Scaling laws of human interaction activity

Diego Rybski[1], Sergey V. Buldyrev[2], Shlomo Havlin[3],

Fredrik Liljeros[4], and Hernán A. Makse[1]

[1]*Levich Institute and Physics Department,*

City College of New York, New York, NY 10031, USA

[2]*Department of Physics, Yeshiva University, New York, NY 10033, USA*

[3]*Minerva Center and Department of Physics,*

Bar-Ilan University, Ramat-Gan 52900, Israel

[4]*Department of Sociology, Stockholm University, S-10691 Stockholm, Sweden*

(Dated: May 15, 2009)

SUPPORTING INFORMATION (SI)

Scaling laws of human interaction activity

Diego Rybski, Sergey V. Buldyrev, Shlomo Havlin,
Fredrik Liljeros, and Hernán A. Makse

I. NOTATION

1. Member j sends his/her nth message at time $t_j(n)$, where $1 \leq n \leq M_j$ and M_j is the total number of messages sent by j in the time of data acquisition T. The sequence of counts defined as the number of messages in the period δt, is given by

$$\mu_j^{\delta t}(t) = \sum_{n, t_j(n) \in [t, t+\delta t]} a_j(n)\,, \tag{1}$$

where $a_j(n) = 1$. In addition, the periods are non-overlapping, $t = i\delta t$ with integer i, and therefore $1 \leq t_j(n) \leq T$. In the case of daily resolution $\delta t = 1$ day.

2. The cumulative number of messages that a member sends until time t is:

$$m_j^{\delta t}(t) = \sum_{t'=1}^{t} \mu_j^{\delta t}(t')\,. \tag{2}$$

In particular, $m_j(1) = \mu_j(1)$ and $m_j(T) = M_j$.

3. The displacement of the random walk is the cumulative sum of the normalized $\mu_j^{\delta t}(t)$:

$$Y_j^{\delta t}(t) = \sum_{t'=1}^{t} (\mu_j^{\delta t}(t') - \langle \mu_j^{\delta t}(t) \rangle)\,, \tag{3}$$

where $\langle \mu_j^{\delta t}(t) \rangle$ is the average of $\mu_j^{\delta t}(t)$ in time t. The root-mean-square displacement after Δt is defined as

$$F_j^{\delta t}(\Delta t) = \sqrt{\langle [Y_j^{\delta t}(t+\Delta t) - Y_j^{\delta t}(t)]^2 \rangle_t}\,, \tag{4}$$

where the average is performed over the time t. Additionally, we perform an average over members j with activity level M and define

$$(F^{\delta t}(\Delta t))_M^2 = \langle (F_j^{\delta t})^2 | M \rangle_j\,. \tag{5}$$

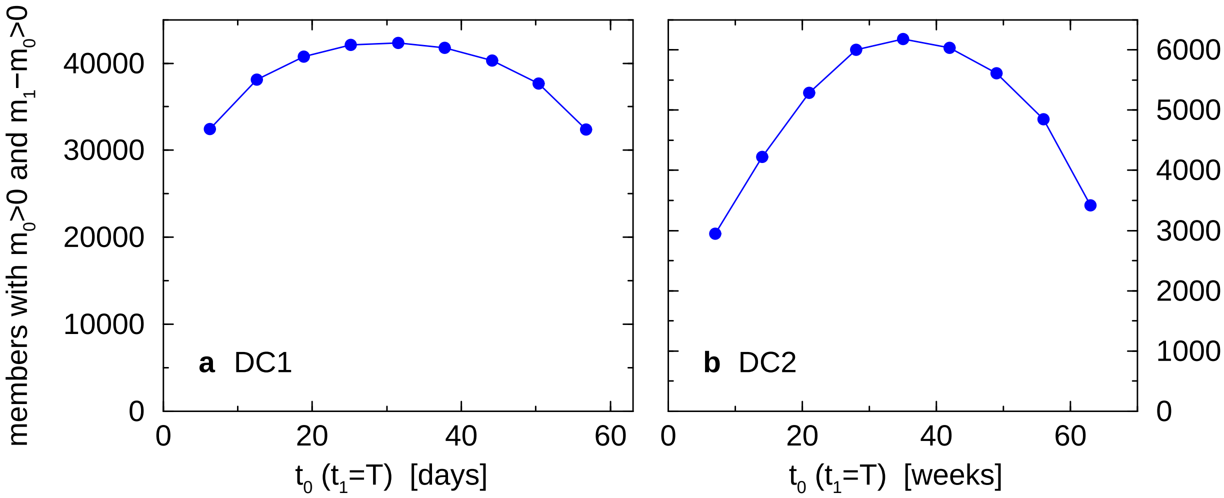

FIG. 1: Optimal times t_0 and t_1. The panels show for **a**, OC1, and **b**, OC2, the number of members with both, $m_0 > 0$ and $m_1 - m_0 > 0$. While t_1 obviously is optimal at the end of the period, t_0 is varied to find the value for which the number of members – with at least one message until t_0 and at least one new message between t_0 and t_1 – is maximal.

4. For simplicity, in the main text we skip the index j as well as δt and write $\mu(t)$, $m(t)$, $Y(t)$, as well as $F(\Delta t)$.

5. To investigate the growth in the number of messages we use the quantities $r = \ln \frac{m_1}{m_0}$, $\langle r(m_0)\rangle$, $\sigma(m_0)$ and the exponents β_{OC1}, β_{OC2}, β_{G}, β_{rnd}.

6. To investigate the growth of the degree we use the quantities $r_k = \ln \frac{k_1}{k_0}$, $\langle r_k(k_0)\rangle$, $\sigma_k(k_0)$ and the exponents $\beta_{k,\mathrm{OC1}}$; $\beta_{k,\mathrm{OC2}}$.

7. For the growth of the degree in the preferential attachment model we use the quantities $r_{\mathrm{PA}} = \ln \frac{k_1}{k_0}$, $\langle r_{\mathrm{PA}}(k_0)\rangle$, $\sigma_{\mathrm{PA}}(k_0)$ and the exponent β_{PA}.

II. OPTIMAL TIMES t_0 AND t_1

Figure 1 displays the optimal times t_0 and t_1 to calculate the growth rates for OC1 (panel a) and OC2 (panel b).

III. DETAILS ON THE QUANTIFICATION OF LONG-TERM CORRELATIONS USING DETRENDED FLUCTUATION ANALYSIS

Statistical dependencies between the values of a record $\mu(t)$ with $t = 1, \ldots, T$ can be characterized by the auto-correlation function

$$C(\Delta t) = \frac{1}{\sigma_\mu^2 (T - \Delta t)} \sum_{t=1}^{T-\Delta t} \left[\mu(t) - \langle \mu(t) \rangle\right] \left[\mu(t+\Delta t) - \langle \mu(t) \rangle\right] , \tag{6}$$

where T is the length of the record $\mu(t)$, $\langle \mu(t) \rangle$ its average, and σ_μ its standard deviation. For *uncorrelated* values of $\mu(t)$, $C(\Delta t)$ is zero for $\Delta t > 0$, because on average positive and negative products will cancel each other out. In the case of *short-term correlations* $C(\Delta t)$ has a characteristic decay time $\Delta t_\times$. A prominent example is the exponential decay $C(\Delta t) \sim \exp(-\Delta t / \Delta t_\times)$. *Long-term correlations* are described by a slower decay, e.g. diverging $\Delta t_\times$, namely a power-law,

$$C(\Delta t) \sim (\Delta t)^{-\gamma} , \tag{7}$$

with the correlation exponent $0 < \gamma < 1$.

Detrended Fluctuation Analysis (DFA) is a well studied method to quantify long-term correlations in the presence of non-stationarities [1]. The analysis of a considered record $\mu(t)$ of length T consists of 5 steps:

1. Calculate the cumulative sum, the so-called profile:

$$Y(t) = \sum_{t'=1}^{t} \left(\mu(t') - \langle \mu(t) \rangle\right) . \tag{8}$$

2. Separate the profile $Y(t)$ into $T_{\Delta t} = \mathrm{int} \frac{T}{\Delta t}$ segments of length Δt. Often, the length of the record is not a multiple of Δt. In order not to disregard information, the segmentation procedure is repeated starting from the end of the record and one obtains $2T_{\Delta t}$ segments.

3. Locally detrend each segment ν by determining best polynomial fits $p_\nu^{(n)}(t)$ of order n and subsequently subtract it from the profile:

$$Y_{\Delta t}(t) = Y(t) - p_\nu^{(n)}(t) . \tag{9}$$

4. Calculate for each segment the variance (squared residuals) of the detrended $Y_{\Delta t}(t)$

$$F^2_{\Delta t}(\nu) = \frac{1}{\Delta t} \sum_{j=1}^{\Delta t} \left(Y^2_{\Delta t} \left[(\nu - 1)\Delta t + j \right] \right) \tag{10}$$

by averaging over all values in the corresponding νth segment.

5. The DFA fluctuation function is given by the square-root of the average over all segments:

$$F(\Delta t) = \left[\frac{1}{2T_{\Delta t}} \sum_{\nu=1}^{2T_{\Delta t}} F^2_{\Delta t}(\nu) \right]^{1/2} . \tag{11}$$

The averaging of $F^2_{\Delta t}(\nu)$ is additionally performed over members of similar activity level M.

If the record $\mu(t)$ is long-term correlated according to a power-law decaying auto-correlation function, Eq. (7), then $F(\Delta t)$ increases for large scales Δt also as a power-law:

$$F(\Delta t) \sim (\Delta t)^H , \tag{12}$$

where the fluctuation exponent H is analogous to the well-known Hurst exponent [2]. The exponents are related via

$$H = 1 - \gamma/2 \ , \qquad \gamma = 2 - 2H . \tag{13}$$

When $\gamma = 1$ then $H_{\mathrm{rnd}} = 1/2$, that is the case of uncorrelated dynamics. If the correlations decay faster than $\gamma > 1$ then the random exponent $H_{\mathrm{rnd}} = 1/2$ is still recovered. Long-term correlations imply $0 < \gamma < 1$ and $1/2 < H < 1$. In practice, one plots $F(\Delta t)$ versus Δt in double-logarithmic representation, determines the exponent H on large scales and quantifies the correlation exponent γ. The order of the polynomials $p_\nu^{(n)}$ determines the detrending technique which is named DFAn, DFA0 for constant detrend, DFA1 for linear, DFA2 for parabolic, etc.

The subtraction of the average in Eq. (8) is only necessary for DFA0. By definition the corresponding fluctuation function is only given for $\Delta t \geq n + 2$. The detrending order determines the capability of detrending. Since the local trends are subtracted from the profile, only trends of order $n - 1$ are subtracted from the original record $\mu(t)$. Throughout the paper we show the results using DFA2 which we found to be sufficient in terms of detrending.

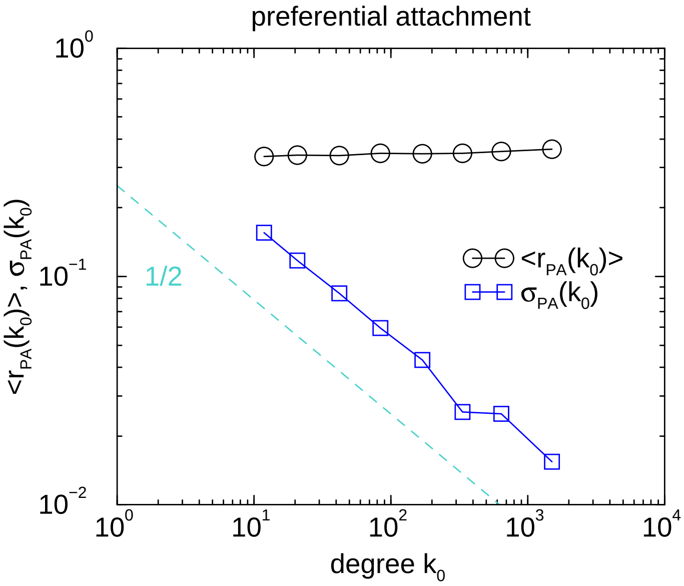

FIG. 2: Growth properties of the preferential attachment model [3] discussed in the main text. We plot the average (black circles) and standard deviation (blue squares) of the growth rate $r_{\rm PA}$ conditional to k_0, the degree of the corresponding nodes at the first stage.

Since the fluctuation functions $F(\Delta t)$ for single users are very noisy, it is useful to average fluctuation functions among various members. Thus, we first group the members in logarithmic bins according to their activity level, the total number of messages M sent. Namely, we group all members that send 1-2, 3-7, 8-20, . . . messages in the period of data acquisition by using bins determined by $b = \text{int}\,(\ln M)$. Next we average the fluctuation function among all members from each group b and obtain for every activity level of the members one DFA fluctuation function. The error bars in Fig. 3a,c of the main text were obtained by subdividing each group and determining the standard deviations of the fluctuation exponents from different groups of the same activity level.

IV. GROWTH IN THE DEGREE

Figure 2 shows the results of the average growth rates and fluctuations of the growth rates as a function of the initial degree for the preferential attachment model [3]. We find a constant average growth rate and a standard deviation decreasing as a power law with exponent $\beta_{\rm PA} = 1/2$ in Eq. (7) in the main text.

The PA network model has been described analytically. In particular, it has been shown that each nodes' degree increases as

$$k(t) \sim \left(\frac{t}{t^*}\right)^b , \tag{14}$$

where t^* is the time when the corresponding node was introduced to the system and b is the dynamics exponent in growing network models ($b = 1/2$ for the standard PA) [4]. Accordingly, here the growth rate, Eq. (6) in the main text, is $r_{\rm PA} = \frac{1}{2}\ln\frac{t_1}{t_0}$, which we also find in Fig. 2.

To obtain $\sigma_{\rm PA}(k_0)$ one can use analogous considerations as for $\sigma(m_0)$ in the main text. Due to Eq. (6) in the main text, here we have

$$r_{\rm PA} \approx \frac{1}{k_0}\sum_{t=1}^{\Delta t}\kappa(t) , \tag{15}$$

where $\kappa(t)$ are small increments analogous to $\mu(t)$, whereas Eq. (14) implies

$$\kappa(t) \sim (\Delta t)^{-1/2} . \tag{16}$$

As before, the conditional standard deviation of the growth rate is

$$\langle [r_{\rm PA}(k_0) - \langle r_{\rm PA}(k_0)\rangle]^2\rangle \approx \frac{1}{k_0^2}\sum_{i}^{\Delta t}\sum_{j}^{\Delta t}\sigma_\kappa^2 C(j-i) . \tag{17}$$

In the uncorrelated case $C(j-i) = \delta_{ij}$, the double sum can be reduced to a single one:

$$\sigma_{\rm PA}^2(k_0) = \frac{1}{k_0^2}\sum_{i}^{\Delta t}\sigma_\kappa^2(i) . \tag{18}$$

As shown below, $\sigma_\kappa(i) \sim i^{-1/4}$, and integration leads to

$$\sigma_{\rm PA}^2(k_0) \sim \frac{1}{k_0^2}\int^{\Delta t} i^{-1/2}\mathrm{d}i \tag{19}$$

$$\sim \frac{1}{k_0^2}(\Delta t)^{1/2} . \tag{20}$$

Eliminating Δt using $k \sim t^{-1/2}$, Eq. (14), one obtains

$$\sigma_{\rm PA}(k_0) \sim k_0^{-1/2} . \tag{21}$$

That is, we obtain $\beta_{\rm PA} = 1/2$ as found numerically.

Remains to show $\sigma_\kappa(t) \sim t^{-1/4}$. We assume new links are set according to a Poisson process, whereas every new link of a node represents an event. The intervals between these events (asymptotically) follow an exponential distribution $p(\tau) = \lambda e^{-\lambda\tau}$. Accordingly, $\kappa(t)$ is a sequence of zeros and only one when a new link is set to the corresponding node. The standard deviation of this sequence is

$$\sigma_\kappa \sim \lambda^{1/2}\,. \tag{22}$$

Due to Eq. (14) the rate parameter decreases like

$$\lambda(t) \sim t^{-1/2}\,. \tag{23}$$

Accordingly,

$$\sigma_\kappa(t) \sim t^{-1/4}\,. \tag{24}$$

In order to extend the standard PA model, a fitness model has been introduced [5] taking into account different fitnesses of the nodes of acquiring links and therefore involving a distribution of b-exponents. The spread of growth rates r could be related to the distribution of fitness. On the other hand, the growth according to Eq. (14) is superimposed with random fluctuations that we characterize with the exponent β.

---

[1] Peng C-K, *et al.* (1994) Mosaic organization of DNA nucleotides. *Phys Rev E* 49:1685–1689.

[2] Feder J (1988) *Fractals*, Physics of Solids and Liquids. (Plenum Press, New York).

[3] Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512.

[4] Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74:47–97.

[5] Bianconi G, Barabási, A-L (2001) Competition and multiscaling in evolving networks. *Europhys Lett* 54:436-442.