Supporting Information

Mathiesen et al. 10.1073/pnas.1304179110

SI Text

Details are provided on the data fitting and the derivation of model properties.

Data Fitting. The scaling exponent of the tail in the distribution of the fluctuations of the tweet and trading volume rates γ was estimated from a best power-law fit for the part of the distribution where $(\gamma - \overline{\gamma})/\sigma_\gamma > 1$. Here $\overline{\gamma}$ and σ_γ denotes the mean and SD, respectively. The distribution of the fitted exponents is shown in Fig. S1.

The Bursty User Activity Is Primarily Excited by Exo-Social Media Events. Reposts, also called retweets, typically contain information from the original tweet ("seed") together with the letters "RT" and a unique identification code of the retweet. We collect data on retweets by sampling random reposts from the public stream of tweets. For a given retweet, we find other retweets that originate from the same tweet. We estimate the retweet rate γ_{RT} of individual tweets and users by using the number n_{RT} of similar retweets returned by a query and time over which they appeared.

$$\gamma_{RT}(t) = \frac{n_{RT}(t)}{t - t_1}. \qquad \textbf{[S1]}$$

In the lack of access to a complete stream of new tweets, we estimate the distribution of retweet rates conditional on that a given retweet is observed $p(\gamma_{\mathrm{RT}}|\mathrm{observed})$. The distributions of expected retweet rates to a given user and to specific tweets are shown in Fig. S2*A*. We notice that both distributions have a broad, power-law tail over several decades with an exponent $\alpha \approx 1.7 \pm 0.1$ (SD), i.e., $p(\gamma_{\mathrm{RT}}|\mathrm{observed}) \propto \gamma_{RT}^{-1.7}$. The probability density of retweet rates $p(\gamma_{\mathrm{RT}})$ can be determined from Bayes' theorem as

$$p(\gamma_{\mathrm{RT}}) = \frac{p(\gamma_{\mathrm{RT}}|\mathrm{observed})}{p(\mathrm{observed}|\gamma_{\mathrm{RT}})} p(\mathrm{observed}) \sim \gamma_{\mathrm{RT}}^{-2.7}, \qquad \textbf{[S2]}$$

using that the conditional probability to observe a repost $p(\mathrm{observed}|\gamma_{\mathrm{RT}})$ is proportional to the frequency of RTs γ_{RT}.

First, we look at the subset of tweets that is associated with how information is spread on the Twitter network as a result of direct interaction between users, i.e., how often users repost information from other users. The scale-free distribution of retweets implies that whereas few tweets/users are more popular and, thus, massively retweeted, the vast majority of tweets/users have very few or absent reposts. Interestingly, the tail of $p(\gamma_{RT})$ has an exponent ~ -2.7 that is close to the out-degree distribution of users on Twitter. This suggests that the rate of information spread by reposting and the celebrity of a user depend on the user's followers, namely the degree of connectivity on the Twitter network. The more followers a user has, the more likely it is that his/her post gets reposted. 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 across Twitter's users.

Retweets are a smaller subset of tweets in the stream. Thus, there are more information pathways that are not only related to the network topology, but, to a larger degree, they correspond to many users tweeting at the same time about the same thing. These tweets quantify the flux of information from the outside world to Twitter, and thus could be used as sensors of real-time events in the world. We find that on Twitter, the community's interest in a given subject is highly intermittent with long quiescent periods interrupted by relatively few bursts of online activity, as seen in a time snapshot of γ_t in Fig. 1. This intermittency is modulated by the day/night activity.

A measure of the rare events is to look at tail distribution of the tweet rate increments $\Delta\gamma_t = |\gamma_{t+\Delta t} - \gamma_t|$ measured in hourly intervals, which has a clear stretched exponential tail

$$p(\Delta\gamma_t) \propto \exp\left(-\kappa|\Delta\gamma_t|^\beta\right), \qquad \textbf{[S3]}$$

where $\beta = 0.3 \pm 0.1$ as determined from Fig. S2*B*.

List of Brands. Google, IBM, Apple, Microsoft, Coca-Cola, McDonald's, Marlboro, General Electric, Vodafone, Hewlett-Packard, Walmart, Blackberry, Amazon, UPS, Tesco, Visa, Oracle, Verizon, SAP, BMW, Nintendo, BP, Exxon, Disney, Carrefour, Nokia, Accenture, Intel, L'Oreal, American Express, Mercedes, Citi, T-Mobile, Pepsi, Nike, H&M, Porsche, Dell, MasterCard, Samsung, FedEx, Baidu, eBay, Zara, Siemens, Starbucks, Nissan, Hermès, Barclays, Sony, Gucci, Ikea, Adidas, Puma, Dolce & Gabbana, Louis Vuitton, Cisco, Honda, Toyota, Nestlé, Nescafé, Budweiser, Smirnoff, Canon, Kellogg's, Reuters, Colgate, Volkswagen, Danone, Audi, Santander, Panasonic, Hyundai, Shell, Nivea, Corona, Carlsberg, Adobe, Credit Suisse, Allianz, Cartier, Rolex, Ferrari, Heineken, Lancome, Burberry, UBS, Yahoo, Philips, Kleenex, Moët & Chandon, and Goldman Sachs.

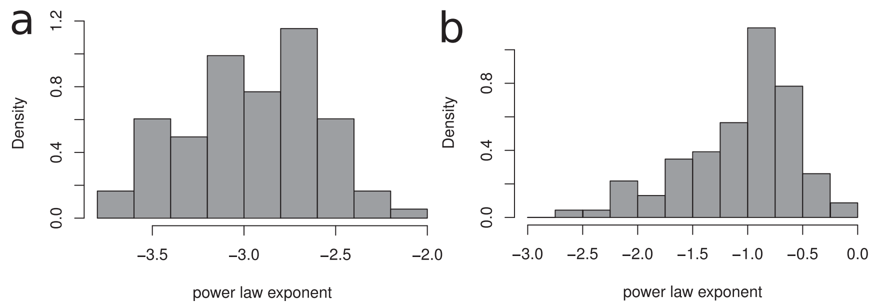

Fig. S1. Distribution of the best estimates of (*A*) the scaling exponents of the rate distributions and (*B*) power spectra for the individual brands presented in Fig. 2 of the rate distributions. The mean values are for *A*, $-3.0 \pm 0.4(\mathrm{SD})$ and for *B*, $-0.9 \pm 0.4(\mathrm{SD})$.

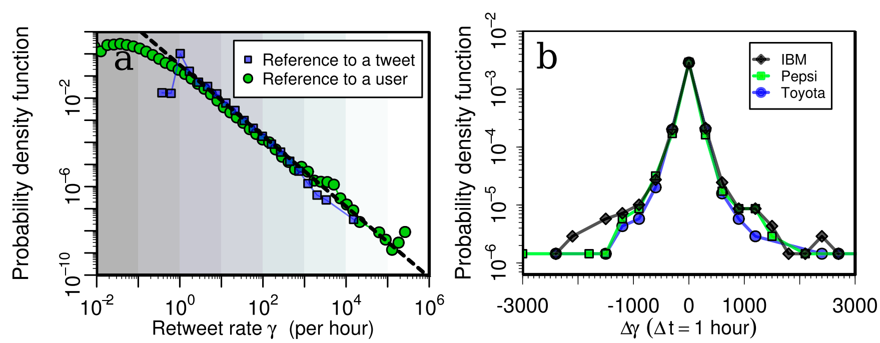

Fig. S2. Intermittent dynamics of user communication on Twitter. (*A*) Probability distribution function of retweet rates of individual tweets (blue squares) and users (green circles). The distribution is presented using double logarithmic axes. The dashed line is a best fit to the scale-free distribution with an exponent $\alpha = 1.70 \pm 0.05$ (SD). (*B*) Distributions of the tweet rate change $\Delta\gamma_t$ measured in hourly intervals for the intermittent time series shown in Fig. 1.