# Social signature identification of dynamical social networks

Ren-De Li [a,b], Jian-Guo Liu [c,d,*], Qiang Guo [a], Yi-Cheng Zhang [d]

[a] *Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, People's Republic of China*
[b] *Library, University of Shanghai for Science and Technology, Shanghai 200093, People's Republic of China*
[c] *Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China*
[d] *Department of Physics, University of Fribourg, CH-1700 Fribourg, Switzerland*

### HIGHLIGHTS

- We empirically investigate the evolution characteristics of Facebook and Wiki users' social signature.
- There are robust social signatures for collective egos based on interactions.
- Individual social signature remains stable, no matter how alters change over time.
- Alters identified from structural information show the characteristic of social signature.
- The accuracy of the prediction in terms of embeddedness ranges from 33.47% to 66.90%.

*Keywords:*
Social signature
Ego networks
Relation identification

Understanding the social relation of dynamical online social networks (OSNs) is significant for identifying the strong and weak ties. In this paper, we empirically investigate the evolution characteristics of Facebook and Wiki users' *social signature*, capturing the distribution of frequency of interactions between different alters over time in ego network. The statistical results show that there are robust social signatures for collective egos based on interactions. Notably, individual social signature remains stable, no matter how alters change over time. Furthermore, we use structure information of OSNs to predict the social closeness in each interval in terms of embeddedness measurement, and find that alters identified from structural information still show the characteristic of social signature. The accuracy of the prediction in terms of embeddedness ranges from 33.47% to 66.90%. This work enables an effective method for identifying potential highly-related friends based on the regularity of social signature in online social network.

## 1. Introduction

Online social networks (OSNs) [1] have attracted hundreds of millions users to communicate and interact with each other through the past years, which is of great significance for identifying the human relation patterns [2–5]. Although

* Corresponding author at: Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China.
*E-mail address:* liujg004@ustc.edu.cn (J.-G. Liu).
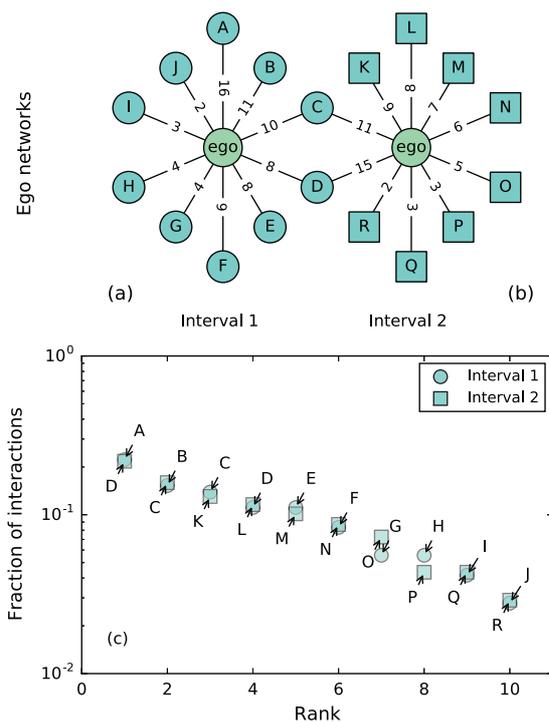
**Fig. 1.** The evolution properties of a specific user. We extract top-10 alters of the target ego according to tie strength. Subplot (a) and (b) show the tie strength in ego network, which is the number of interactions between the target (ego) and friends (alters) in two consecutive time intervals 1 and 2, where there are two alters (alters C and D) coexist in two intervals. The turnover is defined by the *Jaccard* measurement, which is defined as the ratio of common alters over total alters in two intervals, in this case, $2/18 = 0.1111$. Subplot (c) shows two social signatures of the ego in each interval, where the social signature is defined as the fraction of interactions to the alter of each rank, which approximately has a form of exponential distribution. Persistence of two social signatures is defined as the similarity between two probability distributions.

it is convenient for users to build connection with more friends by OSNs, due to the cognitive ability [6,7] and memory capacity [8,9] constraints, the communication pattern of OSNs is found similar to the offline face-to-face networks [10]. Dunbar argued that, limited to human cognitive ability, average speaking, each individual could only maintain 150 friends, who has only 5 closest friends [11], which has been verified by face-to-face networks [12,13], mobile communication relationships [14,15], and online social networks [16–19].

Since the scale of meaningful relations is limited in OSNs, recognizing the closest friends and meaningful interactions are crucial tasks for spreading social influence [20–22], discovering online user's behavioral preference [23–25], and providing better online service in recommendation systems [26,27]. Bond et al. [21] investigated the Facebook network with 61 million users for political mobilization and found that close friends exerted more influence on voters mobilized than the message itself. Hence, identifying the closest relationship has attracted much attention [28]. Statistical properties of the static social ties, such as tie strength [29–31,21] and embeddedness [32,29], have been extensively studied. The mechanisms like preferential return and memory effect have given some explanations of activities in OSNs [33,34]. However, online social networks are evolving dynamically, and evolving characteristics of the users' relations are still missing.

In this paper, we identify *social signature* of OSNs users' relations, which is defined as a distribution calculated by the fraction of interactions between ego and top-$R$ alters in ego network, capturing a distinctive and robust interactive pattern [14]. By investigating the constant pattern for 30 students over 18 months period during the transition from school to university or work, Saramäki et al. found that social signatures are similar with each other for different alters [14]. Different from face-to-face network, the online social network provides an effective way to build social connections. We investigate the evolving characteristics of social signature (Fig. 1) in two OSNs, namely Facebook-Wall (short for FW) [35,36] and Wiki-Talk (short for WT) [37,38]. The empirical results show that the persistence of social signatures exists on both collective and individual level over time. Even for the users who only exist in one time interval, their social signatures are also significant. Combining each ego network, one could have a global social network, which provides a way to explore the correlation between the network structure and the evolution properties of social relations. Inspired by this idea, we present one way to predict the social closeness in terms of the structure property.

**Table 1**

Basic statistical properties of the Facebook-Wall (short for FW) and Wiki-Talk (short for WT) dataset, including the time span of each interval $i$ and the number of nodes $N$, edges $E$, long-active egos $n_L$, short-lived egos $n_S$.

| Interval | FBWall | | | | |
|---|---|---|---|---|---|
| | Time span | $N$ | $E$ | $n_L$ | $n_S$ |
| $i = 1$ | 2006-09-05~2007-01-24 | 8,544 | 58,549 | 56 | 385 |
| $i = 2$ | 2007-01-24~2007-06-14 | 11,768 | 106,160 | 56 | 814 |
| $i = 3$ | 2007-06-14~2007-11-02 | 15,653 | 136,806 | 56 | 1174 |
| $i = 4$ | 2007-11-02~2008-03-22 | 17,663 | 119,296 | 56 | 907 |
| $i = 5$ | 2008-03-22~2008-08-10 | 22,991 | 140,348 | 56 | 1201 |
| $i = 6$ | 2008-08-10~2008-12-29 | 36,344 | 230,847 | 56 | 2313 |
| Interval | Wikitalk | | | | |
| | Time span | $N$ | $E$ | $n_L$ | $n_S$ |
| $i = 1$ | 2005-09-09~2006-01-28 | 81,860 | 470,296 | 618 | 2056 |
| $i = 2$ | 2006-01-28~2006-06-18 | 192,916 | 957,599 | 618 | 3736 |
| $i = 3$ | 2006-06-18~2006-11-06 | 207,167 | 1,094,482 | 618 | 4502 |
| $i = 4$ | 2006-11-06~2007-03-27 | 293,397 | 1,449,596 | 618 | 5449 |
| $i = 5$ | 2007-03-27~2007-08-15 | 338,571 | 1,623,190 | 618 | 5523 |
| $i = 6$ | 2007-08-15~2008-01-03 | 320,861 | 1,540,296 | 618 | 4902 |

## 2. Empirical analysis

### 2.1. Materials

The FW dataset spans from September 14, 2004 to January 22, 2009. It should be noticed that a user can post comments on his(her) friends' walls, and these comments can be seen by visitors. The WT dataset spans from August 30, 2001 and January 6, 2008. The interactions in the WT data represent that a user edited another user's talk page. We extract the WT data from September 9, 2005 to January 3, 2008, which is in the same time span of the FW data. Each interval includes 141 days. Here, we treat interactions as undirect links since we only need to know the connection of ego and alters when calculating the social signature. The information of each interaction consists three parts: The wall/talk page owner (ego), the user who posted/edited (alter) and the corresponding posted/edited time. Datasets are separated into $I$ intervals evenly, here $I = 6$, and each interval includes 141 days. Then, in each interval, ego networks are constructed as the list of egos who has at least $R$ alters and take top-$R$ closest alters by counting interactions, here $R = 10$. Specifically, we define the long-term egos $L$ as those who appear in all 6 intervals, and short-term egos $S$ are those who only occur in one interval. For ego $j$, the number of long-term ego $n_L$ is 56 and 618 in the FW and WT datasets separately, and the number of short-term ego $n_S$ varies in each interval.

For the Facebook and Wiki-talk datasets, we divided two datasets into 6 equal time intervals (see Table 1 for basic statistical properties). In each interval, we construct ego networks, where egos and alters are tied to each other by online social relations. Tie strength is the number of interactions of each ego $j$ to alters in each interval $i$ and the alters are ranked on descending order by the number of interactions. Then, we calculate the social signature $P_{i,j}$ of ego $j$ in the interval $i$, defined by the fraction of interactions to the alter of each rank (Eqs. (1) and (2)). Here, only those who have at least 10 alters are taken into account to calculate social signatures. Meanwhile, two kinds of egos are separately compared. The long-term egos $L$ are those appeared in all 6 intervals. The short-term egos $S$ are those who have more than 10 alters in one interval, while in the other intervals, their alters are less than 10. To quantify the similarity of social signatures, we use the Jensen–Shannon Divergence (short for $JSD$) to measure the distance of the social signatures between two egos (see Measurements). $JSD = 0$ stands for the fact that two social signatures are identical. The smaller the $JSD$ value is, the more similar two social signatures are, which means fraction of interactions has persistence over time. We use $Jaccard$ (short for $J$) to quantify the turnovers of two different sets of alters (see Measurements). $J = 0$ represents that there are no common alters, while $J = 1$ represents the exact same alters staying in consecutive intervals. As illustrated in Fig. 1, the number of common alters equals $20 \cdot J/(1 + J)$, in this case, $J = 0.1111$, showing that there are two common alters between two consecutive intervals. Empirically, we compare the characteristics of social signatures' similarity $JSD$ and turnover $J$ on both collective and individual levels over 6 time intervals.

### 2.2. Measurements

**Social signature** $P$. The social signature of ego $j$ in the interval $i$, saying $P_{i,j}$ (specifically, denoting $P_{i,j}^L$ for long-term ego, $P_{i,j}^S$ for short-term ego), is defined as fraction of interactions of each alters after ranking and taking top $R$ alters (here $R = 10$), which can be read as,

$$P_{i,j} = \{p_{i,j}(r), r = 1, 2, \ldots, R\}, \tag{1}$$

3

$$p_{i,j}(r) = \frac{f_{i,j}(r)}{\sum_{r=1}^{R} f_{i,j}(r)}, \tag{2}$$

where $f_{i,j}(r)$ is the frequency of interactions given by an ego $j$ to the alter $r$ in the interval $i$, and $p_{i,j}(r)$ is the fraction of interactions to the alter of rank $r$. The rank $r$ ranges from 1 to $R$. Meanwhile, to measure collective social signature in each interval $i$, we define the average social signature $\langle P_i \rangle$ as:

$$\langle P_i \rangle = \{\langle p_{i.}(r) \rangle, r = 1, 2, \ldots, R\}, \tag{3}$$

$$\langle p_{i.}(r) \rangle = \frac{1}{n} \sum_{j=1}^{n} p_{i,j}(r), \tag{4}$$

where $\langle p_{i.}(r) \rangle$ is the average fraction of interactions to the alter of rank $r$ in the interval $i$, $i = 1, 2, \ldots, 6$. For long-term ego, $n = n_L$, and average social signature in each interval $i$ is denoted as $\langle P_i^L \rangle$. For short-term ego, $n = n_S$, and average social signature in each interval $i$ is denoted as $\langle P_i^S \rangle$.

**Jensen–Shannon Divergence** *JSD*. Jensen–Shannon Divergence (short for *JSD*) [14] is introduced to measure the similarity between two different social signatures $P_1$ and $P_2$, defined as:

$$JSD(P_1, P_2) = H[\frac{1}{2}P_1 + \frac{1}{2}P_2] - \frac{1}{2}[H(P_1) + H(P_2)], \tag{5}$$

$$H(P) = -\sum_{r=1}^{R} p(r) \log p(r), \tag{6}$$

where $P_1$ and $P_2$ are two social signatures to be compared, which are the form of the interaction distributions. If comparing a long-term ego's social signatures between two consecutive time intervals, $P_1 = P_{i,j}^L$ and $P_2 = P_{i+1,j}^L$. If comparing collective egos signatures between two consecutive time intervals, $P_1 = \langle P_i^L \rangle$ and $P_2 = \langle P_{i+1}^L \rangle$. If comparing two average social signatures of between short-term and long-term egos in the same interval, $P_1 = \langle P_i^S \rangle$ and $P_2 = \langle P_i^L \rangle$. Additionally, $H(P)$ represents the Shannon entropy, where $p(r)$ refers to Eq. (2) if calculating *JSD* value of individual ego $j$, and refers to Eq. (4) calculating *JSD* value of collective egos. From the definition of *JSD*, the lower bound is *JSD* = 0 only when two distributions are identical. The smaller *JSD* value is, the more persistence of two social signatures are.

Specifically, when calculating the individual *JSD* value averaged though intervals, we denote the mean value of *JSD* as:

$$\langle JSD_j \rangle = \frac{1}{I-1} \sum_{i=1}^{I-1} JSD(P_{i,j}, P_{i+1,j}). \tag{7}$$

where $P_{i,j}$ and $P_{i+1,j}$ standing for social signatures of a long-term ego $j$ in the interval $i$ and $i + 1$, and $I = 6$.

**Jaccard** *J*. We define $X_{i,j}$ as the set of alters, which contains 10 alters here, for ego $j$ in the interval $i$. The difference between two sets $X_{i,j}$ and $X_{i+1,j}$ of alters in two consecutive intervals $i$ and $i+1$, namely *turnover*, is measured by the Jaccard coefficient *J*,

$$J(X_{i,j}, X_{i+1,j}) = \frac{|X_{i,j} \cap X_{i+1,j}|}{|X_{i,j} \cup X_{i+1,j}|}. \tag{8}$$

The value of turnover *J* lies in [0, 1]. *J* = 1 means that all alters of target ego are the same over time, and *J* = 0 stands for no common alters. The low *Jaccard* value means that two sets of alters have little common alters, in another word, turnover of alters is high.
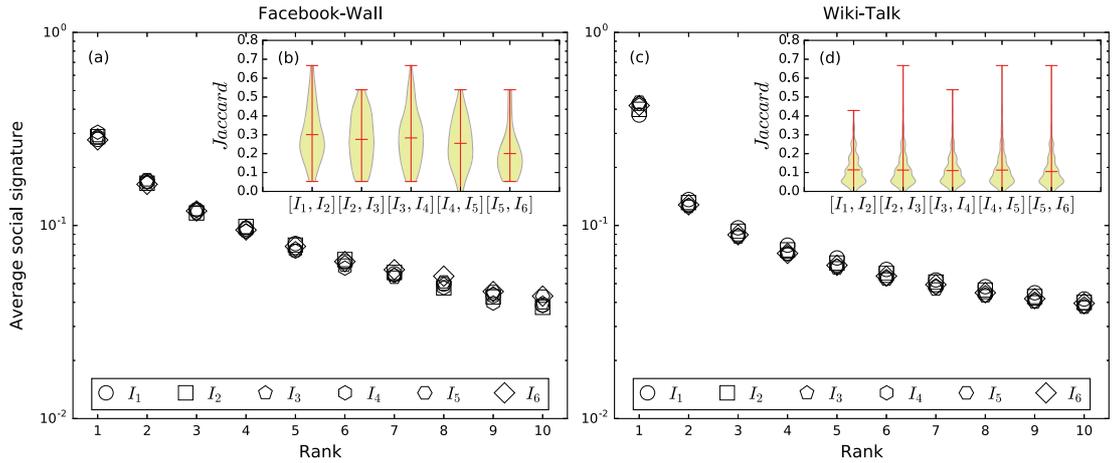
Then, we calculate the *Jaccard* averaged over collective egos in two consecutive intervals $i$ and $i + 1$:

$$\langle J(X_{i,\cdot}, X_{i+1,\cdot}) \rangle = \frac{1}{n_L} \sum_{j=1}^{n_L} J_j(X_{i,j}, X_{i+1,j}). \tag{9}$$

where the number of long-term ego $n_L = 56$ for FB dataset and $n_L = 618$ for WT dataset.

Specifically, when calculating the individual *J* value averaged though intervals, we denote the mean value of *J* for a ego $j$ as:

$$\langle J_j \rangle = \frac{1}{I-1} \sum_{i=1}^{I-1} J_j(X_{i,j}, X_{i+1,j}). \tag{10}$$

4

**Fig. 2.** The average social signature $\langle P_i^L \rangle$ and turnover $J(X_{i,j}, X_{i+1,j})$ of collective long-term egos for the FW and WT datasets. The shapes of average social signatures are invariant (seen in subplot (a) and (c) for the distributions in interval $I_1, I_2, \ldots, I_6$), although alters display different turnovers as indicated by the *Jaccard* between two consecutive intervals ($[I_i, I_{i+1}]$ ($i = 1, 2, \ldots, 5$)) for two datasets (seen in subplot (b) and (d) as violin plots). For the FW users, the average *Jaccard* in consecutive intervals are 0.30,0.28,0.28,0.25, and 0.20 respectively. For the WT users, average *Jaccard* equals 0.11 throughout time intervals.

It should be noticed that, since the *Jaccard* is the value of alters compared over time, the value is only available for long-term egos.

**Slope** $\lambda$. The average social signature $\langle P_i \rangle$ is approximately the form of exponential distribution, as is shown in Fig. 2, we fit the distribution by the following probability density function:

$$f(x) = \lambda \cdot e^{-\lambda x}, \tag{11}$$

where the parameter $\lambda$ is the *slope* of social signature and variable $x$ is the rank $1, 2, \ldots, R$. The larger the parameter $\lambda$ is, the steeper the distribution is.

**Null model.** In order to examine social signatures deviating from randomness, a null model is introduced to compare with the empirical results, which is constructed as follows: (i) Randomly reshuffle alters, reconnecting the reshuffled alter to ego and remain timestamp, (ii) Construct the ego network based on reshuffled data, and then calculate the average social signature $\langle P_i \rangle$ according to Eq. (3).

### 2.3. The empirical results

**The persistence of collective social signature**

For collective long-term egos in each interval, first, we calculate the average social signature $\langle P_i^L \rangle$ by averaging over the set of all long-term egos $L$'s fraction of interactions in each interval $i$ (Eqs. (3) and (4)). As shown in Fig. 2(a) and (c), the shape of average social signature $\langle P_i^L \rangle$ of long-term egos in each interval exhibits the form of exponential distribution in both datasets. Then, we calculate the similarity of two average social signatures between two consecutive time intervals $JSD(\langle P_i^L \rangle, \langle P_{i+1}^L \rangle)(i = 1, 2, \ldots, 5)$ by Eq. (5). The similarity values of $JSD$ are all smaller than 0.0005 for both datasets, which suggests that the interactive frequency of ego to alters remain stable between two consecutive time intervals. Second, to measure the turnover of alters through time, we calculate the *Jaccard* of each long-term ego between two consecutive intervals $J(X_{i,j}, X_{i+1,j})$ (Eq. (8)), where $X_{i,j}$ and $X_{i+1,j}$ are the two sets of the alters interacted with the ego $j$ in time interval $i$ and $i + 1$ separately. The distributions of collective *Jaccard* in each interval are shown as violin plot in Fig. 2(b) and (d), with the mean value calculated by Eq. (9). For the FW users, the average *Jaccard* is at the range of [0.2, 0.3], which means that there are average 3–5 common alters over time. For the WT users, the average *Jaccard* is stable around 0.11, which means there are average 2 common alters over time. From the information above, there exists the persistence of average social signatures over time, despite turnovers of alters.

**The persistence of individual social signature**

Besides the persistence of collective interaction behaviors, we investigate the individual behaviors by averaging each individual ego's $JSD$ though intervals, denoted as $\langle JSD_j \rangle$ (calculated by Eq. (7)), and the turnover of each individual over time, denoted as $\langle J_j \rangle$ (calculated by Eq. (10)). Then, we rank the egos based on the average values of the turnover of individuals $\langle J_j \rangle$. Fig. 3 shows that the turnovers $\langle J_j \rangle$ are different among individuals with the average *Jaccard* value ranging from 0.1 to 0.5 for both datasets. One can find that, for the long-term egos in FW dataset, the maximum average *Jaccard* value is 0.47, which

5
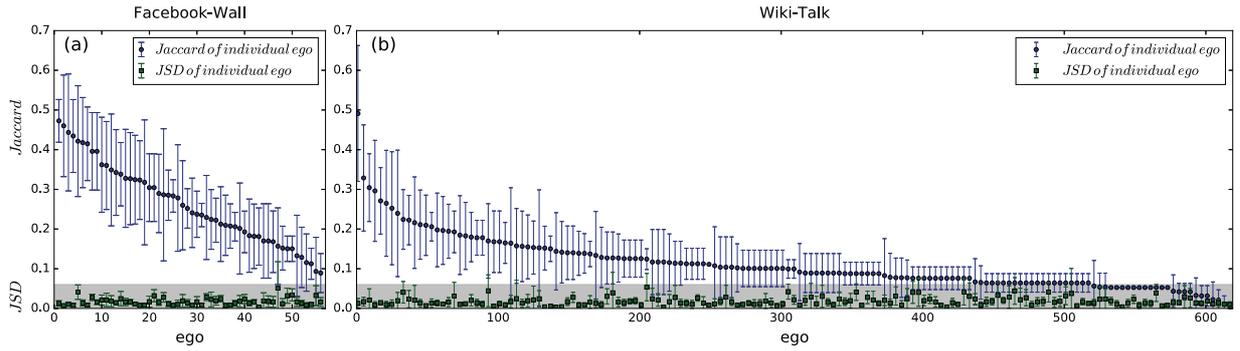
**Fig. 3.** The individual social signature and turnover of long-term egos for the FW and WT datasets. The *x*-axis represents individual ego, where egos are ranked according to the mean value of *Jaccard* of ego (averaged over time, displayed by error bar with circle markers) in the descending order. The corresponding values of *JSD* of ego (averaged over time, displayed by error bar with square markers) beneath the *Jaccard* error bar, are smaller than 0.07. The comparisons indicate that despite the wide range of difference of turnover, individual's social signature remains persistence over time.
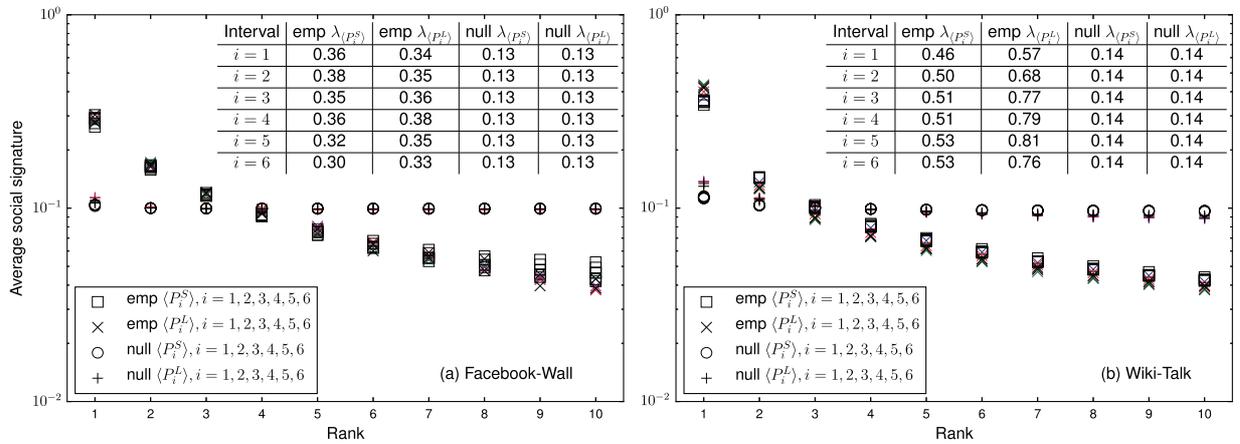


**Fig. 4.** Slope $\lambda$ is the parameter of exponential distribution used to depict the shape of average social signature of short-term egos $\langle P_i^S \rangle$ denoted by $\lambda_{\langle P_i^S \rangle}$, and long-term egos $\langle P_i^L \rangle$ denoted by $\lambda_{\langle P_i^L \rangle}$ in each interval $i$, compared between empirical results (short for emp) and the null model (short for null). Each marker stands for 6 average social signatures (6 distributions) of each group of results. The comparisons show that average social signatures of empirical results are steeper than that of the null model, which indicate that social signature has heterogeneous characteristic.

indicates the ego's alters are very stable, in another word, there are 6 alters frequently communicated over time; While the minimum average *Jaccard* value is 0.09, indicating alters of the ego changes a lot, only 1.7 alters remaining in constant relations. The WT users also show the wide difference of individual's turnover over time. Despite the difference of turnovers, the *JSD* values of egos over time are smaller than 0.07, which indicates the stable social signature for individual ego over time.

### The similarity of long-term and short-term ego's social signatures

In addition, for those short-term egos who merely leave their interactive footprints in one time interval, we compare their social signatures with the long-term egos' social signatures in their time intervals. For the FW users, the similarity of average social signatures between short-term and long-term egos, denoted as $JSD(\langle P_i^S \rangle, \langle P_i^L \rangle)$ (calculated by Eq. (5)), is on average $0.0004 \pm 0.0002$ in each interval. For the WT users, the corresponding value is on average $0.0018 \pm 0.0008$. The results show that it still exists high similarity between social signatures of short-term and long-term egos, indicating that even those who have no evolving characteristic, there still exist the social signatures. As the matter of fact, in the following structural analysis, for these short-term egos with social signature, it is much easier to predict their alters.

### The heterogeneity of social signature

The social signature exists in both long-term and short-term egos in both OSNs datasets. We introduce a null model to regenerate the data for comparison. The null model reshuffles all alters into random order and reallocate alters to each ego, so that the egos interactive patterns no longer exist. As shown in Fig. 1, individual and collective signatures approximately follow exponential distributions. Then, we use the parameter $\lambda$ of exponential distribution to depict the shape of signature (Eq. (11)). Fig. 4 shows the value of slope $\lambda$ for empirical results and null model results. We calculate the slopes $\lambda$ of four

groups of average social signatures, which are average social signatures of collective short-term egos from empirical results (denoted as emp $\langle P_i^S \rangle$), average social signatures of collective long-term egos from empirical results (denoted as emp $\langle P_i^L \rangle$), average social signatures of collective short-term egos from null model (denoted as null $\langle P_i^S \rangle$), average social signatures of collective long-term egos from null model (denoted as null $\langle P_i^L \rangle$). Meanwhile, in each group, there are six distributions, which represent six average social signatures in six time intervals (calculated according to Eq. (3)). One can find that the values of $\lambda$ in the FW dataset are twice as much as that of null model, while the values of $\lambda$ in the WT dataset are triple larger than that of the null model, which indicates that empirical social signatures have steeper slope. The results indicates that ego's tie strength with alters decreases with the frequency of interactions, much faster than the null model, namely, the social signature has heterogeneous characteristic.

## 3. The structural analysis

### 3.1. A structural method

Empirically, social signatures have been investigated based on frequency of interactions. However, it is hard for us to access the interactive relations comparing with the structural information. For instance, in the OSNs, communications of each pair of users are private. Then, the question is, whether one could predict the social closeness in term of the network structure. Here, we present a simple procedure to predict the most closest friends by the structure information: (1) Draw the unweighted network from online social network in each time interval; (2) Calculate the embeddedness $E$ between each pair of nodes, which is defined as the ratio of the common neighbors over the total first-order neighbors of the pair nodes (Eq. (12)); (3) Take and rank top-$R$ first-order neighbors of target ego based on the value of embeddedness $E$. Then, one can infer a set of ego networks based on the structure information with the same egos as empirical ego networks based on frequency of interactions, where close relation between ego and alters is denoted as the value of embeddedness $E$. For distinguishing from the empirical results, we call ego network based on tie strength as *empirical ego network*, and ego network based on embeddedness as *structural ego network*. The comparisons between two kinds of ego networks are in the same time interval. Correspondingly, social signatures are divided into two classes: *the empirical signature $P_{i,j}$* and *the structural signature $Q_{i,j}$* of ego $j$ in the interval $i$.

We still use Fig. 1 as schematic illustration, but the purpose here is to evaluate the predictive characteristics of structural ego network, rather than evolving characteristics of empirical ego network. So the elements in Fig. 1 have been changed. For an individual ego $j$, assuming subplot (a) of Fig. 1 is the empirical ego network based on frequency of interactions in the interval 1. Subplot (b) of Fig. 1 is the structural ego network, where the close relation between ego and alters is the value of embeddedness $E$, while the interval 1 and 2 are the same and denoted by the interval $i$. Then, one can get two kinds of social signatures in subplot (c) of Fig. 1. For individual $j$ in the interval $i$, the empirical signature $P_{i,j}$ is calculated by fraction of interactions, while the structural signature $Q_{i,j}$ is calculated by fraction of embeddedness $E$. The similarity of two kinds of social signatures is still measured by the *JSD*. The smaller *JSD* value is close to 0, the more consistent the structural social signature is to the empirical signature. Besides, we use *precision* as the measurement of prediction, which is defined as the number of identified alters divided by the number of top-$R$ alters, in this example, $2/10 \times 100\% = 20\%$. As a baseline, the embeddedness measurement could achieve a predictable performance of 24.7% in OSNs [29].

**Embeddedness** $E$. Embeddedness is served as the measurement in structural analyses for identifying close relation, capturing how much the two friends' social circles 'overlap' [39,40]. The embeddedness measurement $E$ can be expressed as,

$$E_{uv}^\alpha = \left[ \frac{n_{uv}}{(k_u - 1)(k_v - 1) - n_{uv}} \right]^\alpha, \tag{12}$$

where $u$ is an ego, $v$ is a neighbor of $u$, $k$ is the degree of the nodes and $n_{uv}$ is the number of common neighbors. Specifically, the parameter $\alpha$ is tunable for different type of egos' signatures. To get the best fitness between empirical signature and the structural signature, when calculating embeddedness in unweighted network of each interval, we adjust the value of the parameter $\alpha = 1.3$ for long-term ego and $\alpha = 0.75$ for short-term ego in the FW dataset, and the parameter $\alpha = 0.66$ for long-term ego and $\alpha = 0.64$ for short-term ego in the WT dataset. Notably, the parameter $\alpha$ is independent of the results of prediction.

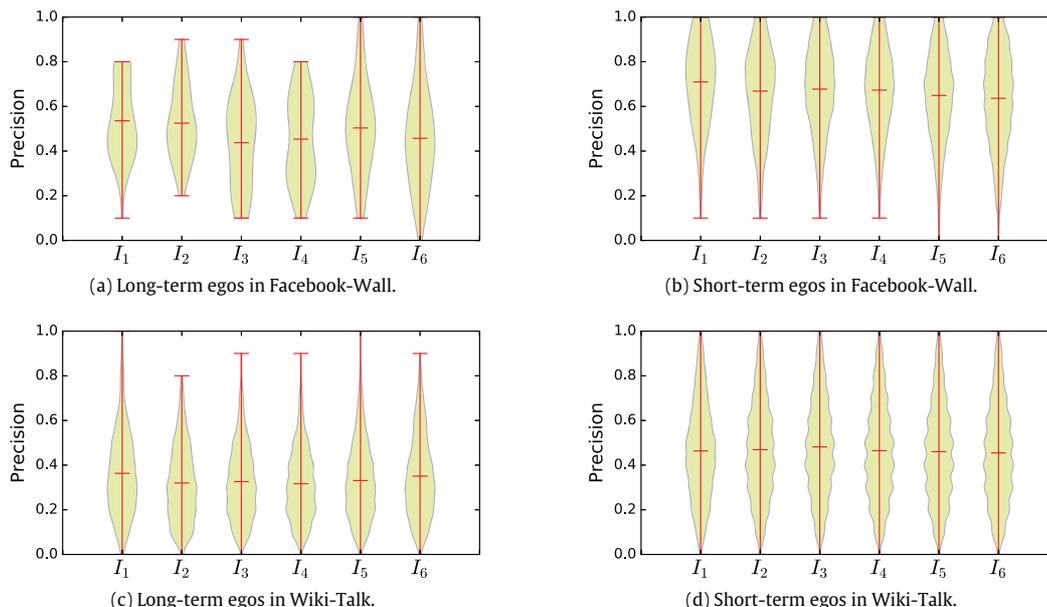**Precision.** To measure the predictive effect of alters identified by embeddedness $E$, we define

$$Precision = \frac{n_E}{R}, \tag{13}$$

where $n_E$ is the number of alters identified by embeddedness, and $R = 10$. Here, we take top-$R$ closest alters as the results of prediction, which is 10 alters here. So, in another words, $n_E$ is the number of intersection between the set of predicted alters and the set of actual alters with social signature.

**Table 2**

The *JSD* values of the average empirical signature and the structural signature for long-term and short-term egos in each time intervals. The comparisons show the similarity of two kinds of social signatures.

| Interval | FW | | WT | |
|---|---|---|---|---|
| | $JSD(\langle P_i^L \rangle, \langle Q_i^L \rangle)$ | $JSD(\langle P_i^S \rangle, \langle Q_i^S \rangle)$ | $JSD(\langle P_i^L \rangle, \langle Q_i^L \rangle)$ | $JSD(\langle P_i^S \rangle, \langle Q_i^S \rangle)$ |
| $i = 1$ | 0.0035 | 0.0037 | 0.0055 | 0.0033 |
| $i = 2$ | 0.0021 | 0.0041 | 0.0055 | 0.0037 |
| $i = 3$ | 0.0017 | 0.0030 | 0.0052 | 0.0035 |
| $i = 4$ | 0.0021 | 0.0034 | 0.0051 | 0.0038 |
| $i = 5$ | 0.0015 | 0.0025 | 0.0052 | 0.0039 |
| $i = 6$ | 0.0006 | 0.0025 | 0.0045 | 0.0037 |



(a) Long-term egos in Facebook-Wall.

(b) Short-term egos in Facebook-Wall.

(c) Long-term egos in Wiki-Talk.

(d) Short-term egos in Wiki-Talk.

**Fig. 5.** Precision of alters identified by embeddedness, shown as violin plot for long-term and short egos in each time interval.

## 3.2. The results analysis

Firstly, we calculate average structural signatures $\langle Q_i^L \rangle$ and $\langle Q_i^S \rangle$ by averaging over the collective long-term and short-term ego's social signatures according to Eq. (3), comparing with the average empirical social signatures $\langle P_i^L \rangle$ and $\langle P_i^S \rangle$ respectively by the measurement of *JSD* in each interval $i$. The values of $JSD(\langle P_i^L \rangle, \langle Q_i^L \rangle)$ and $JSD(\langle P_i^S \rangle, \langle Q_i^S \rangle)$ are all less than 0.005 in each interval for the FW dataset. While in the WT dataset, the *JSD* of long-term egos are around 0.005 and less than 0.005 for short-term egos (see Table 2 for detail values in each interval). The results indicate that, alters identified from structural information still show the characteristic of social signature, which is consistent with empirical ego network. Namely, the social signature presents the intrinsic characteristic embedded in network structure.

Secondly, we calculate the precision (Eq. (13)) of long-term and short-term egos predicted by embeddedness in all 6 intervals (Fig. 5). In the FW dataset, the mean values of precision in each time interval are 53.57%, 52.50%, 43.75%, 45.35%, 50.36%, 45.71% respectively for long-term egos, and 70.96%, 66.87%, 67.73%, 67.30%, 64.91%, 63.62% respectively for short-term egos. In the WT dataset, the mean value of precision in each time interval are 36.33%, 32.02%, 32.62%, 31.68%, 33.07%, 35.08% respectively for long-term egos, and 46.35%, 46.92%, 48.17%, 46.48%, 46.03%, 45.48% respectively for short-term egos. The results show the predictive ability of structural measurement of embeddedness, where precisions of the FW users are larger than the WT users, and short-term egos' are larger than long-term egos'. The reason of lower precision for long-term egos may lie in that long-term ego's alters are constantly changing over different time intervals and their network structure are also changing, which cause the difficulty of prediction. While the short-term egos only appear in one time interval, which provides stable friendships and network relations. If the social signature exists, it will be easier to predict short-term egos' alters than long-term egos'.

## 4. Conclusion and discussions

In summary, we empirically investigate the evolution characteristics of online users' social signatures in dynamical online social networks. The FW and the WT datasets are divided into 6 equal time intervals. In each interval, we construct the ego networks in which each ego has at least 10 alters. Two kinds of egos are separately investigated. Long-term egos are defined as those whose ego networks exist in all 6 intervals, and short-term egos are defined as those whose ego networks only appear in one interval. The evolution properties are based on the tie strength of interactions, measured by social signature $P$, persistence $JSD$, turnover $J$ and slope $\lambda$. Empirical results show that: (1) Persistence of social signatures exists in long-term egos, where the $JSD$ values are all smaller than 0.0005 for both datasets; (2) Individual ego's signature is stable over time despite the wide range of turnover of alters, where turnover $J$ ranges widely from 0 to 0.5 while persistence $JSD$ values are smaller than 0.07; (3) Short-term egos also have the characteristic of social signatures, where $JSD$ values are smaller than 0.0005 by comparing with long-term egos' social signatures in the FW dataset, and smaller than 0.005 in the WT dataset. Besides, we use a null model for comparison. The results show the heterogeneous characteristic of the social signature $P$, where the slope $\lambda$ of empirical social signatures are 2 times larger than that of null model for the FW users and 3 times for the WT users. Furthermore, we investigate the correlation between the network structure and the social close relation. By using the embeddedness $E$, one could calculate the social closeness in terms of the network structure. Comparing with empirical tie strength based on interactive frequency, one could find that alters identified from structural information still show the characteristic of social signature. Moreover, embeddedness measurement could predict top-10 alters with the average precision of 48.54%, 66.90%, 33.47%, 46.57% respectively for long-term egos in the FW data, short-term egos in the FW data, long-term egos in the WT dataset, short-term egos in the WT dataset.

Promising further directions include exploring the intrinsic mechanism of human online communication by constructing models to accurately reproduce the empirical observations, such as random-walk-based model [41,42] and structural-based network model [43,44]. Besides, the universal pattern of social relation is still need to be empirically investigated. For instance, could the relation of scientists' collaboration be captured by social signature? Considering the heterogeneity of human behaviors, could preferential selection mechanism in game theory [45] be introduced to explain the social signature evolution? Moreover, the effect of Dunbar number should be examined when calculating social signatures. Since human layers of social closeness approximately contain 5, 15, 50 and 150 individuals [11], the size of close friends may affect predictive results, which has not been investigated in our paper. In addition, individual social attributes [46,47] may have significant influence to the way of people's communication patterns. Finding the answers of these questions could help in deep understanding the pattern of human social relation and improve the ability to identify important people.

## References

[1] S.P. Borgatti, A. Mehra, D.J. Brass, G. Labianca, Network analysis in the social sciences, Science 323 (2009) 892.
[2] R.I. Dunbar, Coevolution of neocortical size, group size and language in humans, Behav. Brain Sci. 16 (1993) 681.
[3] R.I. Dunbar, J. Bever, Neocortex size predicts group size in carnivores and some insectivores, Ethology 104 (1998) 695.
[4] R.I. Dunbar, S. Shultz, Understanding primate brain evolution, Philos. Trans. R. Soc. B 362 (2007) 649.
[5] H. Kudo, R.I. Dunbar, Neocortex size and social network size in primates, Anim. Behav. 62 (2001) 711.
[6] J. Stiller, R.I. Dunbar, Perspective-taking and memory capacity predict social network size, Soc. Net. 29 (2007) 93.
[7] J. Powell, P.A. Lewis, N. Roberts, M. García-Fiãana, R.I. Dunbar, Orbital prefrontal cortex volume predicts social network size: an imaging study of individual differences in humans, Philos. Trans. R. Soc. B 279 (2012) 2157.
[8] H.R. Bernard, P.D. Killworth, On the social structure of an ocean-going research vessel and other important things, Soc. Sci. Res. 2 (1973) 145.
[9] G.A. Miller, The magical number seven plus or minus two: some limits on our capacity for processing information, Psychol. Rev. 63 (1956) 81.
[10] R.I. Dunbar, Do online social media cut through the constraints that limit the size of offline social networks? Roy. Soc. Open Sci. 3 (2016) 150292.
[11] R.I. Dunbar, V. Arnaboldi, M. Conti, A. Passarella, The structure of online social networks mirrors those in the offline world, Soc. Net. 43 (2015) 39.
[12] R.A. Hill, R.I. Dunbar, Social network size in humans, Hum. Natu. 14 (2003) 53.
[13] R.I. Dunbar, S. Shultz, Evolution in the social brain, Science 317 (2007) 1344.
[14] J. Saramäki, E.A. Leicht, E. López, S.G. Roberts, F. Reed-Tsochas, R.I. Dunbar, Persistence of social signatures in human communication, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 942.
[15] T. Aledavood, E. Lõpez, S.G. Roberts, F. Reed-Tsochas, E. Moro, R.I. Dunbar, J. Saramäki, Daily rhythms in mobile telephone communication, PLoS ONE 10 (2015) e0138098.
[16] Y.Y. Ahn, S. Han, H. Kwak, S. Moon, H. Jeong, Analysis of topological characteristics of huge online social networking services, in: ACM In Proceedings of the 16th international conference on World Wide Web, 2007, pp. 835–844.
[17] S.A. Golder, D.M. Wilkinson, B.A. Huberman, Rhythms of social interaction: Messaging within a massive online network, Commun. Technol. 41 (2007).
[18] B. Gonçalves, N. Perra, A. Vespignani, Modeling users' activity on twitter networks: Validation of dunbar's number, PLoS ONE 6 (2011) e22656.
[19] Q. Guo, F. Shao, Z.L. Hu, J.G. Liu, Statistical properties of the personal social network in the Facebook, EPL 104 (2013) 28004.
[20] D. Centola, The spread of behavior in an online social network experiment, Science 329 (2010) 1194.

[21] R.M. Bond, et al., A 61-million-person experiment in social influence and political mobilization, Nature 489 (2012) 295.
[22] S. Aral, D. Walker, Identifying influential and susceptible members of social networks, Science 337 (2012) 337.
[23] L. Hou, X. Pan, Q. Guo, J.G. Liu, Memory effect of online user preference, Sci. Rep. 4 (2014) 6560.
[24] L. Ji, J.G. Liu, L. Hou, Q. Guo, Identifying the role of common interests in online user trust formation, PLoS ONE 10 (2015) e0121105.
[25] Q. Guo, L. Ji, J.G. Liu, J.T. Han, Evolution properties of online user preference diversity, Physica A 468 (2017) 698.
[26] T. Zhou, et al., Solving the apparent diversity-accuracy dilemma of recommender systems, Proc. Natl Acad. Sci. USA 107 (2010) 4511.
[27] Q. Guo, W.J. Song, L. Hou, Y.L. Zhang, J.G. Liu, Effect of the time window on the heat-conduction information filtering model, Physica A 401 (2014) 15.
[28] S. Aral, D. Walker, Tie strength, embeddedness, and social influence: A large-scale networked experiment, Manag. Sci. 60 (2014) 1352.
[29] L. Backstrom, J. Kleinberg, Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook, ACM, 2014, p. 831.
[30] M. Granovetter, Getting a Job: A Study of Contacts and Careers, Harvard University Press, 1974.
[31] P.V. Marsden, K.E. Campbell, Measuring tie strength, Soc. F. 63 (1984) 482.
[32] D. Easley, J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, 2010.
[33] Z.D. Zhao, Y.C. Gao, S.M. Cai, T. Zhou, Dynamic patterns of academic forum activities, Physica A 461 (2016) 117.
[34] Z.D. Zhao, S.M. Cai, Y. Lu, Non-Markovian character in human mobility: Online and offline, Chaos 25 (2015) 063106.
[35] J.G. Liu, Z.M. Ren, Q. Guo, D.B. Chen, Evolution characteristics of the network core in the Facebook, PLoS ONE 9 (2014) e104028.
[36] B. Viswanath, A. Mislove, M. Cha, K.P. Gummadi, On the evolution of user interaction in facebook, in: Proceedings of the 2nd ACM workshop on Online social networks, 2009, pp. 37–42.
[37] A. Paranjape, A.R. Benson, J. Leskovec, Motifs in temporal networks, in: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017.
[38] J. Leskovec, D.P. Huttenlocher, J.M. Kleinberg, Governance in social media: A case study of the Wikipedia promotion process, in: ICWSM, 2010.
[39] D.H. Felmlee, No couple is an island: A social network perspective on dyadic stability, Soc. Forc. 79 (2001) 1259.
[40] M. Kalmijn, Shared friendship networks and the life course: an analysis of survey data on married and cohabiting couples, Soc. Net. 25 (2003) 231.
[41] M. Starnini, A. Baronchelli, R. Pastor-Satorras, Model reproduces individual, group and collective dynamics of human contact networks, Soc. Net. 47 (2016) 130.
[42] T. Jia, D. Wang, B.K. Szymanski, Quantifying patterns of research-interest evolution, Nat. Hum. Behav. 1 (2017) 0078.
[43] W.X. Wang, B. Hu, T. Zhou, B.H. Wang, Y.B. Xie, Mutual selection model for weighted networks, Phys. Rev. E 72 (2005) 046140.
[44] Z. Pan, X. Li, X. Wang, Generalized local-world models for weighted networks, Phys. Rev. E 73 (2006) 056109.
[45] W.B. Du, X.B. Cao, L. Zhao, M.B. Hu, Evolutionary games on scale-free networks with a preferential selection mechanism, Physica A 388 (2009) 4509.
[46] S. Centellegher, E. López, J. Saramäki, B. Lepri, Personality traits and ego-network dynamics, PLoS ONE 12 (2017) e0173110.
[47] H.H. Jo, R.I. Dunbar, J. Saramäki, K. Kaski, Dynamics of close relationships for the life-course migration, Sci. Rep. 4 (2014) 6988.