# Connections between Human Dynamics and Network Science

# Supplemental Material

Chaoming Song, Dashun Wang, and Albert-László Barabási

## Contents

# S1 Data Description

We compiled four independent datasets that together capture most aspects of digital communication that humans are involved in lately:

- *Mobile phone*: The dataset represents one year of call patterns from 4 million anonymized mobile phone users in 2008, consisting of $\approx$ 100 million calls each month [1].

- *Email*: The dataset is extracted from the log files of a university mail server during a period of 83 days, consisting around 3,000 users interchanging 0.3 millon messages [2].

- *Twitter*: The dataset consists of 54 million users who produced a total of 1.7 billion tweets between Sep-2006 and Aug-2009 (http://twitter.mpi-sws.org/) [3], among which we select 0.7 million users at random. This subset contained over 8 million direct user-to-user communications spanning over 8 months time frame from Aug 2009 to Mar 2010.

- *Online Messages*: The data covers more than 500,000 messages sent during a 492 days period among almost 30,000 members of the Swedish online community pussokram.com. The user base is primarily composed of teenagers, of which 70% are women [4, 5].

Table S1 summarizes the ensemble sizes used to determine $P_k$, $P_w$, $P_\tau$ and $P_C$ distributions for each dataset.

Table S1: **Ensemble sizes for each distribution.**

| | Mobile phone | Email | Twitter | Message |
|---|---|---|---|---|
| $P_k(k)$ | $4,174,344$ | $2,429$ | $732,220$ | $28,876$ |
| $P_w(w)$ | $53,982,936$ | $39,256$ | $1,923,918$ | $174,151$ |
| $P_\tau(\tau)$ | $567,021,128$ | $44,180$ | $2,935,997$ | $14,561$ |
| $P_C(C)$ | $4,174,344$ | $2,429$ | $732,220$ | $28,876$ |
| $\overline{C}$ | 307.86 | 53.16 | 227.01 | 20.83 |

## S2 Corrections for Multiple Email Recipients

An important feature of email, distinguishing it from other forms of communication, is the ability to send a message to multiple recipients at the same time. To quantify the impact of this feature on the scaling, we define $n_i$ as the average number of recipients per email among all the emails sent by individual $i$. As we show in Fig. S1a, the distribution $p(n)$ across the user base follows a fat tailed distribution with exponent $\sim 2.0$, representing a significant population heterogeneity in the number of email recipients. The larger $n_i$ for an individual, the more friends he/she contacts each time. To account for this effect, Eq. (1) should be revised to

$$k_i(t_1,t_2) \sim n_i(t_1,t_2) C_i(t_1,t_2)^{\alpha_i} \tag{S1}$$

where $n_i = 1$ for phone calls/twitter/online messages in which case (S1) reduces to (1). In Fig. S1b, we find that the growth of $\langle k_i/n_i \rangle$ indeed collapses for different $C$ values, implying that the sociability $\alpha$ is statistically independent of $(C,n)$, and thus

$$P_k(k) = \int \delta(k - nC^{\alpha}) P_{C,n}(C,n) P_{\alpha}(\alpha) d\alpha dn dC, \tag{S2}$$

where $P_{C,n} \equiv N^{-1} \sum_i \delta(C - C_i)\delta(n - n_i)$ is the joint activity/multi-recipients distribution. The validity of Eq. (S2) is confirmed for Email dataset through numerical integration (Fig. 3A).

If the system falls in the *Case 1*, we can further simplify Eq. (S2) by replaing $\alpha_i$ with the population average $\overline{\alpha}$,

$$P_k(k) = \int \delta(k - nC^{\overline{\alpha}}) P_{C,n}(C,n) dn dC. \tag{S3}$$

To test the validity of Eq. (S3), we measured $x_i = n_i C_i^{\overline{\alpha}}$ for each email user and computed the $P_x(x)$ distribution, finding that it indistinguishable from the empirically observed $P_k$ (Figs. 3A, S1C).

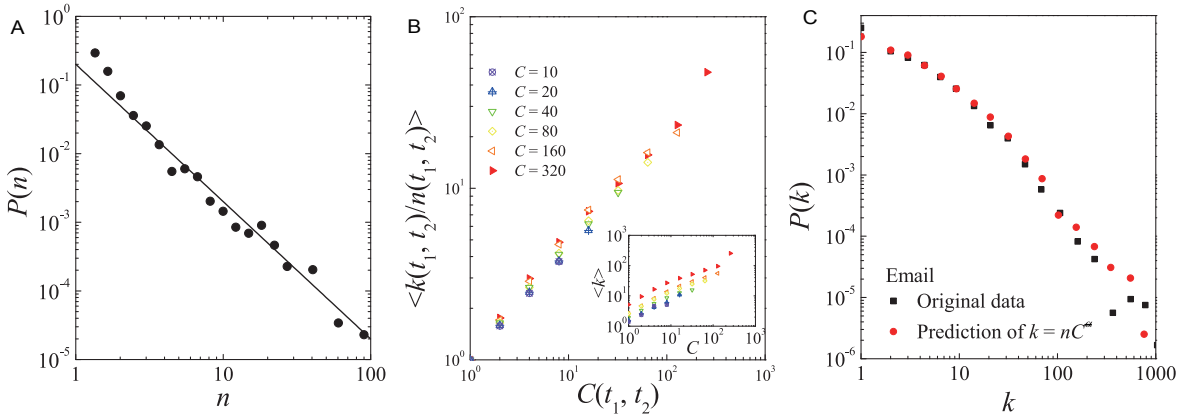


Figure S1: (A) The multi-recipient number distribution $P(n)$ across the population for the email datasets, indicating the inhomogeneous nature of email recipients. The straight line represents a power law with exponent $-2$. (B) The average $k(t_1,t_2)/n(t_1,t_2)$ vs. $k$ for different $C$-value groups, validating the scaling relationship $k \sim nC^{\alpha}$. (C) The distribution of $x_i \equiv n_i C_i^{\overline{\alpha}}$, showing that $P(x) \sim x^{-\beta_x}$ collapses with the $P(k)$.

If $P_x(x) \sim x^{-\beta_x}$, we have

$$\gamma_k = \min\left[\beta_x, 1 + \frac{\sigma}{\overline{\beta}_\tau \ln \overline{C}}\right]. \tag{S4}$$

Again, for $n_i = 1$ we have $x_i = C_i^{\overline{\alpha}}$ and $\beta_x = 1 + (\beta_C - 1)/\overline{\alpha}$, and Eqs. (S2-S4) reduce to Eqs. (5),(6) and (8), respectively.

# S3 Systematic Analysis Across All Datasets

Some results are omitted in the main text and in some cases only the results for the Mobile phone dataset were shown for brevity. Here we present the analysis for all datasets systematically, showing that the results are consistent across the whole data corpus.

## S3.1 Interevent Time Distribution $P_{\Delta t}$

To characterize the individual communication activity, we measured the interevent time distribution $P(\Delta t)$ for the studied datasets (Fig. S2).

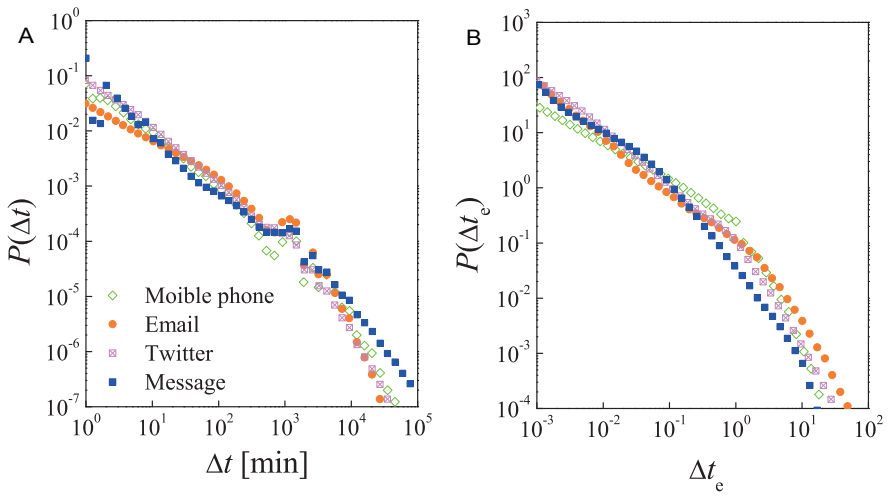


Figure S2: The distribution $P(\Delta t)$ and $P(\Delta t_e)$ for the interevent time between two calls using the (A) real time and (B) event time unit (normalized by the total number of users).

## S3.2 $\langle k \rangle$ vs. $C$

As shown in Eq. (S1), the degree of an individual has a power law scaling relationship with its activity level, $k_i \sim n_i C_i^{\alpha_i}$, where $\alpha_i$ characterizes the individual's affinity to translate its level of activity into new contacts, and $n_i \geq 1$ for email dataset and $n_i = 1$ for other datasets. Figure S3 shows that in average $\langle k(t_1,t_2) \rangle \approx C(t_1,t_2)^{\overline{\alpha}}$.

## S3.3 $p_r$ vs. $r$

We denote with $p_r \equiv p_{i\to j} \equiv w_{i\to j}/C_i$ the probability that user $i$ communicates with user $j$, and $r$ is the rank of $p_{i\to j}$ across all friends $j$ of user $i$. We find that $p_r$ is well approximated by Zipf's law $p_r \sim r^{-\zeta_i}$ (Fig. 3a) [6] for the Mobile phone data. We find that for users with similar activity level ($C$), their $\zeta$ exponents are different, yet $p_r$ collapses for users with similar sociability ($\alpha$), leading to the scaling relationship $\alpha_i \zeta_i = 1$. Figure S4 shows the results for the other three datasets, indicating that our analysis is robust across the whole data corpus.

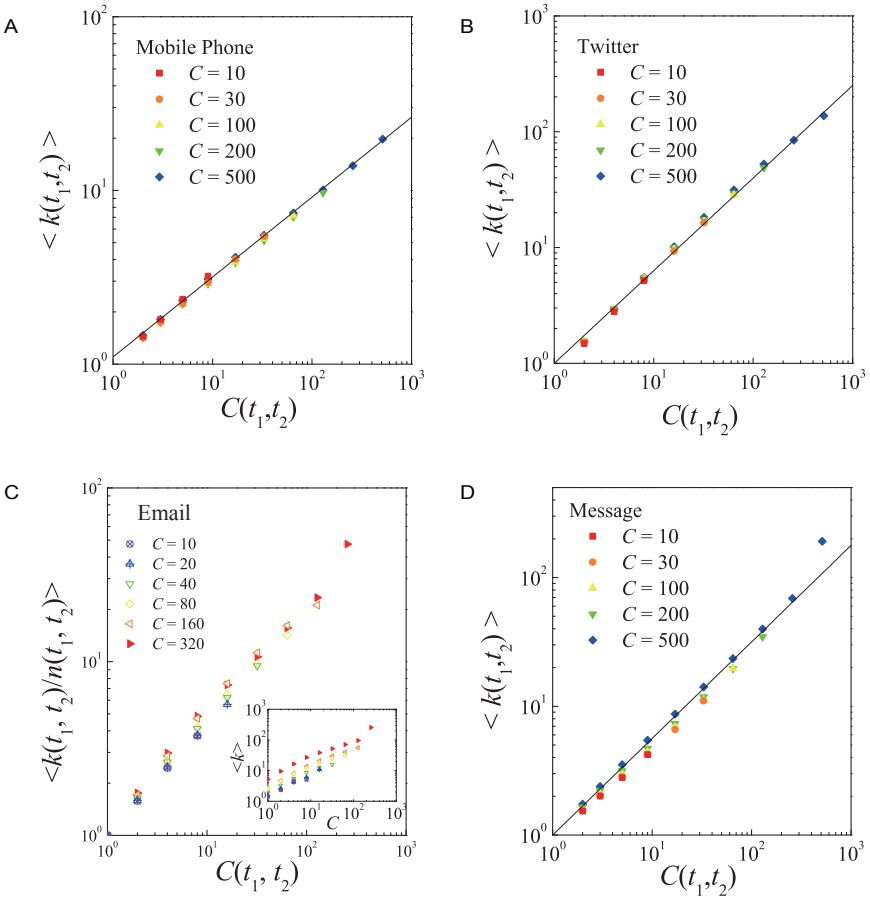


Figure S3: $\langle k(t_1,t_2)\rangle \approx C(t_1,t_2)^{\overline{\alpha}}$ for all datasets, showing the scaling relationship between $k$ and $C$ is robust across all datasets.

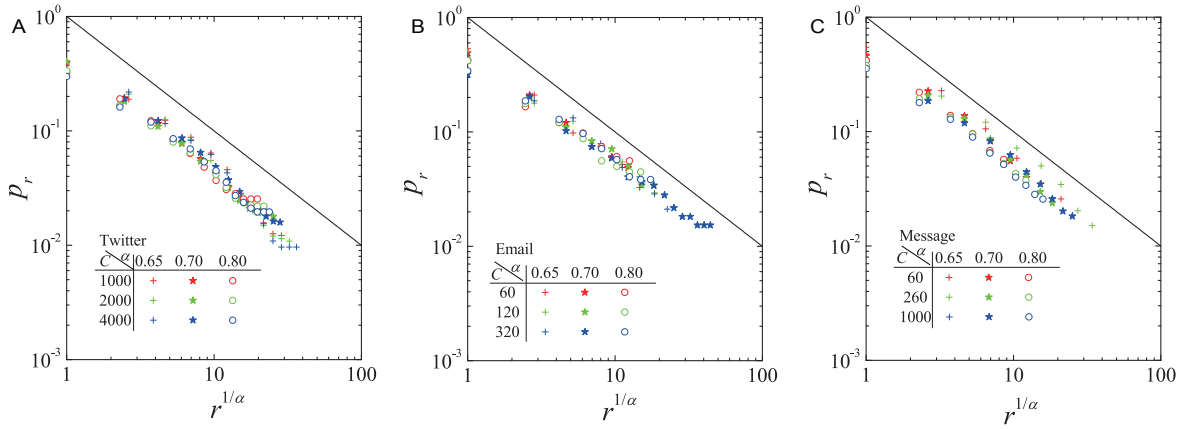


Figure S4: Zipf's plot of the communication frequency $p_r$ versus $r^{1/\alpha}$ of (A) Twitter, (B) Email and (C) Online Messages, showing the data collapse over the different sociability groups, as predicted by $\alpha_i\zeta_i = 1$, derived in (11). The different colors and symbols represent different activities and sociabilities, respectively.

# S4 Prediction of Degree Distribution

In this section, we offer a detailed derivation of the $P_k(k)$ for different potential functional forms of $P_C(C)$.

## S4.1 Case 1 [Email, Twitter and Online Messages]

In the *Case 1* limit $P_k(k)$ is dominated by differences in the users' activity level, i.e., the activity distribution $P_C(C)$. We ignore the inherent population variability in $\alpha$, hence we represent the individuals' sociability ($\alpha_i$) by its mean value $\overline{\alpha}$. In this case, Eq. (5) leads to

$$P_k(k) \sim k^{1/\overline{\alpha}-1} P_C(k^{1/\overline{\alpha}}), \tag{S5}$$

#### S4.1.1 Power law $P_C$

Substituting $P_C(C) \sim C^{-\beta_C}$ into (S5), we find,

$$P_k(k) \sim k^{-\beta_C/\overline{\alpha}} k^{1/\overline{\alpha}-1}, \tag{S6}$$

yielding,

$$\gamma_k = 1 + \frac{\beta_C}{\overline{\alpha}} = 1 + \frac{\beta_C}{1-\overline{\beta_\tau}}. \tag{S7}$$

#### S4.1.2 Log-normal $P_C$

Substituting $P_C(C) \sim C^{-1} e^{-(\ln C - \mu_C)^2/(2\sigma_C^2)}$ into (S5), we find

$$P_k(k) \sim k^{-1} e^{-(\ln k - \mu_k)^2/(2\sigma_k^2)}, \tag{S8}$$

where

$$\sigma_k = \overline{\alpha}\sigma_C \tag{S9}$$

#### S4.1.3 Stretched exponential $P_C$

Substituting $P_C(C) \sim C^{\beta_C - 1} e^{-(\lambda_C x)^{\beta_C}}$ into (S5), we find

$$P_k(k) \sim k^{-(1-\beta_k)} e^{-(\lambda_k k)^{\beta_k}}, \tag{S10}$$

where

$$\beta_k = \beta_C/\overline{\alpha}. \tag{S11}$$

Note that the stretched exponentials are in fact power laws with cutoffs, and therefore the scaling (S11) is actually equivalent with (S7).

### S4.2 Case 2 [Mobile Phone Calls]

In the *Case 2* limit, $P_k(k)$ is dominated by the sociability distribution $P_\alpha(\alpha)$. In this case each individual's activity level ($C_i$) can be approximated with a mean $\overline{C}$, and the tail of $P_\alpha(\alpha)$ plays the leading role. Hence the particular functional form is not relevant. Therefore, Eq. (5) leads to

$$P_k(k) \approx \int \delta(k - \overline{C}^\alpha) P_\alpha(\alpha) d\alpha, \tag{S12}$$

leading to

$$P_k(k)dk = P_\alpha(\alpha)d\alpha = P_\alpha(\ln k/\ln \overline{C})d(\ln k/\ln \overline{C}). \tag{S13}$$

As $P_\alpha(\alpha) \sim \exp(-\alpha\sigma/\overline{\beta}_\tau)$ we find,

$$P_k(k) \sim \exp\left(-\frac{\sigma}{\overline{\beta}_\tau}\frac{\ln k}{\ln \overline{C}}\right)\frac{1}{k}, \tag{S14}$$

or

$$P_k(k) \sim k^{-\gamma_k}, \tag{S15}$$

where

$$\gamma_k = 1 + \frac{\sigma}{\overline{\beta}_\tau \ln \overline{C}}. \tag{S16}$$

## S5 Estimating the Fitting Parameters

Given the empirically observed fat tailed nature of the distributions, it is important to see if the data is statistically consistent with the best fits reported in Table 1 in the main text. In this section, we describe our procedures to obtain the reported exponents.

## S5.1 Maximum Likelihood Estimation

The Maximum Likelihood Estimation (MLE) is a powerful method to determine the fitting parameters that best describe an empirical distribution. It is especially useful when the distribution is characterized by a fat tail, as least-squares fitting for this class of distributions are known to often lead to systematic errors [7]. Assume that we have a sample $x_1, x_2, \ldots, x_n$ of $n$ data points, and a testing model $P(x|\alpha)$, where $\alpha$ are the parameters of the model. The maximum likelihood estimation selects the parameters $\alpha$ such that the following joint probability (known as likelihood function),

$$L(\alpha) \equiv \prod_{i=1}^{n} P(x_i|\alpha), \tag{S17}$$

is maximized. In practice it is often more convenient to work with the logarithm of the likelihood function,

$$\ln L(\alpha) \equiv \sum_{i=1}^{n} \ln P(x_i|\alpha), \tag{S18}$$

or its scaled version, the average log-likelihood

$$\hat{l}(\alpha) \equiv \frac{1}{n} \ln L(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \ln P(x_i|\alpha). \tag{S19}$$

The method of maximum likelihood estimates $P(x|\alpha)$ by finding a value of $\alpha$ that maximizes $\hat{l}(\alpha)$.

The error bars of the fitting parameter $\alpha$ can be obtained using the bootstrapping method, offering an accurate assessment of the variance of the parameter estimates [8]. To do this, we generated a series of bootstrap resamples drawn randomly from the empirically observed histogram, and performed a fitting, using the algorithms discussed above for each random sample. By measuring the variance of the fitting parameters of these resamples, we are able to calculate the error bars precisely.

## S5.2 Power Law Ansatz

Mathematically, a quantity $x$ obeys a power law if it is drawn from a probability distribution

$$P(x) \propto x^{-\alpha}, \tag{S20}$$

where $\alpha$ is the scaling parameter of the distribution, known as the exponent. If $x_i \in [x_{min}, \infty)$ are *continuous* random variables, the optimization of the average log-likelihood (S19) leads to [7]

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right]. \tag{S21}$$

If, however, $x_i$ are *discrete* random variables, which is usually the case for empirically measured quantities, no analytical form is applicable. Therefore in this case we applied the steepest descent algorithm to optimize Eq. (S19) directly.

All the exponents reported in Table 1 (main text) are obtained by applying the fitting procedures described in [7]. In Table S2, we report the power law fitting parameters of $P_k(k)$, $P_C(C)$ and $P_w(w)$ for all datasets through MLE. Table S3 summarizes the fitting parameters for the Mobile phone datasets for different time periods, where the cutoff $k_{min}$ is obtained by minimizing Kolmogorov-Smirnov (KS) goodness of fit test [7].

Table S2: Power law fitting parameters

| | Mobile phone | Email | Twitter | Message |
|---|---|---|---|---|
| $k_{min}$ | $21 \div 40$ | 26 | 1 | 1 |
| $\gamma_k$ | $4.19_{\pm 0.01} \div 3.205_{\pm 0.007}$ | $2.27_{\pm 0.01}$ | $1.241_{\pm 0.001}$ | |
| $KC_k$ | 0.003 | 0.023 | 0.19 | 0.11 |
| $C_{min}$ | $190 \div 1,926$ | 57 | 1 | 1 |
| $\beta_C$ | $3.36_{\pm 0.01} \div 3.39_{\pm 0.01}$ | $0.82_{\pm 0.01}$ | $0.147_{\pm 0.001}$ | $0.430_{\pm 0.002}$ |
| $KC_C$ | 0.02 | 0.046 | 0.15 | 0.11 |
| $w_{min}$ | 1 | 1 | 1 | 1 |
| $\gamma_w$ | $1.51335_{\pm 0.00006}$ | $1.637_{\pm 0.003}$ | $1.8483_{\pm 0.0006}$ | $1.930_{\pm 0.002}$ |
| $KC_w$ | 0.036 | 0.068 | 0.046 | 0.071 |

Table S3: Fitting parameters for the Mobile phone dataset for different time frames

| | 1-month | 3-months | 6-months | 1-year |
|---|---|---|---|---|
| $k_{min}$ | 21 | 28 | 29 | 40 |
| $\gamma_k$ | $4.19_{\pm 0.01}$ | $3.67_{\pm 0.01}$ | $3.391_{\pm 0.005}$ | $3.205_{\pm 0.007}$ |
| $KC_k$ | 0.0030 | 0.0046 | 0.0053 | 0.0032 |
| $C_{min}$ | 190 | 543 | 1,138 | 1,926 |
| $\beta_C$ | $3.37_{\pm 0.02}$ | $3.46_{\pm 0.02}$ | $3.53_{\pm 0.02}$ | $3.39_{\pm 0.01}$ |
| $KC_C$ | 0.020 | 0.018 | 0.018 | 0.022 |
| $\ln \overline{C}$ | 3.39 | 4.31 | 5.03 | 5.73 |

Table S4: **Quantify networks and human dynamics.** The scaling exponents characterizing the networks and human dynamics in the four studied datasets, as well as the most studied human dynamics models. The reported $\overline{\alpha}$ and $\overline{\beta_\tau}$ represent average values over the population for empirical data, where $\overline{\beta_\tau}$ is measured from $P_\tau(\tau) \sim \tau^{-(1+\overline{\beta_\tau})}$ as a first order approximation. The error of $\overline{\beta_\tau}$ and $\beta_C$ are derived from the error of $1+\overline{\beta_\tau}$ and $1+\beta_C$, respectively. Note that the small error bars of exponents are due to the large population size.

| | Mobile phone | Email | Twitter | Message | Queueing Models | |
|---|---|---|---|---|---|---|
| | | | | | Fixed Length [9] | Variable Length [10] |
| $\gamma_k$ | $4.19_{\pm 0.01} \div 3.205_{\pm 0.007}$ | $2.27_{\pm 0.01}$ | $1.241_{\pm 0.001}$ | $1.624_{\pm 0.003}$ | – | – |
| $\gamma_w$ | $1.51335_{\pm 0.00006}$ | $1.637_{\pm 0.003}$ | $1.8483_{\pm 0.0006}$ | $1.930_{\pm 0.002}$ | – | – |
| $\overline{\beta_\tau}$ | $0.53823_{\pm 0.00001}$ | $0.431_{\pm 0.002}$ | $0.3162_{\pm 0.0001}$ | $0.360_{\pm 0.002}$ | 0 | 0.5 |
| $\beta_C$ | $3.39_{\pm 0.01}$ | $0.82_{\pm 0.01}$ | $0.147_{\pm 0.001}$ | $0.430_{\pm 0.002}$ | – | – |
| $\overline{\alpha}$ | $0.58_{\pm 0.01}$ | $0.68_{\pm 0.02}$ | $0.78_{\pm 0.01}$ | $0.70_{\pm 0.01}$ | 1.0 | 0.5 |
| $\sigma$ | $6.6_{\pm 0.1}$ | $6.8_{\pm 0.2}$ | $6.6_{\pm 0.1}$ | $6.6_{\pm 0.1}$ | – | – |
| $\ln \overline{C}$ | $3.4 \div 5.9$ | 4.8 | 5.4 | 3.0 | – | – |

## S5.3 Log-normal and Stretched Exponential Ansatzs

The datasets for which fat tailed distributions offer a remarkable fit can be often well fitted with log-normal or stretched exponential distributions. Mathematically, a quantity $x$ obeys a log-normal distribution if it is drawn from a probability distribution,

$$P(x) \propto A_{LN}(\mu,\sigma)^{-1} x^{-1} e^{-(\ln x-\mu)^2/(2\sigma^2)}, \tag{S22}$$

and stretched exponential if is drawn from

$$P(x) \propto A_{SE}(\lambda,\beta)^{-1} x^{-(1-\beta)} e^{-(\lambda x)^{\beta}}, \tag{S23}$$

where the normalization factors $A_{LN}(\mu,\sigma)^{-1} \equiv \sum_{x=1}^{\infty} x^{-1} e^{-(\ln x-\mu)^2/(2\sigma^2)}$ and $A_{SE}(\lambda,\beta) \equiv \sum_{x=1}^{\infty} x^{\beta-1} e^{-(\lambda x)^{\beta}}$, respectively.

The log-likelihood functions (S19) are

$$\begin{aligned} \hat{l}_{LN}(\mu,\sigma) &= -\frac{1}{n}\sum_{i=1}^{n}\left(\ln x_i + (\ln x_i-\mu)^2/(2\sigma^2) + \ln A_{LN}(\mu,\sigma)\right) \\ \hat{l}_{SE}(\mu,\sigma) &= -\frac{1}{n}\sum_{i=1}^{n}\left((1-\beta)\ln x_i + (\lambda x_i)^{\beta} + \ln A_{SE}(\lambda,\beta)\right) \end{aligned} \tag{S24}$$

respectively. We performed the numerical maximization of $\hat{l}_{LN}(\mu,\sigma)$ and $\hat{l}_{SE}(\mu,\sigma)$ for $P_k(k)$, $P_C(C)$ and $P_w(w)$ for all datasets, and reported the fitting parameters in Tables (S5-S6), finding that $\sigma_k = \overline{\alpha}\sigma_C$ and $\beta_C = \overline{\alpha}\beta_k$ are also approximately satisfied as predicted by Eqs. (S9) and (S11).

Table S5: Log-normal fitting parameters

| | Mobile phone | Email | Twitter | Message |
|---|---|---|---|---|
| $\mu_k$ | $1.1382_{\pm 0.0007}$ | $1.09_{\pm 0.06}$ | $2.81_{\pm 0.01}$ | $0.30_{\pm 0.02}$ |
| $\sigma_k$ | $0.8128_{\pm 0.0006}$ | $1.71_{\pm 0.05}$ | $1.7_{\pm 0.5}$ | $1.632_{\pm 0.007}$ |
| $KC_k$ | 0.014 | 0.026 | 0.045 | 0.011 |
| $\mu_C$ | $2.681_{\pm 0.002}$ | $1.9_{\pm 0.1}$ | $3.60_{\pm 0.02}$ | $1.01_{\pm 0.04}$ |
| $\sigma_C$ | $1.297_{\pm 0.002}$ | $2.40_{\pm 0.07}$ | $2.24_{\pm 0.01}$ | $2.02_{\pm 0.02}$ |
| $KC_C$ | 0.039 | 0.040 | 0.052 | 0.023 |
| $\mu_w$ | $-3.45085$ | $-0.48_{\pm 0.05}$ | $-1.96_{\pm 0.01}$ | $-0.65_{\pm 0.03}$ |
| $\sigma_w$ | 3.42321 | $2.00_{\pm 0.02}$ | $2.06_{\pm 0.001}$ | $1.48_{\pm 0.01}$ |
| $KC_w$ | 0.015 | 0.015 | 0.0030 | 0.0060 |

To compare the statistical appropriateness of the three fat tailed models (power law, log-normal and stretched exponential in Tables S2-S6), we summarize the goodness of fit ($KS$ measures) in Table S7, where the bold number indicates the best and comparable fits. We find that $P_k$ of the mobile phone dataset is best fitted by a power law, evidenced by a remarkably small $KS$ value

Table S6: Stretched exponential fitting parameters

| | Mobile phone | Email | Twitter | Message |
|---|---|---|---|---|
| $\lambda_k$ | $0.259_{\pm 0.001}$ | $0.55_{\pm 0.08}$ | $0.03260_{\pm 0.0005}$ | $1.4_{\pm 0.1}$ |
| $\beta_k$ | $1.013_{\pm 0.005}$ | $0.37_{\pm 0.01}$ | $0.563_{\pm 0.003}$ | $0.38_{\pm 0.01}$ |
| $KC_k$ | 0.035 | 0.028 | 0.045 | 0.0029 |
| $\lambda_C$ | $0.0410_{\pm 0.0005}$ | $0.24_{\pm 0.04}$ | $0.0121_{\pm 0.0003}$ | $0.53_{\pm 0.04}$ |
| $\beta_C$ | $0.782_{\pm 0.001}$ | $0.27_{\pm 0.01}$ | $0.436_{\pm 0.003}$ | $0.335_{\pm 0.006}$ |
| $KC_C$ | 0.0093 | 0.035 | 0.017 | 0.013 |
| $\lambda_w$ | $8000_{\pm 1000}$ | $9_{\pm 1}$ | $290_{\pm 30}$ | $11_{\pm 1}$ |
| $\beta_w$ | $0.133_{\pm 0.002}$ | $0.271_{\pm 0.004}$ | $0.215_{\pm 0.002}$ | $0.336_{\pm 0.006}$ |
| $KC_w$ | 0.016 | 0.012 | 0.0016 | 0.0071 |

compared to the other models, supporting our theoretical predictions (S15). For the Email, Twitter and Online Message datasets, the stretched exponential form offers the best fits, indicating the fact that these distributions are best described by a power law with finite cutoff, an effect widely recognized for social networks. Yet, despite these variations in the functional form of the best fits, the scaling relationships (S7, S9 and S11) derived for these models are equally applicable, implying that the underlying connections between network topology and human dynamics are independent of the particular functional forms of the underlying data.

Table S7: Comparing KS goodness for different models

| | | Mobile phone | Email | Twitter | Message |
|---|---|---|---|---|---|
| $P_k$ | Power Law | **0.003** | **0.023** | 0.19 | 0.11 |
| | Log-normal | 0.014 | **0.026** | **0.045** | 0.011 |
| | Stretched exponential | 0.035 | **0.028** | **0.045** | **0.0029** |
| $P_C$ | Power Law | 0.02 | **0.046** | 0.15 | 0.11 |
| | Log-normal | 0.039 | **0.040** | 0.052 | 0.023 |
| | Stretched exponential | **0.0093** | **0.035** | **0.017** | **0.013** |
| $P_w$ | Power Law | 0.036 | 0.068 | 0.046 | 0.071 |
| | Log-normal | **0.015** | **0.015** | 0.0030 | **0.0060** |
| | Stretched exponential | **0.016** | **0.012** | **0.0016** | **0.0071** |

# S6 Waiting Time Distribution

In this section, we offer a detailed discussion about the waiting time distribution $P_{\tau,i}$.

## S6.1 Definition

We define a link-specific interevent time $\tau_{i\to j}$ as the total number of communication events initiated by user $i$ between two consecutive communications from $i$ to $j$ [11]. For example, $\tau_{A\to C} = 3,4,5$ in Fig. S5, and we define

$$P_{\tau,i}(\tau) \propto \sum_{j,k} \delta(\tau - \tau^k_{i\to j}), \tag{S25}$$

where the superscript $k$ denotes the $k$-th communication between $i$ and $j$.

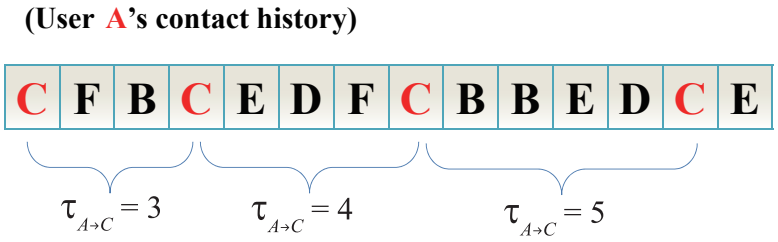


Figure S5: The definition of $\tau_{A\to C}$, the interevent time captures communication intervals between two individuals, A and C. Note that $\tau_{A\to C}$ measures time in terms of the number of events, a feature that corrects for daily fluctuations in the communication volume, but has the same asymptotic scaling as the real interevent time [2].

## S6.2 $P_{\tau,i}$ and $k_i(C_i)$

Denote with $\Pi_i(t_1,t_2)$ the probability that individual $i$ contacts a *new* friend $j$, representing someone that $i$ did not contact in the previous $[t_1,t_2]$ time frame. This requires that the waiting time $\tau_{i\to j}$, that characterizes the communication between $i$ and $j$, be greater than $C_i(t_1,t_2)$. That is, the communication from $i$ to $j$ did not occur in any of the $C_i(t_1,t_2)$ number of previous communications.

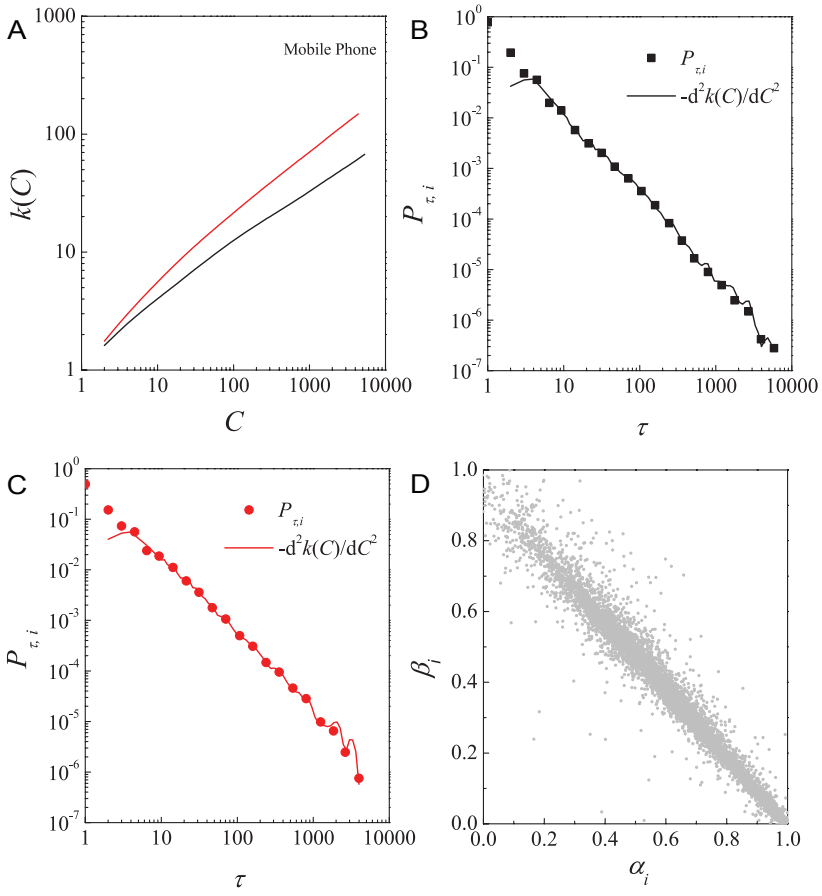


Figure S6: (A) Heaps' plot $k_i(C_i)$ vs. $C_i$ for two mobile phone users. (B-C) The measured individual waiting distribution $P_{\tau,i}(\tau)$ (dots), compared to the predictions of Eqs. S28 (lines) for each user. (D) $\alpha_i$ vs. $\beta_{\tau,i}$ measured for the Mobile phone dataset for users with $C_i > 1000$ and $k_i > 100$, supporting the validity of Eq. (S29).

This implies

$$\Pi_i(t_1, t_2) = 1 - \int_0^{C_i} P_{\tau,i}(\tau)d\tau = \int_{C_i}^{\infty} P_{\tau,i}(\tau)d\tau. \tag{S26}$$

On the other hand, we notice that

$$\Pi_i = dk_i/dC_i. \tag{S27}$$

Comparing Eq. (S26) with Eq. (S27), we obtain the scaling relationship

$$P_{\tau,i}(\tau) = -\left.\frac{d^2k_i(C_i)}{dC_i^2}\right|_{C_i=\tau}. \tag{S28}$$

In Fig. S6, we test the validity of Eq. (S28) for two mobile phone users by comparing $P_{\tau,i}$ to the numeric second derivative of $K_i(C_i)$, demonstrating that despite the variations of $P_{\tau,i}$ and $k_i(C_i)$ Eq. S28 is robust. Equation (S28) also predicts, together with Heaps' law (1) $P_{\tau,i}(\tau) \sim \tau^{-(1+\beta_{\tau,i})}$, satisfying

$$\alpha_i + \beta_{\tau,i} = 1. \tag{S29}$$

In Fig. S6D, we plot $\alpha_i$ vs. $\beta_{\tau,i}$ measured independently for 8,000 mobile phone users with $C_i > 1000$ and $k_i > 100$, finding that while $\alpha_i$ and $\beta_{\tau,i}$ span from 0 to 1 widely they are strongly correlated, validating Eq. (S29).

## S6.3 Activity Dependence

To explore the correlation between $P_{\tau,i}$ and $C_i$, we grouped users based on their activity level $C_i$ and computed $\langle P_{\tau,i}\tau\rangle$ for each user group. Figure S1 shows that by rescaling $\tau$ with the the activity level $C_i$ as $C_i^{-1}\langle P_{\tau,i}\tau\rangle$ and $\tau/C_i$ the obtained distributions collapse approximately into a single curve that follows a power law (spanning over 3 magnitude) with a sharp bounded by one due to the fact $\tau_{i\to j} < C_i$, suggesting that $\left\langle C_i^{-1}P_{\tau,i}\tau/C_i\right\rangle$ is a appropriate population average of $P_{\tau,i}$ (Fig. 1d).

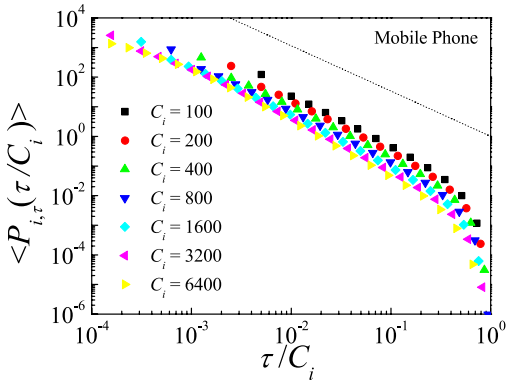


Figure S7: Rescaled waiting time distribution $\left\langle C_i^{-1} P_{\tau,i}(\tau) \right\rangle$ vs. $\tau/C_i$, for different activity groups of mobile phone users.

# S7 Time Evolution of the Studied Distributions

In this section, we report the time evolution of the activity and degree distributions for all datasets and provide further empirical support for our theory.

## S7.1 Activity Distribution

We generated networks for variable observation time frame for each dataset. The activity distributions $P_C(C|t_1,t_2)$ for networks of different time frames $(t_1,t_2)$ are shown in Fig. S8. We find $P_C(C|t_1,t_2)$ collapses into a single curve curve for for Mobile phone, Email and Twitter datasets (Fig. S9A-C) after rescaled $C$ by the time window $\Delta T \equiv t_2 - t_1$, indicating

$$P_C(C|\Delta T) = (\Delta T)^{-1} P_f(C/\Delta T), \tag{S30}$$

where $P_f$ capturing the distribution of individual activity rate $f_i \equiv C_i/\Delta T$ is independent of choice of time frames, implying that the activity patterns are stationary for these three datasets. However, the communication process of Online Message dataset is non-stationary as it breaks the time-translation invariance (Fig. S9D). Instead we find that the Online Message dataset preserves a

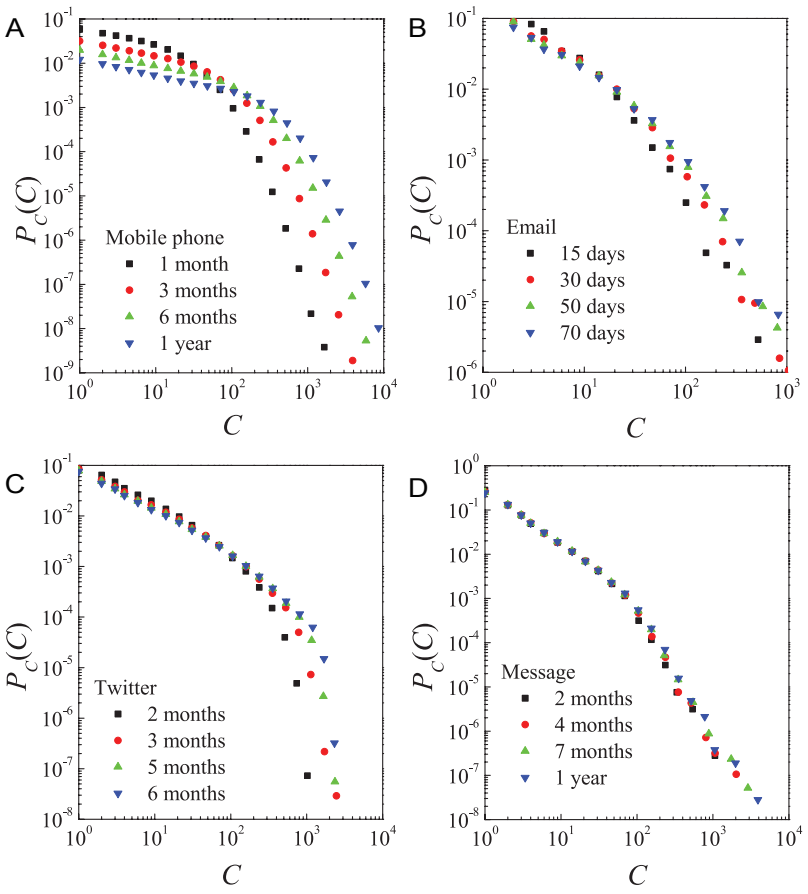


Figure S8: Time evolution of $P_C(C)$ for (A) Mobile phone, (B) Email, (C) Twitter and (D) Online Message, respectively.

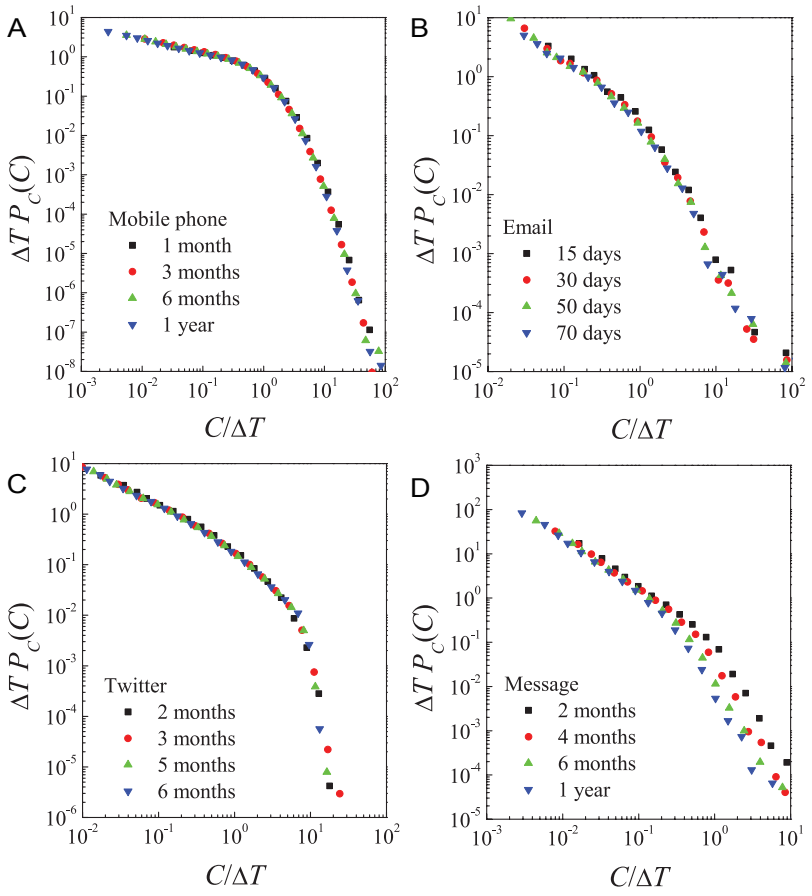


Figure S9: Rescaled $\Delta^T P_C(C)$ vs. $C/\Delta T$ for different time frames for (A) Mobile phone, (B) Email, (C) Twitter and (D) Online Message, respectively.

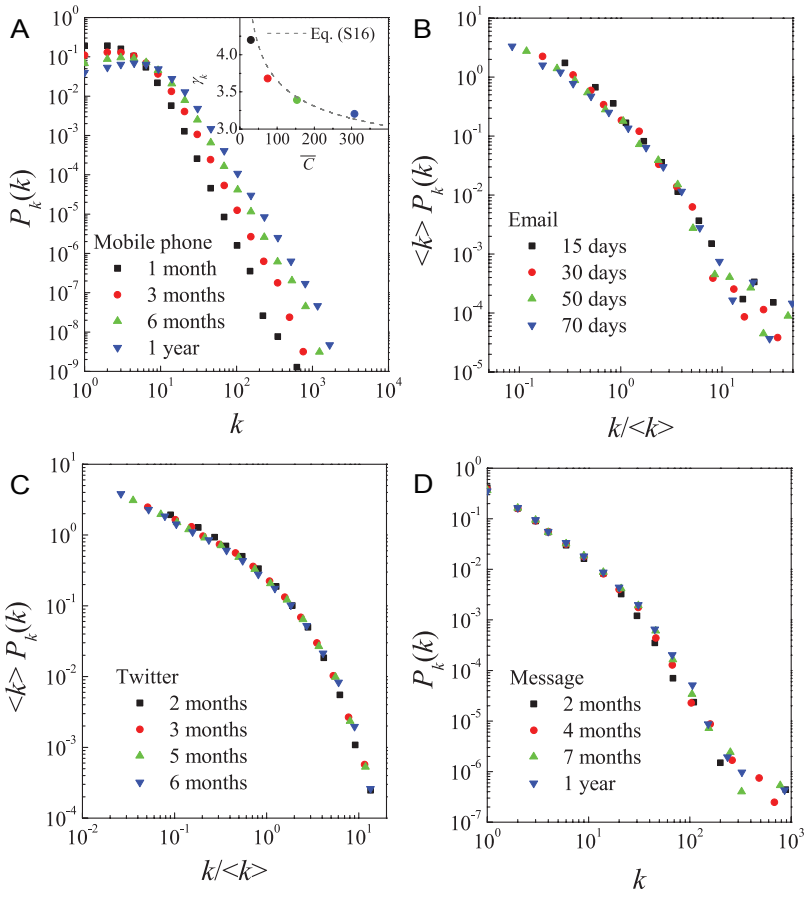


Figure S10: Time evolution of $P_k(k)$ for (A) Mobile phone, (B) Email, (C) Twitter and (D) Online Message, respectively.

scaling invariance since $P_C$ of different time frames simply collapses and the time window $\Delta T$ controls only the maximum activity $C_{max}$ (Fig. S8D).

## S7.2 Degree Distribution

In this section, we report the time evolution of the degree distribution for all dataset and provide further empirical support for our prediction. The degree distributions for networks of different time frames are shown in Fig. S10.

We find that the Mobile phone dataset falls in the *Case 2*, evidenced by a time dependent $P_k$ (Fig. S10A) despite the temporal stationarity of the underlying activity process (Eq. (S30)

and Fig. S9A). The degree distribution has a power law tail $P_k \sim k^{-\gamma_k}$ for different time frames, where the scaling regime spans from 1.5 to 2.5 order of magnitude for the various time frames. We measured the average activity level ($\overline{C}$) for each time frame, and looked at the relationship between $\overline{C}$ and $\gamma_k$, as shown in the inset of Fig. S10A, where the dashed line represents the prediction of Eq. (S16), finding that the empirically observed degree distribution is in good agreement with our prediction.

The $P_k$ of *Case 1* systems (Email, Twitter and Online Message) is determined by Eq. (S5). If the underlying activity process is stationary, by substituting Eq. (S30) we obtain

$$P_k(k) \sim k^{-1} \left( \frac{k}{\Delta T^{\overline{\alpha}}} \right)^{1/\overline{\alpha}} P_f \left( \left( \frac{k}{\Delta T^{\overline{\alpha}}} \right)^{1/\overline{\alpha}} \right), \tag{S31}$$

indicating that the average degree follows $\langle k(\Delta T) \rangle = K \Delta T^{\overline{\alpha}}$, where the constant is $K \equiv \int x^{\overline{\alpha}} P_f(x) dx$. Hence,

$$\langle k \rangle P_k(k) = G(k/\langle k \rangle), \tag{S32}$$

with the universal function

$$G(x) \sim x^{1/\overline{\alpha}-1} P_f \left( (Kx)^{1/\overline{\alpha}} \right). \tag{S33}$$

Equation (S32) indicates that $P_k$ of stationary-activity driven systems corresponding different time frames can be collapsed into a single master curve by rescaling $k$ with the average degree $\langle k \rangle$, as confirmed in Fig. S10BC for the Email and Twitter datasets. The activity pattern of the Online Message datasets, however, is non-stationary, preserving a scaling invariance (Fig. S8D, Fig. S9D), Eq. (S5) predicting a simple time-dependence of the degree distribution. That is, $P_k$ from different time frames simply collapse without rescaling where the time window $\Delta T$ affects only the maximum degree $k_{max}$, as confirmed in Fig. S10D. Moreover, the functional forms of master curves in Fig. S10BCD are also well predicted by Eq. (S5) (see Fig. 3A-C).

Therefore, our theory correctly predicts not only the static measures, but also the temporal

evolution of the underlying social network characteristics.

# References


[1] 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. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.

[2] J.P. Eckmann, E. Moses, and D. Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences*, 101(40):14333, 2004.

[3] M. Cha, H. Haddadi, F. Benevenuto, and K.P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.

[4] Petter Holme, Christofer R. Edling, and Fredrik Liljeros. Structure and time evolution of an internet dating community. *Social Networks*, 26(2):155 – 174, 2004.

[5] D. Rybski, S.V. Buldyrev, S. Havlin, F. Liljeros, and H.A. Makse. Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences*, 106(31):12640, 2009.

[6] Bernat Corominas-Murtra, Jordi Fortuny, and Ricard V. Solé. Emergence of zipfs law in the evolution of communication. *Physical Review E*, 83:32767, 2011.

[7] Aaron Clauset, Cosma R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(2):661–703, 2009.

[8] B. Efron and R. Tibshirani. *An introduction to the bootstrap*, volume 57. Chapman & Hall/CRC, 1993.

[9] A. Cobham. Priority assignment in waiting line problems. *Journal of the Operations Research Society of America*, 2:70–76, 1954.

[10] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[11] M. Karsai, K. Kaski, A.L. Barabási, and J. Kertész. Universal features of correlated bursty behaviour. *Scientific Reports*, 2(397):397, 2012.