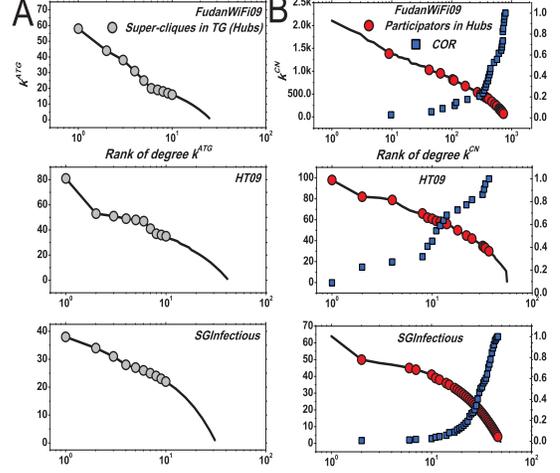
Table 1: Three empirical data sets of human close proximity interactions(CPIs)

	FudanWiFi09	HT09	SGInfectious
Area	Campus	Conference	Mesume
Technology	WiFi	RFID	RFID
Collection duration	84 days	3 days	62 days
Number of individuals	17897	113	10970
Number of CPIs	884800	9865	198198
Spatial resolution(meters)	< 8	< 2	< 2
Types of CPIs	Acquaintances with repeat	Strangers with repeat	Acquaintances without repeat


Figure 2: The degree distributions of the contact networks and the corresponding aggregated transmission graphs.

We visualize the transformed networks from the three empirical data sets in Figure 2, illustrating the difference among the contact networks (static networks) and transmission graphs (temporal networks). The cumulative probability distributions of k^{CN} in Figure 2(A) from three empirical data sets are all exponential distributions although the corresponding number of sampled individuals are different (Table 1), i.e., the contact networks are homogeneous, implying that the static patterns of human CPIs are random and well-mixed. While in the transmission graph, the probability distributions of k^{ATG} exhibit the form of power-law (see Figure 2(B)), indicating that the aggregated transmission graphs are heterogeneous, and the temporal patterns of human CPIs contain many preference linking[1].

In the study of spreading processes, scale-free networks with the form of power-law degree probability distributions have been the focus for years, where the thresholds of spreading processes are approximate to zero[21, 22]. While in a homogeneous network such as an ER random graph, there exists a fixed threshold. Recent study further states that the hubs of such networks play a self-survived role in the dynamical spreading processes[3]. Therefore, we further compare the hubs in contact networks and transmission graphs which are named as ‘super-connectors’ and ‘super-connecting groups’, respectively. Figure 3(A) illustrates the relations between the degrees of the aggregated transmission graph(ATG) and the corresponding ranks, where the top 10 nodes ranked in the ATG are the ‘super-connecting groups’. Each super-


Figure 3: The analyses of hubs in the aggregated transmission graphs and contact networks.

connecting group consists of at least two individuals. Naturally, we may regard these involved individuals as those ‘super-connectors’ in the contact networks. However, Figure 3(B) shows the opposite situation. The black line represents the relation between the degrees of individuals and the corresponding rank in the contact network, while the red circles exhibit the individuals of super-connecting groups in Figure 3(A). Strikingly, most of the members in the super-connecting groups have low degrees in the contact network. The blue rectangles represent the cumulative occurrence rate (COR) of each member, and around 90% of the individuals in the super-connecting groups are ‘leaf’ nodes in the contact network. Therefore, the degree distribution and the rank-degree relations of hubs present the obvious differences between static networks and temporal networks.

3.2 The analyses of contact network

In Definition 3, we define the participation rate of a given individual (r). Moreover in Definition 4 we define the degree of individuals in a contact network (k^{CN}). In Figure 4(A) we observe that k^{CN} increases with the growth of r . In other words, the potential of an individual becoming a ‘super-connector’ is due to his (her) participation capability.

With the defined self-activity potential PS_A of an individual, Figure 4(B) shows the individuals’ degrees in the contact network is positive correlated with their self-activity

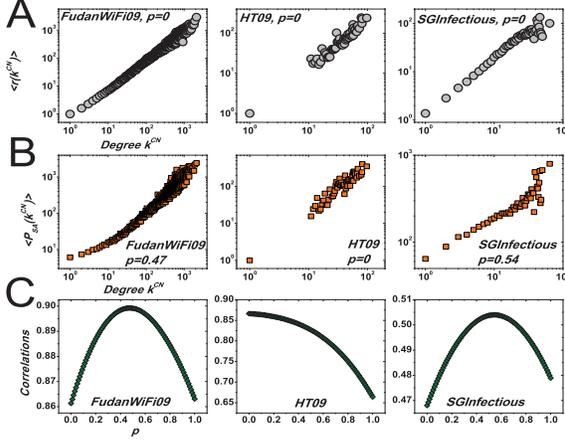


Figure 4: The individuals' self-activity potentials with their degree centrality in contact networks.

potentials, where a higher self-activity potential tells that the individual is more active. Therefore, the possibility of an individual becoming a 'super-connector' is also dependent on his (her) self-activeness.

Moreover, when $p = 0$, the self-activity potential degenerates to $P_{SA}(j) = n^{SA}(j)$, which is equal to the self-activity frequency. When $p = 1$, the self-activity potential degenerates to the total self-activity duration $\Delta t_{sum}^{SA}(j)$. Figure 4(C) exhibits the Pearson's Coefficient (PCC)[9] between the individuals' self-activity potentials $P_{SA}(j)$ and degrees k^{CN} , where p belongs in $[0, 1]$. The data of 'FudanWiFi09' and 'HT09' show their PCC can achieve as high as 0.9 and 0.85, respectively, indicating that the self-activity potentials of individuals have a very strong positive correlation with their degrees in the contact network. However, the PCC of the 'SGInfectious' data is less than 0.5, whose correlation is weaker. As summarized in Table 1, the types of CPIs of three empirical data sets are different: 'FudanWiFi09' and 'HT09' are the repeatable interactions while 'SGInfectious' represents unrepeatable interactions. In other words, the individuals in 'FudanWiFi09' and 'HT09' both build a time-invariant community, which does not hold in the case of 'SGInfectious'. In a time-invariant community, a larger self-activity potential may help the individual increase the possibility of contacting with other individuals. While in a time-variant community, new individuals join the community along with the old individuals leave, and every individual has an active lifetime, out of which the individuals will lose the capability to contact with other individuals in the community. Therefore, we conclude that in such (static) contact networks, super-connectors own two essential characteristics: high self-activity and sociability (participation capability). The 'leaf' individuals in the contact network are those with low self-activity and sociability. However, why can these 'leaf' individuals gather into the super-connecting groups in the (temporal) transmission graph? we need further explore the temporal transmission graphs in more detail.

4. PARTICIPATION ACTIVITY POTENTIAL

In a transmission graph, an EI has a self-activity potential P_{SA}^{EI} and a degree k^{ATG} . The crucial role of an individual in the transmission graph is featured by the degree of the corresponding EI. That is to say, when an individual participates in an EI having a higher degree, the individual is more important in the temporal network. However, since most of the individuals' participation rates are larger than one, they may participate in r EIs, whose degrees are r at most. Since the role of individuals in a temporal network is featured by the maximal degree, we define a new definition of the individuals' maximal participation degree as follows:

Definition 6. Maximal participation degree κ_{max} : If an individual has the participation rate r , he (she) participates in r EIs: e_1, e_2, \dots, e_r , with the corresponding degrees in the ATG as $k_1^{ATG}, k_2^{ATG}, \dots, k_m^{ATG}$ ($m \leq r$), respectively, the maximal participation degree $\kappa_{max} = \max(k_1^{ATG}, k_2^{ATG}, \dots, k_m^{ATG})$

4.1 Maximal Participation Activity Potential

Given an individual j participates in an EI e_i ($j \in e_i$). Define the participation activity potential of individual j with EI e_i as follows:

$$P_{PA}(j, e_i) = P_{SA}^{EI}(e_i) = \left[\frac{\Delta T_{sum}^{EI}(j, e_i)}{n^{TG}(j, e_i)} \right]^a n^{TG}(j, e_i) \quad (3)$$

The maximal participation activity potential of individual j comes:

Definition 7. Maximal participation activity potential: An individual j participates in a set of EIs

$$\Gamma(j) = [e_1, e_2, \dots, e_i, \dots, e_{r(j)}]$$

The corresponding participation activity potential is:

$$\zeta^{PPA}(j) = [P_{PA}(j, e_1), \dots, P_{PA}(j, e_{r(j)})]$$

and, the maximal participation activity potential of individual j is:

$$P_{PA}^{max}(j) = \max(\zeta^{PPA}(j))$$

From Figure 5(A), we observe that the maximal participation activity potential P_{PA}^{max} increases with the growth of the maximal participation degree κ_{max} , indicating that the maximal participation activity potential features the maximal degree of the individuals in the ATG. In Appendix (Algorithm 3), we provide the algorithms of building transmission graph, and the computational complexity of the algorithms is $O(M^2)$ (M is the size of *EITimeTable*). However, the participation activity potential of individuals is identical with the self-activity potential of EIs, while the computational complexity of algorithms to calculate the self-activity potential is $O(N)$ (N is the size of EIs and $N \ll M$). If we can use the maximal self-activity potential to feature the critical role of individual replacing their degree centrality, we can dramatically reduce the computational complexity.

We further illustrate in Figure 5(B) the accuracy rating to feature the rank of κ_{max} by P_{PA}^{max} . All three empirical data sets show that the accuracy rating increases with the growth of κ_{max} , indicating that the maximal participation activity potential can feature the members of super-connecting cliques with the achieved accuracy rate as high as 100%.

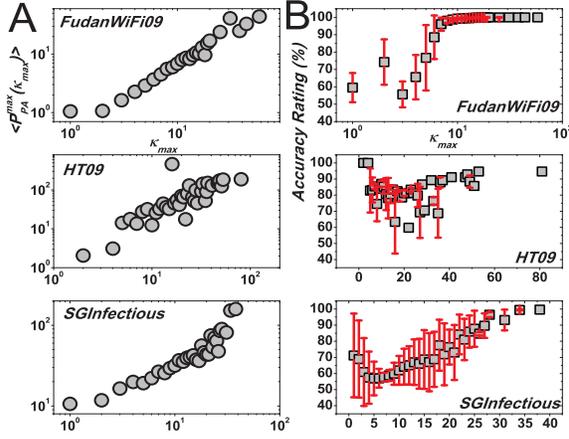


Figure 5: The maximal participation activity potential of three collected data sets.

4.2 Sum of Participation Activity Potential

We further define the sum of participation activity potential of a given individual j as follows:

Definition 8. Sum of participation activity potential P_{PA}^{sum} : Given individual j participates in a set of EIs:

$$\Gamma(j) = [e_1, e_2, \dots, e_i, \dots, e_{r(j)}]$$

The corresponding participation activity potential is

$$P_{PA}(j, e_1), \dots, P_{PA}(j, e_{r(j)})$$

Therefore the total participation activity potential of individual j is defined as:

$$P_{PA}^{sum}(j) = \sum_{m=1}^{r(j)} P_{PA}(j, e_m) \quad (4)$$

Equation 4 shows that when $a = 1$, the sum of participation activity potential degenerates to the total participation duration

$$P_{PA}^{sum}(j) = \sum_{m=1}^{r(j)} \Delta T_{sum}^{EI}(j, e_m)$$

From Figure 1, we find that the total self-activity duration is identical with the sum of participation duration,

$$P_{SA}(j, p = 1) = P_{PA}^{sum}(j, a = 1)$$

While the definition of P_{SA}^{sum} also includes the participation rate and the activity of participation. Therefore, the sum of participation activity potential can replace the self-activity potential to feature the rank of degrees of individuals in the contact network.

As shown in Figure 6(A), in all three data sets, the sum of participation activity potential increases with the growth of corresponding degrees in contact networks, which indicates that super-connectors in the contact networks have higher sum of participation activity potentials. Moreover, we examine the accuracy rate of measuring rank of degrees in the

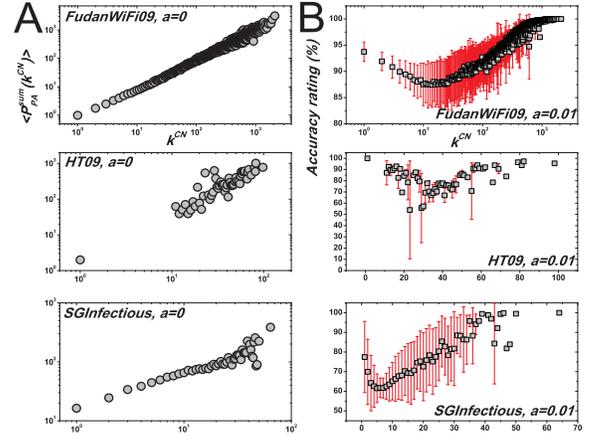


Figure 6: The sum of participation activity potential of three collected data sets.

contact networks by the sum of participation activity potential. In Figure 6(B) all of three empirical data sets support the findings that the sum of participation activity potential can reach the accurate rate as high as 100% to measure the super-connectors. Therefore, the critical role of individuals in the contact networks can be characterized by the sum of participation activity potential.

As discussed above, the maximal participation activity potential of a given individual features the corresponding rank of degree centrality in the transmission graph, and the sum of participation activity potential features the corresponding rank of degree centrality in the contact network. Sometimes, the individual with a high maximal participation activity potential also has the high sum of participation activity potential, e.g., some super-connectors in ‘HT09’ and ‘SGInfectious’. However, most of the individuals with high maximal participation activity potentials do not have the corresponding high sum of participation activity potentials, leading that most of the members of ‘super-connecting groups’ are not the ‘super-connectors’, as shown in Figure 3(B).

5. CONCLUSIONS

In this paper, we have defined a new ‘actor-related’ quantity: *participation activity potential* which can characterize their structural relations. The maximal participation activity potential of the individuals feature their rank of degree centrality in the corresponding temporal network, achieving the accurate rate as high as 100% to measure the members of super-connecting groups. Besides, the sum of participation activity potential is effective to characterize the super-connectors in the contact network. In the view of more detailed spreading processes such as infectious disease prevalence and computer virus propagation, the members of super-connecting groups play a critical role to the spread over the whole networking system, therefore, this new proposed quantity deserves further extensive efforts to understand its significance in dynamical spreading processes of

urban networking systems in near future.

Acknowledgments

We were grateful to the assistance from Bochun Wu, and the Informatization Office of Fudan University for the WiFi Data collection. We acknowledge the SocioPatterns to provide the data of human close proximity interactions collected by the RFID technologies. This work was partly supported by the National Key Basic Research and Development Program (No.2010CB731403) and the NCET program (No.NCET-09-0317) of China.

6. REFERENCES

- [1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [2] Y. Bu, L. Chen, A.-C. Fu, and D. Liu. Efficient anomaly monitoring over moving object trajectory streams. In *Proceedings of the 15th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'09)*, pages 159–168, 2009.
- [3] C. Castellano and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Phys Rev Lett*, 105:218701, 2010.
- [4] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PLoS ONE*, 5(7):e11596, 2010.
- [5] R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Phys Rev Lett*, 91:247901, 2003.
- [6] N. Eagle. *Machine perception and learning of complex social systems*. PhD thesis, Massachusetts Institute of Technology, 2005.
- [7] N. Eagle, A. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *P Natl Acad Sci U S A*, 106(36).
- [8] J. Fernández-Gracia, V. M. Eguíluz, and M. S. Miguel. Update rules and interevent time distributions: Slow ordering versus no ordering in the voter model. *Phys Rev E*, 84:015103, 2011.
- [9] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. The human disease network. *P Natl Acad Sci U S A*, 104(21).
- [10] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.
- [11] P. Holme and J. Saramäki. Temporal networks. *Phys Rep* in press, 2012.
- [12] J. L. Iribarren and E. Moro. Impact of human activity patterns on the dynamics of information diffusion. *Phys Rev Lett*, 103:038702, 2009.
- [13] L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. Van den Broeck, F. Gesualdo, E. Pandolfi, L. Ravà, C. Rizzo, and A. E. Tozzi. Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE*, 6(2):e17144, 2011.
- [14] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *J Theor Biol*, 271:166–180, 2010.
- [15] J.-G. Lee, J. Han, and X. Li. Trajectory outlier detection: A partition-and-detect framework. In *Proceedings of the 24th International Conference on Data Engineering (ICDE'08)*, pages 140–149, 2008.
- [16] S. Lee, L. E. C. Rocha, F. Liljeros, and P. Holme. Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS ONE*, 7(5):e36439, 2012.
- [17] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, page 34, 2008.
- [18] B. Min, K.-I. Goh, and A. Vazquez. Spreading dynamics following bursty human activity patterns. *Phys Rev E*, 83:036102, 2011.
- [19] V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, and V. Latora. Components in time-varying graphs. <http://arxiv.org/abs/1106.2134>, 2012.
- [20] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *P Natl Acad Sci U S A*, 104:18, 2007.
- [21] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys Rev Lett*, 86(14).
- [22] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys Rev E*, 63:066117, 2001.
- [23] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones. A high-resolution human contact network for infectious disease transmission. *P Natl Acad Sci U S A*, 107(51).
- [24] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327:1018–1021, 2010.
- [25] J. Stehlé, A. Barrat, and G. Bianconi. Dynamical and bursty interactions in social networks. *Phys Rev E*, 81:035101(R), 2010.
- [26] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaghiotto, W. Van den Broeck, C. Régis, B. Lina, and P. Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*, 6(8):e23176, 2011.
- [27] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási. Impact of non-poissonian activity patterns on spreading processes. *Phys Rev Lett*, 98:158702, 2007.
- [28] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.
- [29] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- [30] X. Xiao, Y. Zheng, Q. Luo, and X. Xi. Finding similar users using category-based location history. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 442–445, 2010.
- [31] K. Zhao, J. Stehlé, G. Bianconi, and A. Barrat. Social

APPENDIX

A. ALGORITHMS

A.1 Algorithms to build a contact network

Algorithms 1. CNBuilder: building a contact network from the records of human close proximity interactions.

Input: *CPIs*: a set of records about human close proximity interactions (Size: S) with the formalization as $(N_1, N_2, t_{start}, t_{end})$.

Output: *CN*: a list of pairwise nodes.

```

1:  $CN \leftarrow$  an empty set  $\{\}$ ;
2: for Each record of CPIs  $i$  ( $i \in (1, \dots, S)$ ) do
3: If the pairwise nodes in record  $i$  are not included in  $CN$ 

```

```

4:  $CN \leftarrow CN \cup (N_1(j), N_2(j))$ 

```

```

5: end

```

```

6: end for

```

```

7: return  $CN$ ;

```

We assume the size of CN as n , then the computational complexity of the algorithm in the worst case is $O(\frac{(0+(n-1))n}{2} + (S-n)n) \approx O(Sn)$

A.2 Algorithms to build event interactions and a transmission graph

Algorithms 2. EIBuilder: building event interactions from human close proximity interactions.

Input: *CPIs*: a set of records about human close proximity interactions (Size: S) with the formalization as $(N_1, N_2, t_{start}, t_{end})$.

Output: *EITimeTable*: a list of starting and ending time of EIs. *EIs*: a list of event interactions.

```

1:  $timeList \leftarrow$  an empty set  $\{\}$ ;
2: for Each record of CPIs  $i$  ( $i \in (1, \dots, S)$ ) do
3: if the starting and ending time of CPI  $i$  are not including in  $timeList$  then

```

```

4:  $timeList \leftarrow timeList \cup (t_{start}(j), t_{end}(j))$ 

```

```

5: end if

```

```

6: end for

```

```

7: Return  $timeList$ ;

```

```

8: Order  $timeList$  by time ascend.

```

```

9:  $EITimeTable \leftarrow$  a empty set  $\{\}$ ;

```

```

10: for Each record of  $timeList$   $j$  ( $j \in (1, \dots, M)$ )4 do

```

```

11:  $nodeset \leftarrow$  an empty set  $\{\}$ ;

```

```

12: for Each record of CPIs  $i$  ( $i \in (1, \dots, S)$ ) do

```

```

13: if the record  $i$  of CPIs with  $t_{start} \leq t_j$  and  $t_{end} \gg t_{j+1}$  and nodes of  $i$  are not included in  $nodeset$  then

```

```

14:  $nodeset \leftarrow nodeset \cup N_1(i) \cup N_2(i)$ ;

```

```

15: end if

```

```

16: end for

```

```

17:  $EITimeTable \leftarrow EITimeTable \cup nodeset \cup (t_j, t_{j+1})$ ;

```

```

18: end for

```

```

19: Return  $EITimeTable$ ;

```

```

20:  $EIs \leftarrow$  an empty set  $\{\}$ ;

```

```

21: for Each record of EITimeTable  $k$  ( $k \in (1, 2, \dots, M)$ ) do

```

```

22: if the nodeset of record  $k$  is not included in  $EIs$  then

```

```

23:  $EIs \leftarrow EIs \cup nodeset$  of  $k$ ;

```

```

24: end if

```

```

25: end for

```

```

26: Return  $EIs$ ;5

```

The computational complexity of the algorithm to build $timeList$ in the worst case is $O(MS)$. The computational complexity of the algorithm to sort $timeList$ in the worst case is $O((M+1)^2)$. The computational complexity of the algorithm to build *EITimeTable* in the worst case is $O(SM)$. The computational complexity of the algorithm to build *EIs* in the worst case is $O(MN)$. Therefore, the computational complexity of the algorithm to build *EIs* from *CPIs* in the worst case is $O(MS)$

Algorithms 3. TGBuilder: building a transmission graph and the corresponding aggregated transmission graph from the *EITimeTable*

Input: *EITimeTable*.

Output: *TG*: the transmission path from one EI to another. *ATG*: the aggregated transmission graph.

```

1:  $TG \leftarrow$  an empty set  $\{\}$ ;

```

```

2: sort EITimeTable by the ascending order of the starting time.

```

```

3: for Each record of EITimeTable  $e$  ( $e \in (2, \dots, M)$ ) do

```

```

4:  $bridgenodes \leftarrow$  an empty set  $\{\}$ ;

```

```

5: for Each record of EITimeTable  $e_i$  ( $e_i \in (e, \dots, 1)$ ) do
6: if there are common nodes both in record  $e$  and record  $e_i$ , while the common nodes are not included in bridgenodes then

```

```

7:  $TG \leftarrow TG \cup [nodeset$  of  $e_i$ ,  $nodeset$  of  $e$ ,  $t_{start}(e_i)$ ,  $t_{end}(e_i)$ ,  $t_{start}(e)$ ,  $t_{end}(e)]$ ;

```

```

8:  $bridgenodes \leftarrow bridgenodes \cup nodeset$  of  $e_i$ ;

```

```

9: if the size of bridgenodes is equal to the size of  $nodeset$  of  $e_i$  then

```

```

10: break;

```

```

11: end if

```

```

12: end if

```

```

13: end for

```

```

14: end for

```

```

15: Return  $TG$ ;

```

```

16:  $ATG \leftarrow$  an empty set  $\{\}$ ;

```

```

17: for Each record of  $TG$   $t$   $t \in (1, 2, \dots, P)$  do

```

```

18: if the pairwise nodesets of record  $t$  are not included in  $ATG$  then

```

```

19:  $ATG \leftarrow ATG \cup$  the pairwise nodesets of record  $t$ ;

```

```

20: end if

```

```

21: end for

```

```

22: Return  $ATG$ 6;

```

The computational complexity of the algorithm to build a TG in the worst case is $O(M^2)$ and the computational complexity of the algorithm to build an ATG in the worst case is $O(PH)$.

Regarding event interactions as vertices, we further introduce the following rules to link the successive vertices as the edges of the transmission graph: a) in the time series, a source EI is the closest EI prior to the sink EI; b) at least one user coexists in the source and sink EI; c) when there are several sources before one sink, any set of the shared users between the given source and sink EIs never intersect with each other(set).

⁵ N represents the size of EIs.

⁶ H is the size of ATG

⁴ $M+1$ represents the size of $timeList$.