

Uncovering offline event similarity of online friends by constructing null models

Wenkuo Cui(崔文阔), Jing Xiao(肖婧), Ting Li(李婷), Xiaoke Xu(许小可)

Citation: Chin. Phys. B . 2019, 28(6): 068901. **doi:** 10.1088/1674-1056/28/6/068901
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Traffic dynamics considering packet loss in finite buffer networks

Jie Chen(陈杰), Jin-Yong Chen(陈金邕), Ming Li(李明), Mao-Bin Hu(胡茂彬)

Chin. Phys. B . 2019, 28(4): 048901. **doi:** 10.1088/1674-1056/28/4/048901

Time-varying networks based on activation and deactivation mechanisms

Xue-Wen Wang(王学文), Yue-E Luo(罗月娥), Li-Jie Zhang(张丽杰), Xin-Jian Xu(许新建)

Chin. Phys. B . 2017, 26(10): 108902. **doi:** 10.1088/1674-1056/26/10/108902

Link prediction in complex networks via modularity-based belief propagation

Darong Lai(赖大荣), Xin Shu(舒欣), Christine Nardini

Chin. Phys. B . 2017, 26(3): 038902. **doi:** 10.1088/1674-1056/26/3/038902

Multiple-predators-based capture process on complex networks

Rajput Ramiz Sharafat, Cunlai Pu(濮存来), Jie Li(李杰), Rongbin Chen(陈荣斌), Zhongqi Xu(许忠奇)

Chin. Phys. B . 2017, 26(3): 038901. **doi:** 10.1088/1674-1056/26/3/038901

Entropy-based link prediction in weighted networks

Zhongqi Xu(许忠奇), Cunlai Pu(濮存来), Rajput Ramiz Sharafat, Lunbo Li(李伦波), Jian Yang(杨健)

Chin. Phys. B . 2017, 26(1): 018902. **doi:** 10.1088/1674-1056/26/1/018902

Uncovering offline event similarity of online friends by constructing null models*

Wenkuo Cui(崔文阔)[1], Jing Xiao(肖婧)[1], Ting Li(李婷)[1], and Xiaoke Xu(许小可)[1,2,†]

[1] College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China
[2] Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China

(Received 8 December 2018; revised manuscript received 28 February 2019; published online 8 May 2019)

The emergence of Event-based Social Network (EBSN) data that contain both social and event information has cleared the way to study the social interactive relationship between the virtual interactions and physical interactions. In existing studies, it is not really clear which factors affect event similarity between online friends and the influence degree of each factor. In this study, a multi-layer network based on the Plancast service data is constructed. The the user's events belongingness is shuffled by constructing two null models to detect offline event similarity between online friends. The results indicate that there is a strong correlation between online social proximity and offline event similarity. The micro-scale structures at multi-levels of the Plancast online social network are also maintained by constructing 0k–3k null models to study how the micro-scale characteristics of online networks affect the similarity of offline events. It is found that the assortativity pattern is a significant micro-scale characteristic to maintain offline event similarity. Finally, we study how structural diversity of online friends affects the offline event similarity. We find that the subgraph structure of common friends has no positive impact on event similarity while the number of common friends plays a key role, which is different from other studies. In addition, we discuss the randomness of different null models, which can measure the degree of information availability in privacy protection. Our study not only uncovers the factors that affect offline event similarity between friends but also presents a framework for understanding the pattern of human mobility.

Keywords: event-based social networks (EBSNs), event similarity, null models, event chains

PACS: 89.75.Hc, 02.50.Tt, 89.20.Ff **DOI:** 10.1088/1674-1056/28/6/068901

1. Introduction

Online social network services have become the fastest growing applications on the Internet. On these virtual networks, users communicate with each other and share all kinds of information. Actually, the newly appeared data sources, like Event-based Social Networks (EBSNs) data[1] and Location-based Social Networks (LBSNs) data,[2] contain not only online virtual interactions as traditional social networks,but also offline physical interactions of users, which make it possible to combine virtual interactions with physical social ones. How to make use of this kind of interactive relationship to provide users with more convenience is one of the hotspot researches in recent years. As of March 2017, Plancast had more than 50 million registered users and published an average of 10^4 events a day. An event is a special case that is caused by many reasons and conditions, occurs at a certain time and place, and may be accompanied by certain inevitable results. Understanding event similarity has direct potential applications for recommending all kinds of services,[3] building smart cities,[4] and relieving traffic pressures[5]. Therefore, the relevant studies of social networks and their applications are of great significance.[6–9]

The traditional calculation of event similarity is based on event six-section definitions, which only involve the action, object, environment, assertion, language expression, and time factor of the event element itself.[10] However, this neglects the influence of other external factors on event similarity. Therefore, the results are not comparable between different studies. In this work, we use the method by constructing null models, which accurately uncovers the factors that influence the offline event similarity between friends.[11,12] Null networks can provide an accurate reference for the original network, and can accurately describe the non-trivial characteristics of the original network combined with the statistic indicators, which can help us to reveal the origin and the level of complexity. The term 'null model' was proposed by Colwell *et al.* in a conference. Generally, a randomized network with some of the same properties as the real-life network, is called a null network of the original network.[13] Null networks have been widely applied in analyzing clustering coefficient,[14,15] degree distribution,[16] link prediction,[17,18] and community detection.[19,20] In this study, all null networks are generated by a random rewiring algorithm, which randomly shuffles the edges of the original network and makes the original network as random as possible.[21,22]

EBSNs contain not only online social relationship of friends but also their offline events information in actual life. The Plancast is a typical event-based network and the dataset

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61773091, 61603073, 61601081, and 61501107) and the Natural Science Foundation of Liaoning Province, China (Grant No. 201602200).

†Corresponding author. E-mail: xuxiaoke@foxmail.com

© 2019 Chinese Physical Society and IOP Publishing Ltd

is big enough for our purposes. The online social network contains 76665 nodes and 1702058 edges, and the offline network contains 401634 events. In existing studies of the Plancast network, it is not very clear which factors affect event similarity between friends and the influence degree of each factor. For example, we do not know whether there is a relationship between online friends and their offline events. If so, how the online social network affects offline event similarity.[23] Furthermore, the influence of the number and subgraph structure of common neighbors on offline event similarity is not clear.[24,25]

In this study we constructed a double-layer network by utilizing the Plancast data, and measured offline event similarity by Hub Promoted Index (HPI).[26,27] The offline event similarity in Plancast was first detected by constructing two null models which changed the belongingness of events and calculated HPI of each null model. We found that the events of friends showed a strong similarity; that is, once we changed the belongingness of events, and the event similarity of friends would be greatly reduced, which was indicated that there is a strong correlation between online social proximity and offline event similarity. Then, the micro-scale structures at multi levels were maintained by constructing $0k$–$3k$ null networks to study how these characteristics (average degree, degree distribution, assortativity, and clustering coefficient) affected offline event similarity. We found that average degree and degree distribution ($0k$ and $1k$ characteristics) were not enough to maintain offline event similarity, but assortativity ($2k$ characteristic) is a significant characteristic for maintaining offline event similarity. Finally, by calculating the HPI of different number and structure of common friends, we found that the subgraph structure of common friends had no positive impact on event similarity while the number of common friends played a key role, which was different from the results reported in Ref. [25]. In addition, we discussed the randomness of different null models, which can measure the degree of information availability in privacy protection.

2. Detecting offline event similarity in Plancast

2.1. Constructing the multi-layer Plancast network

Recently, event-based online social services, such as Plancast and Eventbrite, have experienced rapid growth. From these services, researchers observe a new type of social network, which makes it possible to combine virtual and physical social interactions. Figure 1 depicts the constructing principle of EBSN based on the online social services data. In Fig. 1, A, B, C, and D represent four users of Plancast services, and $\{e_1, e_2, e_3\}$ are the events that these users participate in. The social relationships between users in Plancast service in Fig. 1(a), forming the online network of the Plancast EBSN, are mapped to green links of Fig. 1(b). Meanwhile, the same events that users co-participate in Plancast service in Fig. 1(a), deriving their offline social connections, collectively form an offline social network in the Plancast EBSN, denoted as blue edges in Fig. 1(b).

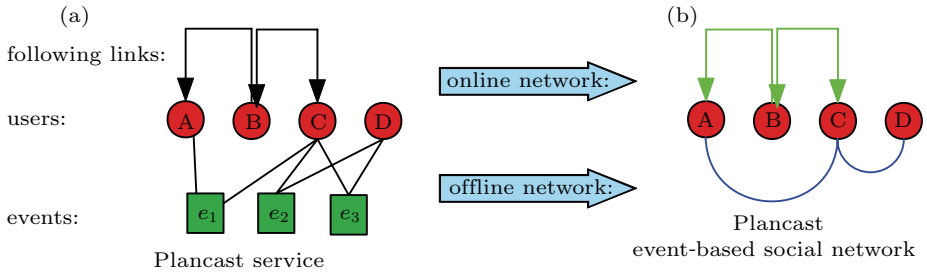


Fig. 1. An example of event-based social service and its corresponding EBSN: (a) the Plancast service, and (b) its corresponding Plancast EBSN.

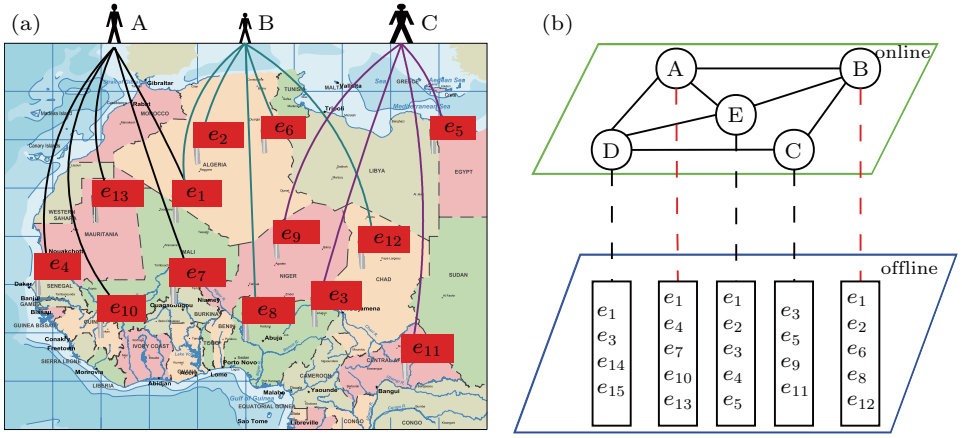


Fig. 2. A schematic illustration of a multi-layer network: (a) the relationship between users and their events in Plancast service, and (b) its corresponding toy model of a special multi-layer network. The online layer represents the social relationship between friends, and the offline layer indicates the chain of events that they participate in.

In this study a multi-layer network is constructed by utilizing the Plancast data, and the constructing principle is shown in Fig. 2. In Fig. 2(a), every user can participate in many different events, while one event can be participated by multiple users. For example, $\{e_1, e_4, e_7, e_{10}, e_{13}\}$ are the events that user A participates in, and e_1 is also participated by user B. A toy model of the special multi-layer network is shown in Fig. 2(b). The Plancast is divided into the online and offline layers. In the online layer, A, B, C, D, and E represent five users. If two users are friends, then there is an edge between them. The offline layer is composed of user events, while the dotted line indicates the corresponding relationship between online users and their event chains.[28–30]

2.2. Indexes to measure offline event similarity

If two users are involved in more events together, then they will be considered to be more similar. In recent years, researchers have developed a number of indexes to measure similarity, such as Jaccard index, Hub Depressed Index (HDI), and Hub Promoted Index (HPI).[26,27] These are common indexes in analyzing user similarity. For a node x, let k_x represent the degree of x and $\Gamma(x)$ denote the set of neighbors of x. The definition of Jaccard, HDI, and HPI are as follows:

$$S_{xy}^{\text{Jaccard}} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}, \tag{1}$$

$$S_{xy}^{\text{HDI}} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k_x k_y}}, \tag{2}$$

$$S_{xy}^{\text{HPI}} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\min\{k_x, k_y\}}, \tag{3}$$

2.3. Null models of exchanging trajectory

Like conventional online social networks, EBSNs provide an online virtual world where users exchange thoughts and share experiences. What distinguishes EBSNs from conventional social networks is that EBSNs also capture the effect of face-to-face social interactions in participating events in the offline physical world. To study the relationship between online social proximity and offline event similarity, we introduce two kinds of null models, which can not only change offline event chains but also maintain the social relationship in online social networks.

The construction of the null model of exchanging trajectory is shown in Fig. 3. Two users from all users are randomly selected. Then, the two event chains of users are obtained. Finally, the offline event marked by them is interchanged. For example, we randomly select two users A and B, and obtain their events in Fig. 3(a), and then we interchange the event chain marked by them. The result of the reconnection is shown in Fig. 3(b), in which the events of A are changed to $\{e_1, e_2, e_6, e_8, e_{12}\}$ and the events of B are changed to $\{e_1, e_4, e_7, e_{10}, e_{13}\}$. The process of randomly rewiring should be repeated until a completely randomized null network is gained.

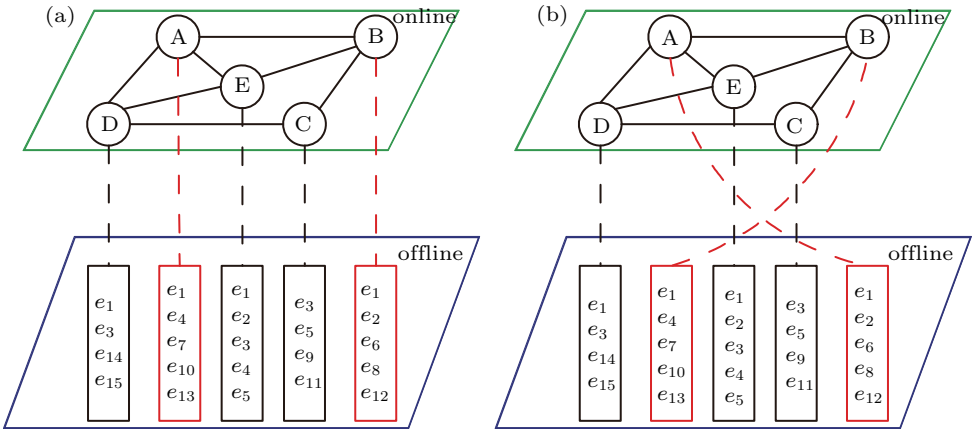


Fig. 3. Constructing the null model of exchanging trajectory: (a) the original toy model, and (b) its corresponding null model of exchanging trajectory.

2.4. Null models of exchanging user events

The construction of the null model of exchanging user events is shown in Fig. 4. The first step is the same as the null model of exchanging trajectory. Then, we obtain the event chains of two users and the number of events that users participate in, respectively. Finally, we randomly select events which have the same number as the original two users from all the events to them respectively. For example, two users A and B are randomly selected, and then we obtain that A has five events $\{e_1, e_4, e_7, e_{10}, e_{13}\}$ and B has five events $\{e_1, e_2, e_6, e_8, e_{12}\}$, and all of the event information contain 9 events. Finally, we randomly select five events from all the events to A and randomly select five events to B. The result of the reconnection is shown in Fig. 4(b), and three events $\{e_2, e_6, e_{12}\}$ of A and four events $\{e_4, e_7, e_{10}, e_{13}\}$ of B are different from their original events. The process of randomly rewiring should be repeated until a completely randomized null network is obtained.

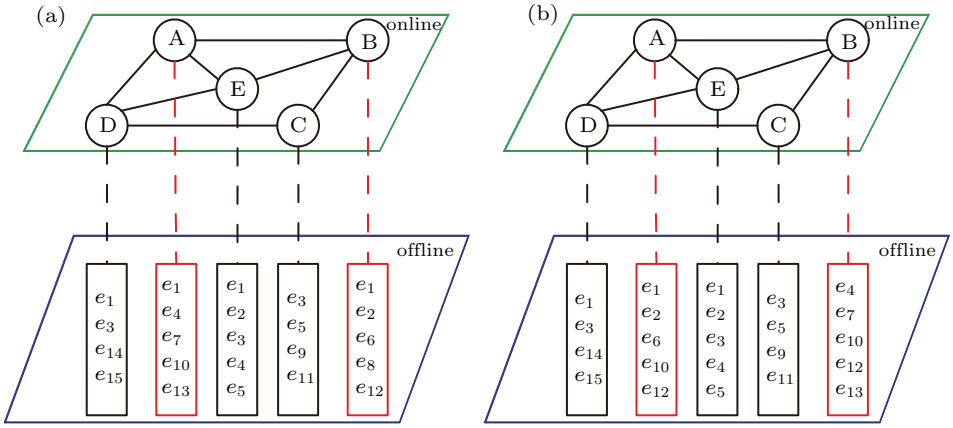


Fig. 4. Constructing the null model of exchanging user events: (a) the original toy model, and (b) its corresponding null model of exchanging user events.

The null model of exchanging trajectory does not change the order of the event sequence but shuffles the ownership of the whole trajectory for each user. The null model of exchanging user events not only changes the ownership of a whole trajectory but also shuffles the set of events belong to each user. Both null models can change the offline event characteristics, and detect the correlation between offline events similarity and online social proximity. This correlation plays an important role in coordinating online and offline resources, such as recommending offline events based on online relationships.

2.5. Offline event similarity between friends

In this study, the indexes above-mentioned (Jaccard, HDI, and HPI) have been applied and the results are similar.

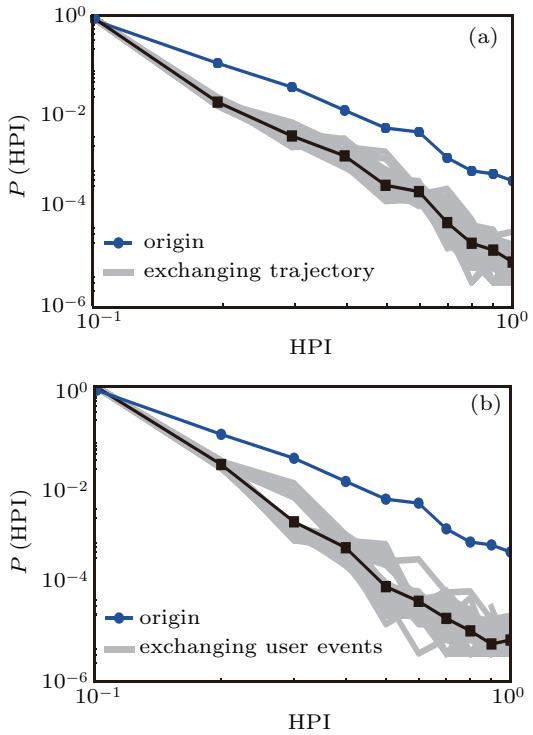


Fig. 5. (a) HPI distribution of null models of exchanging trajectory and (b) HPI distribution of null models of exchanging user events. The x coordinate represents HPI values, and the y coordinate represents the probability of HPI values.

HPI is more observable, so the result of event similarity is shown by HPI in this study. The HPI distribution of the original network and its null models of exchanging trajectory are shown in Fig. 5(a). The blue-dotted line represents the HPI distribution of the original network, and the grey lines indicate the results of 100 null models of exchanging trajectory. To demonstrate the experimental results more clearly, we compute the average HPI value of 100 null models and expressed the HPI distribution with the square line. All of the HPI distribution of null models are less than that of the original network because this type of null model changes the belongingness of the original event chains. The results indicate that there is a strong correlation between online social relationship and offline event similarity. The same method is used to get the HPI distribution of the original network and null models of exchanging user events, which is shown in Fig. 5(b). We obtained the same conclusion as the null models of exchanging trajectory.

3. The impact of online subgraph structures on offline event similarity

The traditional methods have simply studied event similarity by calculating the similarity between the factors of event itself. However, these methods cannot meet the research requirements of actual complex systems because event similarity is often affected by other factors, such as micro-scale structures of online social networks. Actually, $0k$–$3k$ null networks can maintain different micro-scale characteristics (such as average degree, degree distribution, assortativity, and clustering coefficient) of the original network. The null networks of different orders are interrelated, that is, $0k \supseteq 1k \supseteq 2k \cdots\cdots (n-1)k \supseteq nk$. Any nk null network will contain the characteristics of $(n-1)k$ null network.[9,10] Therefore, we uncover how micro-scale structures affect offline event similarity by constructing $0k$–$3k$ null networks.

3.1. $0k$ null network of EBSN

The $0k$ null network is the simplest and most randomized null model, which only possesses the same number of nodes

and the average degree as a given graph $G(V,E)$. Here, V represents the set of nodes, and E represents the set of edges. The average degree refers to the average value of all the nodes' degrees. Suppose n ($n = |V|$) as the number of nodes and m ($m = |E|$) as the number of edges, and the definition of the average degree $\langle k \rangle$ is as follows:

$$\langle k \rangle = \frac{2m}{n}. \tag{4}$$

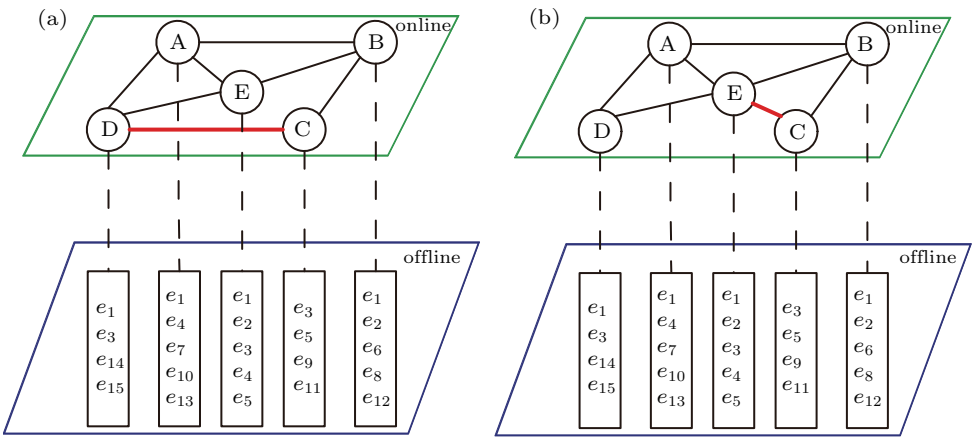


Fig. 6. The constructing process of the 0k null network. We remove link C–D, then add link C–E in the online network: (a) the original toy model, and (b) its corresponding 0k null network.

While constructing the 0k null network for EBSN, the links in the offline network remain unchanged while those in the online network should be shuffled. For example, we randomly select a link C–D in the original network and disconnect it as shown in Fig. 6(a). We then randomly select two nodes C and E. If the two nodes are not connected, we then connect them. The result of the reconnection is shown in Fig. 6(b). The process of randomly rewiring should be repeated until a completely randomized null network is achieved.

3.2. 1k null networks of EBSN

The 1k null network possesses the same degree distribution (or sequence) as the original network. The degree distribution refers to the probability of nodes' degree in the original network. If $n(k)$ is the number of nodes with degree k, then the degree distribution $p(k)$ is defined as

$$p(k) = \frac{n(k)}{n}. \tag{5}$$

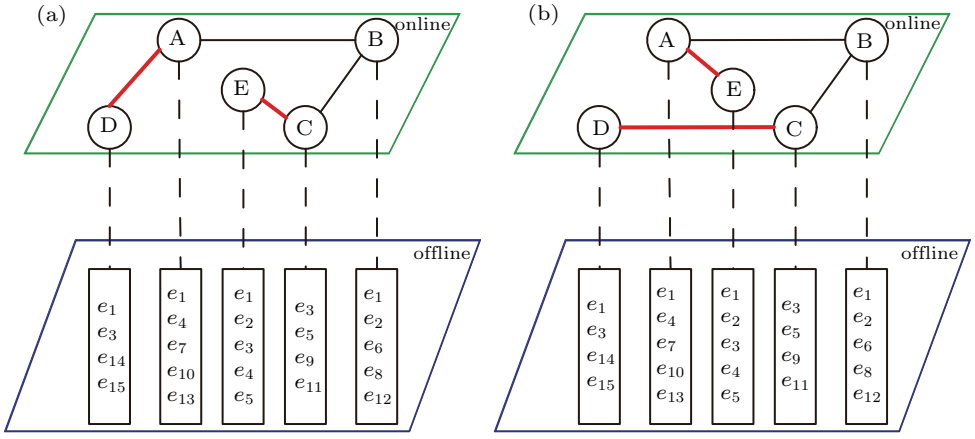


Fig. 7. The construction process of the 1k null network. We remove links A–D and C–E, and then add links A–E and C–D: (a) the original toy model, and (b) its corresponding 1k null network.

While constructing the 1k null network of EBSN, links in the offline network remain unchanged while those in the online network should be shuffled. For example, referring to the constraints of generating 1k null network, we randomly select two links A–D and C–E in the online network and disconnect the links as shown in Fig. 7(a), and then we connect A–E and C–D. The result of the reconnection is shown in Fig. 7(b). The process of random rewiring should be repeated until a completely randomized null network is obtained.

3.3. 2k null network of EBSN

The 2k null network possesses the same joint degree distribution as the original network. The joint degree distribution refers to the number of degree values (probability) of the nodes connected at both ends of each edge. Suppose that k represents the degree of a node and that $m(k_1,k_2)$ is the total number of edges of the nodes whose degrees between k_1 and

k_2. If $k_1 = k_2$, $\mu(k_1,k_2) = 2$; otherwise, $\mu(k_1,k_2) = 1$. The joint degree distribution $p(k_1,k_2)$ is

$$p(k_1,k_2) = \frac{m(k_1,k_2)\mu(k_1,k_2)}{2m}. \quad (6)$$

While constructing the $2k$ null network of EBSN, links in the offline network remain unchanged while those in the online network should be shuffled. The steps of the $2k$ null network are the same as those of the $1k$ method, except that the degree value of four nodes which are connected by the two edges is required to remain the same before and after shuffling. For example, as shown in Fig. 8(a), only when nodes D and E have the same degree value can the edge exchanging be carried out to ensure that the joint degree distribution of the whole network remains unchanged. We disconnect two links A–E and C–D, and we then connect A–D and C–E. The result of the reconnection is shown in Fig. 8(b). The process of randomly rewiring should be repeated until a completely randomized null network is achieved.

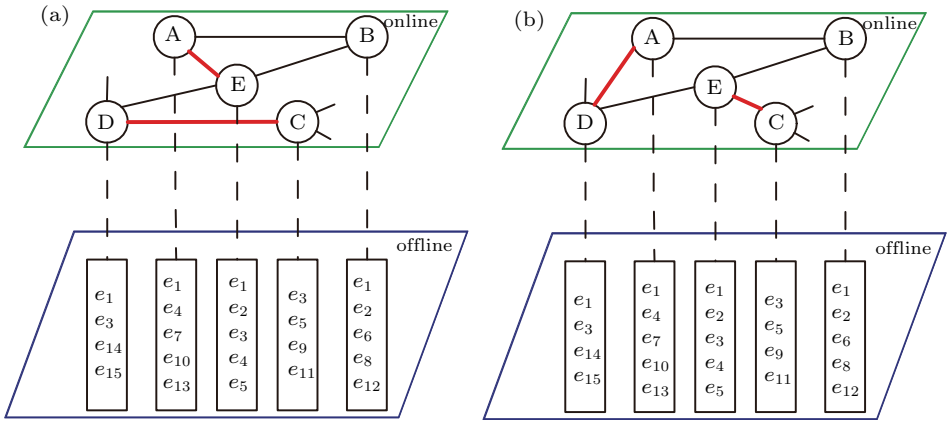


Fig. 8. The construction process of the $2k$ null network. We remove links A–E and C–D, and then add links A–D and C–E: (a) the original toy model, and (b) its corresponding $2k$ null network.

3.4. $3k$ null network of EBSN

The $3k$ null network possesses the same joint edge-degree distribution $p(k_1,k_2,k_3)$ as that of the original network, which means the same number of open triangle and close triangle exists between the original network and its $3k$ null network. The joint edge-degree distribution takes into account the interconnectivity between three nodes, mainly in two cases: the first is an open triangle where three nodes are connected by two edges, which is called $p_\Lambda(k_1,k_2,k_3)$; the second is a closed triangle where three nodes form a ring, which is called $p_\Delta(k_1,k_2,k_3)$.

While constructing the $3k$ null networks, links in the offline network remain unchanged while those in online network should be shuffled. Two edges are randomly select in the original network to ensure that the joint degree distribution remains the same; that is, the basic properties of the $2k$ characteristics remain the same. The number of open triangle and close triangle for each node on both ends of two links and its neighbor nodes are then calculated before and after shuffling, respectively. If the numbers are the same, then we can conclude that the shuffling is successful. For example, referring to the constraints of generating the $3k$ null network, we disconnect two edges A–B and C–D. If the pairs of nodes A–D and B–C are not connected, then we connect them. The result of the reconnection is shown in Fig. 9(b). The process of randomly rewiring should be repeated until a completely randomized null network is obtained.

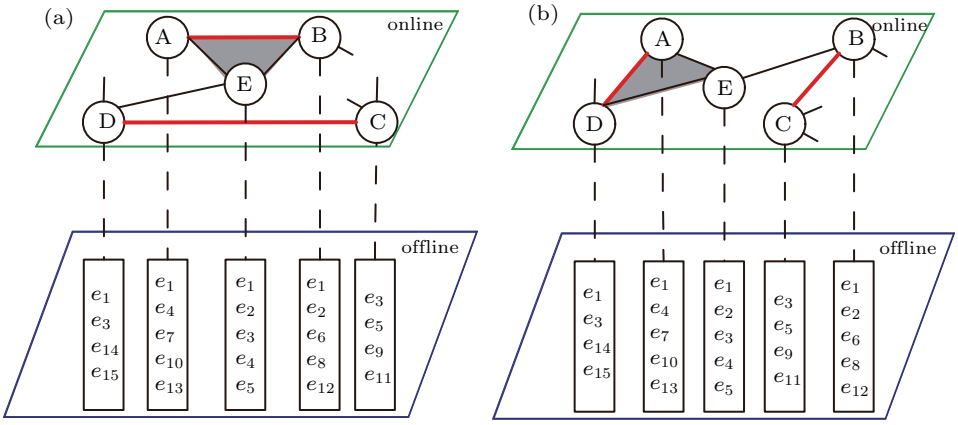


Fig. 9. The construction process of the $3k$ null network. We remove links A–B and C–D, and then add links A–D and B–C: (a) the original toy model, and (b) its corresponding $3k$ null network.

3.5. The impact of online social micro-scale structures on offline event similarity

The HPI distribution of the Plancast network and its 0k–3k null networks are shown in Fig. 10. The red quadrilateral line represents the HPI distribution of the original network, and the grey lines denote the results of 100 null networks. To make the experimental results more clear, the average HPI values of the 100 null networks is computed and we give the HPI distribution in the black-star line.

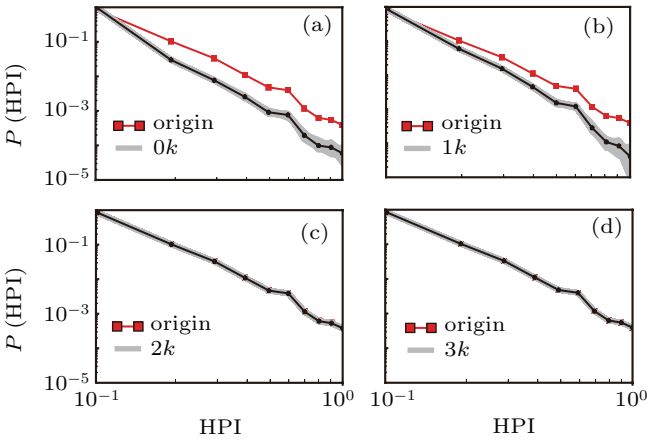


Fig. 10. The impact of online social micro-structures on offline event similarity. The HPI distribution of (a) 0k null networks, (b) 1k null networks, (c) 2k null networks, and (d) 3k null networks.

We find that the HPI distribution of 0k and 1k null network are less than that of the original network. Compared with 0k and 1k null networks, the HPI distribution of 2k null networks are very close to the original network because they maintain more micro-scale structure characteristics of the original (*i.e.*, assortativity). Furthermore, the HPI distribution of 3k null networks are more similar to that of the original network because 3k null networks keep the clustering characteristic of the original. These results suggest that the average degree (0k characteristic) and degree distribution (1k characteristic) are not enough to maintain event similarity between friends, but assortativity (2k characteristic) and clustering coefficient (3k characteristic) are significant micro-scale structures to maintain event similarity between friends.

4. The impact of structural diversity of online friends on offline event similarity

4.1. Indexes to measure online structural diversity

Studies on network structure characteristics can help us to reveal the nature of complex systems, and the relationship between network structure and its function. Therefore, more researchers are devoting themselves to study the diversity of network structures.[25,31] Actually, the number and subgraph structure of common neighbors are the important indexes to reflect the structural characteristics of social networks.[24] However, these two factors have different effects on real networks in existing studies. For example, the researchers in Ref. [31] have shown that the number of common neighbors has greater impact on the similarity of network structure. In contrast, it is found in Ref. [25] that the number of common neighbors has no influence on social networks while the subgraph structure of common neighbors has a positive influence. In this chapter, we study the influence of these two factors on offline event similarity between friends.

Common Neighbors (CN) index is a common indicator to explore the structural similarity of two users in social networks. The definition of CN is

$$S_{xy}^{\mathrm{CN}} = |\Gamma(x) \cap \Gamma(y)|, \tag{7}$$

where $\Gamma(x)$ denotes the set of neighbors of node x. The higher CN is, the more similar two users are.

The other index that can explore the structural diversity is the subgraph structure of common neighbors. We measure this index by the Number of small Connected Components (NCC) in the induced subgraph of common neighbors. For example, in Fig. 11(a) the nodes A, B, C, and D are common neighbors of the nodes U_1 and U_2 and they are wholly connected. Therefore, these four common neighbors are separated into one component. Figure 11(b) shows all of the possible scenarios that four common neighbors may cluster into one, two, three or four connected components with different formations.[25] A higher NCC indicates more groups and higher structural diversity.[32]

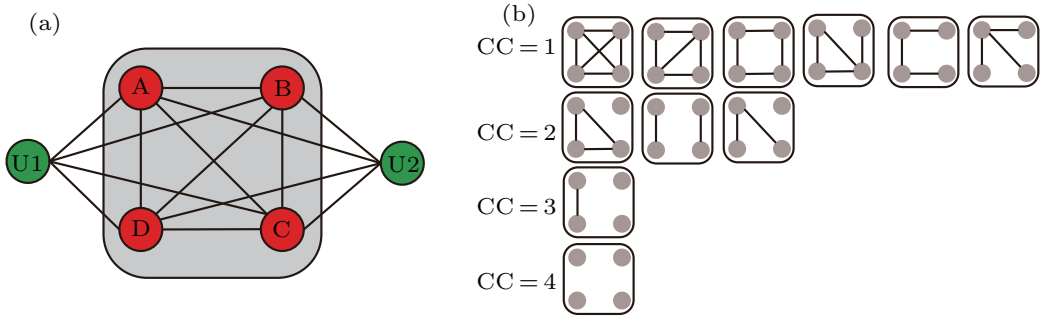


Fig. 11. An illustration of the subgraph structures of common neighbors: (a) the topology of two users and their common four neighbors and (b) an illustration of the subgraph structures of four common neighbors.

4.2. The impact of the number of common neighbors

We calculate the number of common neighbors in the original network and its corresponding $1k$–$3k$ null networks respectively to study event similarity between friends, and the corresponding HPI results are shown in Fig. 12. Here, simulations have been repeated 100 times independently for null models to achieve the average result. We use different types of lines to represent the original network and its corresponding $1k$–$3k$ null networks. The original network and its corresponding null networks show a similar tendency; that is, the HPI values increase gradually as the number of CN increases. In addition, the HPI values of users without common neighbors are less than those with common neighbors. The results indicate that there is a strong correlation between online social proximity and offline event similarity, while the number of CNs has a positive impact on the event similarity between friends.

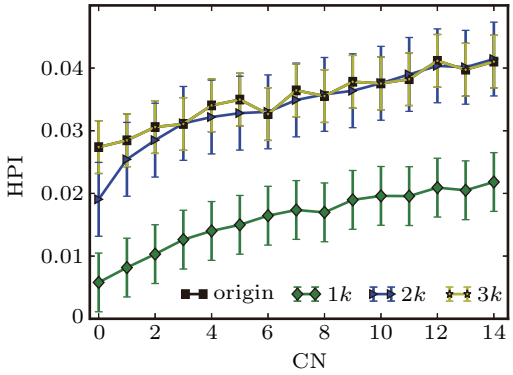


Fig. 12. The event similarity for different number of common neighbors. The x coordinate represents the CN values, and the y coordinate represents HPI values.

4.3. The impact of the subgraph structure of common neighbors

The results in Fig. 12 inspire us to study event similarity from the perspective of the subgraph structure of common neighbors.

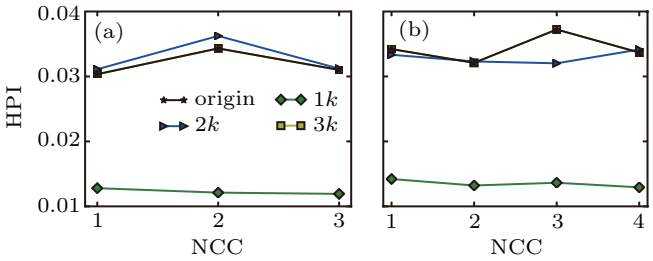


Fig. 13. Event similarity for the different NCC in the induced subgraph of common neighbors: (a) the HPI of NCC of three common neighbors, and (b) the HPI of NCC of four common neighbors. The x coordinate represents the NCC values, and the y coordinate represents HPI values.

In Fig. 13(a), in the case of CN = 3, we calculate the HPI values of different NCC of the original network and its corresponding $1k$–$3k$ null networks, and find that not only the original network HPI values but also null models do not show any regularity. It is indicated that the subgraph structure of common neighbors (motif diversity) has no positive impact on offline event similarity. The same method is used to get the results of the case of CN = 4, which is shown in Fig. 13(b). We can gain the same conclusion as the case of CN = 3.

Figures 12 and 13 show that the subgraph structure of the CN has no positive impact on offline event similarity while the number of CNs plays a key role. This is different from the results in Ref. [25]. From our point of view, the reason for this result may be due to social influence.[31]

4.4. Randomness of offline event similarity

The event data bring not only huge benefits to people but also the harm of personal information leakage.[33] This happens because event data not only directly contains user privacy information but also implies personality habits, health status, social status, and other sensitive information about the users. Once event data are improperly used, all aspects of users' privacy will be seriously threatened. Actually, the randomness of different null networks corresponds to the degree of information available for privacy protection. We calculate the HPI values of six null models, and the results are shown in Fig. 14. Experimental results show that the HPI values of online null networks are higher than those of offline null networks, indicating that the randomness of online null networks is less than that of offline null models. We also find that the HPI values of online null networks go down as the order decreases, and the randomness of online null networks goes down as the order increases. When the construction method combines online and offline networks, the HPI values are closer to those of the offline. This indicates that the offline characteristic has a greater impact on event similarity than online network structures.

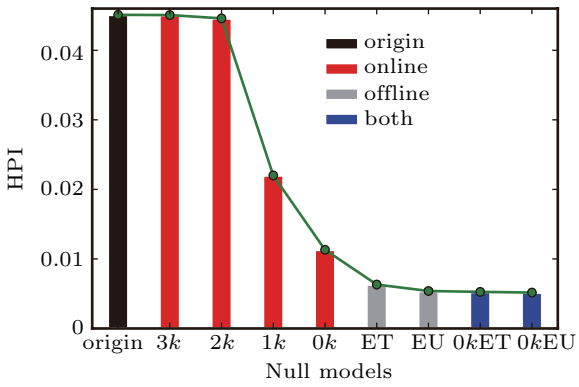


Fig. 14. The influence degree of online social factors and offline event factors. The x coordinate represents the random method, and the y coordinate represents the HPI values. ET denotes the null networks of exchanging trajectory, and EU represents the null networks of exchanging user events. $0k$ET represents the null networks combining $0k$ and exchanging trajectory, and $0k$EU represents the null networks combining $0k$ and exchanging user events.

5. Conclusion

In summary, we use Plancast data to build a double-layer network. By constructing null models of the EBSN network, we study the factors that affect event similarity between

friends and the influence level of each factor. First, two null models of exchanging trajectory and exchanging user events are proposed to change the user's events belongingness to detect offline event similarity between friends. Our experimental results show that there is a strong correlation between online social proximity and offline event similarity. To be specific, the events of friends show a similarity; that is, once we change the belongingness of events, the event similarity of friends will be greatly reduced. This correlation can help us to make offline event recommendations based on online social interactions.

Second, the influence of micro-scale network structures on the event similarity between friends is studied by constructing $0k$–$3k$ null networks of the original network. Our results indicate that the average degree and degree distribution ($0k$ and $1k$ characteristics) are not enough to maintain offline event similarity between friends, but assortativity ($2k$ characteristic) and high order features (such as clustering coefficient) are significant micro-scale characteristics for maintaining offline event similarity between friends. This is helpful for researchers who wish to understand the pattern of human mobility from online network topology.

Finally, we study the influence of structural diversity of online friends on offline event similarity. Our experimental results show that the structural diversity of common neighbors has no positive impact on event similarity, while the number of common friends plays a key role. In addition, the randomness of different null models that can measure the degree of information availability in privacy protection are discussed. Our study not only uncovers the factors that affect offline event similarity between friends but also presents a framework for understanding the pattern of human mobility in cities.

References

[1] Tong Y X, She J Y and Meng R 2016 *World Wide Web* **19** 1151
[2] Lian D, Xie X, Zhang F, Yuan N J, Zhou T and Rui Y 2015 *IEEE Data Engineering Bulletin* **38** 35
[3] Samanthula B K, Lei C, Wei J and Luo S 2015 *ACM Trans. Privacy and Security* **8** 141
[4] Zygiaris S 2013 *Journal of the Knowledge Economy* **4** 217
[5] Shi W, Lu W Z, Xue Y and He H D 2016 *Physica A* **443** 22
[6] Li Z, Tang J and Mei T 2018 *IEEE Trans. Pattern Anal. Mach. Intell.* **1** 1
[7] Li Z and Tang J 2016 *IEEE Trans. Image Process.* **1** 1
[8] Wu G L, Gu C G, Qiu L and Yang H J 2017 *Chin. Phys. B* **26** 128901
[9] Lai D R and Shu X 2017 *Chin. Phys. B* **26** 038902
[10] Liu Z T and Huang M L 2009 *Comput. Sci.* **36** 189
[11] Gjoka M, Kurant M and Markopoulou A 2012 *Proceedings - IEEE INFOCOM* **12** 1968
[12] Krioukov D, Krioukov D, Fall K and Vahdat A 2006 *ACM SIGCOMM Comp. Commun. Rev.* **36** 135
[13] Molloy M and Reed B 1998 *Combinatorics Probability* **7** 295
[14] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[15] Krioukov D, Papadopoulos F, Kitsak M and Vahdat M 2010 *Phys. Rev. E* **82** 036106
[16] Xu X K, Zhang J, Sun J and Small M 2009 *Phys. Rev. E* **80** 056106
[17] Shang K K, Small M, Xu X K and Yan W Sheng 2017 *Eurphys. Lett.* **117** 28002
[18] Liu B, Xu S, Li T, Xiao J and Xu X K 2018 *Entropy* **20** 363
[19] Sarzynska M, Leicht E A, Chowell G and Porter M A 2018 *Journal of Complex Networks* **4** 363
[20] Cui W K, Shang K K, Zhang Y J, Xiao J and Xu X K 2018 *Eur. Phys. J. B* **91** 145
[21] Maslov S and Sneppen K 2002 *Science* **296** 910
[22] Holme P 2005 *Phys. Rev. E* **71** 046119
[23] Xu X K, Wang X and Xiao J 2018 *Physica A* **505** 222
[24] Shang K K, Small M and Yan W S 2017 *Physica A* **469** 767
[25] Fan C, Liu Y, Huang J, Rong Z and Zhou T 2016 *Sci. Rep.* **7** 11975
[26] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabási A L 2002 *Science* **297** 1551
[27] Lv L Y and Zhou T 2010 *Physica A* **390** 1150
[28] Lee Y, Hickman M and Washington S 2007 *Transportation Research Part A* **41** 0–1020
[29] Carlsson G 2009 *Scand. J. Occup. Ther.* **11** 78
[30] Chen J R, Zhang L, Liu W W and Yan Z Z 2017 *Chin. Phys. B* **26** 018901
[31] Schieber T A, Carpi L, Diazguilera A, Pardalos P M, Masoller C and Ravetti M G 2017 *Nat. Commun.* **8** 13928
[32] Ugander J, Backstrom L, Marlow C and Kleinberg J 2012 *Proc. Nat. Acad. Sci. USA* **109** 5962
[33] Wang L and Meng X F 2014 *Journal of Software* **25** 693 (in Chinese)