

# Excitable human dynamics driven by extrinsic events in massive communities

Joachim Mathiesen<sup>a,1</sup>, Luiza Angheluta<sup>b</sup>, Peter T. H. Ahlgren<sup>a</sup>, and Mogens H. Jensen<sup>a</sup>

<sup>a</sup>Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark; and <sup>b</sup>Physics of Geological Processes, Department of Physics, University of Oslo, 0316 Oslo, Norway

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved September 14, 2013 (received for review March 5, 2013)

**Using empirical data from a social media site (Twitter) and on trading volumes of financial securities, we analyze the correlated human activity in massive social organizations. The activity, typically excited by real-world events and measured by the occurrence rate of international brand names and trading volumes, is characterized by intermittent fluctuations with bursts of high activity separated by quiescent periods. These fluctuations are broadly distributed with an inverse cubic tail and have long-range temporal correlations with a  $1/f$  power spectrum. We describe the activity by a stochastic point process and derive the distribution of activity levels from the corresponding stochastic differential equation. The distribution and the corresponding power spectrum are fully consistent with the empirical observations.**

social networks | statistical mechanics | extreme events

Online social networks have emphatically changed the way people interact. The development of network theories and growth in available data on human behavior (1–4) has prompted an explosive interest in research on the evolution of behaviors (5–7) and social structures (8, 9). Among the many forms of online social media, microblogging services such as Twitter (10–12) are characterized by a real-time dynamics with large numbers of user broadcasts related to real-world events. Twitter is a popular microblogging platform where a registered user can submit small pieces of information, named “tweets,” that are either private or made public to the user’s followers. The length of a tweet is limited to 140 characters and its content ranges widely from personal updates to massively distributed advertisements or political messages. Twitter has a global outreach and, hence, is used by an increasing number of companies and political organizations to disseminate news. To a large extent, Twitter users can be seen as direct social sensors to measure the popularity of various topics. A large part of recent research on Twitter uses user activity as predictor for real-world events including the dynamics of stock market prices (13), box office revenues (14), real-time detection of the location of earthquakes hitting populated areas (15), and for opinion mining and political sentiment analysis (16). Large-scale behavioral data from other online media have been shown to have a similar predicting power, e.g., Google query volumes have been used to detect early signs of stock market moves (17) or more general movements in society (18).

Whereas behavioral data from Twitter have been suggested to predict many real-world events or have been used in mapping out social networks, the statistics of the combined user activity are not well understood. Here we suggest a stochastic model for the user activity in massive online communities. Our model sheds light on the statistical properties of the large-scale user activity on Twitter as well as the underlying correlations. Similar to the user activity on Twitter, trading volumes on the stock market reflect the interest that investors have in particular securities or products at given prices. Interestingly, as we will point out below, we find that the trading activity on financial securities is quite similar to the large-scale user activity on Twitter.

We have automatically queried Twitter for tweets containing one or more international brand names (*SI Text*). These tweets

appear at highly irregular time intervals (Fig. 1), reflecting intermittent user activity levels. We consider the broadcasting of tweets to be a random point process with large fluctuations in the time intervals between the online appearances of messages. The number of tweets containing a certain brand “A” as a function of  $t$  is therefore given by a time signal  $g_A(t)$  composed of isolated events occurring at random times  $t_i$ ,  $g_A(t) = \sum_i \delta(t - t_i)$ , where the index  $i$  refers to a specific posting event. For each query, a number of the latest tweets  $n_A \leq 1,500$  is returned. Thus, we determine an average tweet rate for a given query  $k$  as

$$\gamma_A(t_{n_A}) = \frac{1}{t_{n_A} - t_1} \int_{t_1}^{t_{n_A}} dt g_A(t) = \frac{n_A(t_{n_A})}{t_{n_A} - t_1}, \quad [1]$$

where  $t_1$  and  $t_{n_A}$  correspond to the time of the oldest, respectively, latest, tweet returned by the query  $k$ . The time interval  $\tau_k = t_{n_A} - t_1$  for a fixed number  $n_A$  of tweets is a highly fluctuating variable from one query to another. We notice that the appearance of tweets on Twitter resembles a nonhomogeneous Poisson process with random fluctuations in the average tweet rate  $\gamma_A$ .

We have further collected data for the trading volume of selected equities over a period similar to that covered by the Twitter data. In particular, we consider the trading volume in three shares: Apple Inc. (AAPL), Nokia Corporation (NOK), and Green Mountain Coffee Roasters Inc. (GMCR). The number of shares traded for each security was accumulated and sampled in 1-min intervals. Changes in trading volume and price are known to be highly leptokurtic (19), have long-range correlations (20), and intriguing scaling properties (21, 22). We consider the volume as

## Significance

Online social networks have over the last decade influenced the way people interact. Data from Twitter allow for a detailed study of the activity in online massive communities. By studying the frequency by which international brands appear on Twitter and the trade of financial securities on financial markets, we find a characteristic bursty behavior of the activity levels of Twitter users and market participants. We explain the bursty behavior by a simple model of the large-scale human behavior and quantify the correlations in the activity levels. The statistical similarity of the two social systems is an indication that the complex process underlying individual decision-making might not be very different for Twitter users and participants in financial markets.

Author contributions: J.M., L.A., P.T.H.A., and M.H.J. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

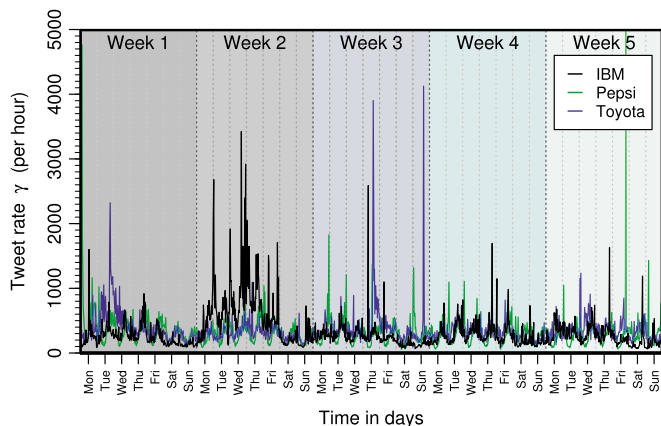
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: mathies@nbi.dk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1304179110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1304179110/-DCSupplemental).



**Fig. 1.** Temporal variation of tweet rates of three international brands, IBM, Pepsi, and Toyota. The time signals are for all brands intermittent, i.e., they have longer periods of relatively steady activity levels interrupted by sudden high-activity spikes. The time signals are all modulated by an underlying periodic variation over days and weeks.

a simple proxy for the temporal interest that the market has in a given security, disregarding more complex effects that might influence the price formation process.

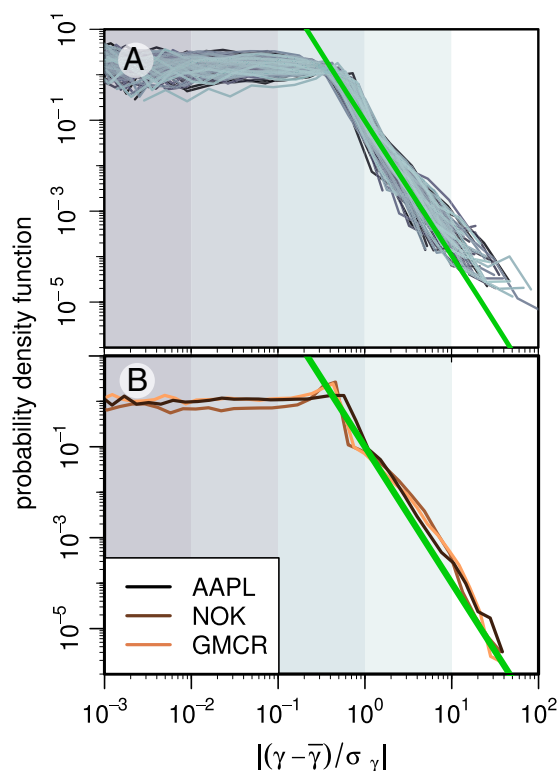
In the Twitter data, we distinguish two types of user activity, one where users post messages independently of other tweets and one where users interact directly, e.g., by reposting information from other users in so-called retweets or by submitting responses to existing tweets. In general, a retweet contains text from the original tweet together with a reference to the author who posted it. Retweets typically form a smaller subset of all tweets, and the frequency by which individual tweets are retweeted follows a power law with an exponent similar to the out-degree distribution of Twitter users (*SI Text*). The out-degree is here measured by the number of followers of a user, i.e., the number of people that directly receive tweets posted by that user. This suggests that the rate of information spread by reposting is proportional with the number of followers and is limited by the local network topology. However, because there are only few tweets that generate a large flux of retweets, this may not be the most efficient way to diffuse information between Twitter's users. There are information pathways that are not only related to the network topology, but, to a larger degree, correspond to many users tweeting at the same time about the same thing triggered by events outside the network.

The intermittent dynamics of individuals has previously been modeled in terms of a timing selection mechanism between different tasks (23). The prioritization of various tasks is suggested to lead to a bursty dynamics with power-law distributed waiting times. This is in contrast with a homogeneous Poisson process, where the waiting time between tasks that are being selected at random follows an exponential distribution. Here, we propose a global measure of collective human behavior and introduce a stochastic model for the global activity rate associated with many interacting individuals in a large social organization. The activity rate is characterized by long-term memory effects as well as nonexponential distributed waiting times.

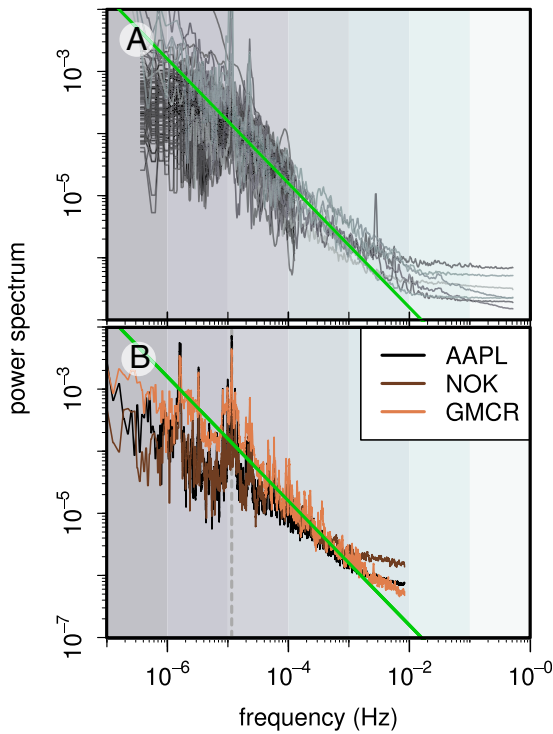
## Results

Interestingly, the intermittent tweet rates of specific brands follow a distribution  $P(\gamma)$  with a power-law tail with an exponent  $\beta$  close to an inverse cube,  $\beta = 2.9 \pm 0.4$  (SD), as seen in Fig. 2 and in *SI Text*. For the trading volumes, we achieve values for the exponent  $\beta_{\text{AAPL}} = 2.9$ ,  $\beta_{\text{NOK}} = 3.1$ , and  $\beta_{\text{GMCR}} = 3.0$ . Moreover, the fluctuations in the flux of tweets are long-range correlated with a power spectrum that decays as  $1/f^\alpha$ , where  $f$  is the

frequency and  $\alpha = 1.0 \pm 0.4$  (SD), in an intermediate frequency window corresponding to timescales from 20 min to 24 h as shown in Fig. 3. This means that tweets posted at a given time are influenced with an equal strength by tweets on all timescales ranging back as far as  $\sim 24$  h. At the same time, high bursts of new tweets, extreme events, occur in the tail of  $P(\gamma)$ . The  $1/f$  noise is a widespread phenomenon observed in a variety of different systems, including voltage fluctuations (24), heartbeats (25), freeway traffic (26), music (27), and trades in financial securities (20), along with many other examples. Although there is not a unified theory that would apply to all systems exhibiting  $1/f$  noise, there are numerous models that reproduce the  $1/f$  fluctuations in temporal signals with fluctuations drawn from different distributions. On the other hand, there are also plenty of studies that focus on the non-Gaussian, power-law statistics of time signals,  $P(\gamma) \sim \gamma^{-\beta}$ , independent of their power spectrum. More recent studies on stochastic point processes investigated the relationship between Pareto-type distributions of the variables and their  $1/f^\alpha$  power spectrum, e.g., refs. 28–30. The idea behind a stochastic point process is to model the average waiting time between random, discrete events by a multiplicative noise process. Essentially, the complexity of scale-free distributed variables with a  $1/f^\alpha$  spectral density emerges from the multiplicative noise. Here we present a stochastic point process that captures both the  $P(\gamma) \sim \gamma^{-3}$  and  $1/f$ -noise features. Other models of correlated human behavior (31) predict similarly a power-law distribution of the activity rates. However, these



**Fig. 2.** Probability density functions of (A) brand tweet rates and (B) trading volume rates for stocks. The mean values have been subtracted from the individual rates and the rates have been normalized by their respective SDs. The density functions collapse and reveal a common scaling behavior for relatively large rates. The green line added to both A and B is a guide to the eye and is consistent with a scaling exponent of  $-3$ . In B, we consider trading activities in the companies Apple Inc. (AAPL), Nokia Corporation (NOK), and Green Mountain Coffee Roasters Inc. (GMCR).



**Fig. 3.** Power spectra of the activity rates  $\gamma(t)$  for the individual brands in *SI Text* (A) and stocks (B). For high frequencies  $0.1\text{--}24\text{ h}^{-1}$ , the spectra have a characteristic  $1/f$  behavior with a crossover to white noise at very high frequencies. The green line is a guide to the eye and corresponds to a  $1/f$  behavior.

models possess a weaker memory effect and do not reproduce the scaling exponent for the power spectrum that we observe.

We assume a scenario where the human activity in the case of no external input is determined by a natural drift toward inactivity. That is, the waiting time  $\tau$  since the last activity increases with time  $\partial_t \tau = 1$ . On the other hand, excitation by external events drives the system toward higher activity levels. Without correlations in the user activity, the rate  $\gamma(t)$  is determined by a balance between the drift toward inactivity and repeated excitation. We assume that the correlation in the human dynamics is controlled by a current waiting time between events in the shape of a multiplicative noise with an amplitude given by  $\sqrt{\tau}$ . The stochastic process for the average waiting time is therefore determined by a stochastic differential equation of the form

$$\frac{d\tau}{dt} = 1 + f(\tau) + \sqrt{\tau}\eta(t), \quad [2]$$

where  $\eta$  is Gaussian noise with zero mean and unit variance, and the deterministic part  $f(\tau)$  is chosen such that the process attains a nontrivial stationary distribution. Collective interactions between users sending messages on Twitter are effectively modeled by the intrinsic, multiplicative noise. The amplitude of the intrinsic noise term is proportional to  $\sqrt{\tau}$ , which implies that if the dynamics was solely driven by noise, the waiting time  $\tau$  would have an absorbing state, i.e.,  $\tau=0$ , corresponding to a tweeting activity that is constant in time and never ceases. The drift term equal to 1 is added to mimic that, if nothing happens, the average waiting time would increase linearly with time as mentioned above. Due to this constant drift term, the absorbing state  $\tau=0$  is never attained, although there are sudden excursions in its neighborhood. The stationary probability distribution function of waiting times  $F(\tau)$  corresponding to Eq. 2 is obtained as the

steady-state solution of the Fokker–Plank equation in the Ito formulation and is given as

$$F(\tau) \sim \tau e^{-2 \int_0^\tau f(\tau')/\tau' d\tau'}, \quad [3]$$

apart from the normalization constant. The divergence at large  $\tau$  is suppressed by the cutoff function which depends on  $f(\tau)$ . However, at small  $\tau$ , corresponding to large tweet rates  $\gamma$ , the function  $f(\tau)$  is irrelevant because  $F(\tau) \sim \tau$ . Thus, in the scaling regime, we can safely ignore  $f(\tau)$ . Using that  $\gamma \sim 1/\tau$ , the stochastic dynamics for  $\gamma$  follows from Eq. 2 by Ito's lemma and is given as

$$\frac{d\gamma}{dt} = \gamma^{3/2}\eta(t), \quad [4]$$

where we ignore the contributions due to  $f(\tau)$  that only determine the range over which  $\gamma(t)$  is power-law distributed as  $P(\gamma) \sim \gamma^{-3}$ . The power spectrum of tweet rate fluctuations is determined by the joint distribution  $P(\gamma, t; \gamma', t')$  associated with the stochastic process in Eq. 4. By the Wiener–Khinchine theorem, the power spectral density of  $\gamma$  is related to the correlation function as

$$S(f) = 4 \int_0^\infty dt \langle \gamma(0)\gamma(t) \rangle \cos(2\pi ft), \quad [5]$$

where the correlation function  $\langle \gamma(t)\gamma(0) \rangle$  is defined as

$$\langle \gamma(t)\gamma(0) \rangle = \int d\gamma d\gamma' P(\gamma, t; \gamma', 0) \gamma \gamma'. \quad [6]$$

As with other point processes, one may assume that the transition distribution has an eigenfunction expansion, and therefore

$$P(\gamma, t; \gamma', 0) \approx \sum_n P_n(\gamma) P_n(\gamma') e^{-\lambda_n t} \quad [7]$$

where the unnormalized probability eigenfunctions  $P_n(\gamma)$  satisfy the master equation corresponding to Eq. 1 and are given by

$$-\lambda_n P_n(\gamma) = \frac{1}{2} \frac{d^2}{d\gamma^2} (\gamma^3 P_n(\gamma)). \quad [8]$$

Combining Eqs. 6 and 7, we have that

$$\langle \gamma(t)\gamma(0) \rangle = \sum_n \gamma_n^2 e^{-\lambda_n t}, \quad [9]$$

where  $\gamma_n = \int_0^\infty d\gamma \gamma P_n(\gamma)$  is the first moment of the probability eigenfunction. By its Fourier transform as in Eq. 5, the power spectrum can be written as a sum of Lorentzian spectra

$$S(f) \approx 4 \sum_n \frac{\gamma_n^2 \lambda_n}{\lambda_n^2 + 4\pi^2 f^2}. \quad [10]$$

The relation between  $\gamma_n$  and the eigenvalues  $\lambda_n$  follows from the structure of the unnormalized eigenfunction obtained from Eq. 7 and given as  $P_n(\gamma) = \gamma^{-5/2} J_1(2\sqrt{2\lambda_n/\gamma})$ , where  $J_1(x)$  is the Bessel function of the first kind. Hence, the first moment of it is  $\gamma_n = (2\lambda_n)^{-1/2}$ . The regime  $S(f) \sim 4 \int_0^\infty \frac{d\lambda}{\lambda^2 + 4\pi^2 f^2} = 1/f$  is obtained when  $\gamma_n \sim \lambda_n^{-1/2}$ .

The joint appearance of the  $1/f$  fluctuations distributed with an inverse cubic law for the flux of tweets and volume of trades has interesting implications. Although there is no consensus on

a unique underlying process generating  $1/f$  noise, we associate this kind of fluctuations with complex systems that exhibit an increase in structures and information due to a long-term memory. In general, the collective interactions in relation to trading and tweeting exhibit the characteristics of an emergent phenomenon.

Intermittent fluctuations in the rate of tweets on Twitter can happen by two types of information pathways: (i) the influx of tweets triggered from the outside world onto Twitter (many independent users tweet about the same thing) and (ii) cascades of retweets or replies to existing tweets. Because there is a larger influx of new tweets compared with the rate of retweets, we conclude that most of the information spread across the online network happens in a “one-step” cascade when many unrelated people tweet about the same thing.

The activity on Twitter may not be very different from the dynamics of stock trades on financial markets because both are influenced by the social behavior inside massive communities combined with simple rules on the interaction set by, e.g., the

platform through which the individuals interact. Furthermore, the similarity on large scales indicates a common feature in the complex process underlying the decision-making of users on Twitter and participants in financial markets.

## Materials and Methods

From the Twitter Application Programming Interface (API), we automatically query for tweets on Twitter containing one or more of 92 preselected international brand names. The full list of brand names is given in *SI Text*. For each query to Twitter a maximum number of 1,500 tweets is returned by the API. Each returned tweet has a time stamp, which can be used to estimate an average tweet rate. The dataset used in this study was created by monitoring the public timeline for a period of 4 mo, November 2010 to February 2011; 2 mo, January–February 2012; and 2 mo, October–November 2012. During these periods, we computed the tweet rates of selected international brands with a sampling rate down to half a minute.

**ACKNOWLEDGMENTS.** Suggestions and comments by Anders Blok are gratefully acknowledged. This study was supported by the Danish National Research Foundation through the Center for Models of Life.

- King G (2011) Ensuring the data-rich future of the social sciences. *Science* 331(6018): 719–721.
- Ginsberg J, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012–1014.
- Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323(5916):892–895.
- Oliveira JG, Barabási A-L (2005) Human dynamics: Darwin and Einstein correspondence patterns. *Nature* 437(7063):1251.
- Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464(7291):1025–1028.
- Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2009) Scaling laws of human interaction activity. *Proc Natl Acad Sci USA* 106(31):12640–12645.
- Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197.
- Arenas A, Danon L, Diaz-Guilera A, Gleiser PM, Guimerà R (2004) Community analysis in social networks. *Eur Phys J B* 38(2):373–380.
- Blondel V, Guillaume J, Lambiotte R, Lefebvre E. (2008) Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp* 2008:P10008.
- Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web (Association for Computing Machinery, New York)*, pp 591–600.
- Mandavilli A (2011) Peer review: Trial by Twitter. *Nature* 469(7330):286–287.
- Huberman BA, Romero DM, Wu F (2009) Crowdsourcing, attention and productivity. *J Inf Sci* 35:758–765.
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: Real-time event detection by social sensors. *Proceedings of the 19th International Conference on World Wide Web (Association for Computing Machinery, New York)*, pp 851–860.
- Tumasjan A, Sprenger TO, Sandner PG, Welpe IM (2010) Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM* 10:178–185.
- Bollen J, Mao H, Zeng J (2011) Twitter mood predicts the stock market. *Comp. Sci.* 2:1–8.
- Asur S, Huberman BA (2010) Predicting the future with social media. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference (IEEE, Los Alamitos, CA)*, Vol 1, pp 492–499.
- Preis T, Moat HS, Stanley HE (2013) Quantifying trading behavior in financial markets using Google Trends. *Sci Rep* 3:1684.
- Choi H, Varian H (2012) Predicting the present with Google Trends. *Econ Rec* 88:2–9.
- Cont R (2001) Empirical properties of asset returns: Stylized facts and statistical issues. *Quant Finance* 1:223–236.
- Bonanno G, and Lillo, F. and Mantegna, R.N. (2000), Dynamics of the number of trades of financial securities. *Physica A* 280(1):136–141.
- Mantegna RN, Stanley HE (1995) Scaling behaviour in the dynamics of an economic index. *Nature* 376:46–49.
- Gabaix X, Gopikrishnan P, Plerou V, Stanley HE (2003) A theory of power-law distributions in financial market fluctuations. *Nature* 423(6937):267–270.
- Barabási A-L (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207–211.
- Voss RF, Clarke J (1976) Flicker ( $1/f$ ) noise: Equilibrium temperature and resistance fluctuations. *Phys Rev B* 13:556.
- Kobayashi, M and Musha, T (1982),  $1/f$  fluctuation of heartbeat period, *IEEE Trans Biomed Eng.* 6:456–457.
- Zhang X, Hu G (1995)  $1/f$  noise in a two-lane highway traffic model. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 52(5):4664–4668.
- Voss RF, Clarke J (1978)  $1/f$  noise in music: Music from  $1/f$  noise. *J Acoust Soc Am* 63:258.
- Kaulakys B, Gontis V, Alaburda M (2005) Point process model of  $1/f$  noise vs a sum of Lorentzians. *Phys Rev E Stat Nonlin Soft Matter Phys* 71(5 Pt 1):051105.
- Ruseckas J, Kaulakys B (2010)  $1/f$  Noise from nonlinear stochastic differential equations. *Phys Rev E Stat Nonlin Soft Matter Phys* 81(3 Pt 1):031105.
- Erland S, Greenwood PE, Ward LM (2011)  $1/f^{\alpha}$  noise is equivalent to an eigenstructure power relation. *Europhys Lett* 95:60006.
- Karsai M, Kaski K, Barabási A-L, Kertész J (2012) Universal features of correlated bursty behaviour. *Sci Rep* 2:397.