Spontaneous emergence of social influence in online systems

Jukka-Pekka Onnela[a,b,c,d,e] and Felix Reed-Tsochas[e,f,1]

[a]Harvard Medical School, Harvard University, Boston, MA 02115; [b]Harvard Kennedy School, Harvard University, Cambridge, MA 02138; [c]Department of Physics, University of Oxford, Oxford OX1 3PU, United Kingdom; [d]Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, FIN-02015 HUT, Finland; [e]CABDyN Complexity Centre, Saïd Business School, University of Oxford, Oxford OX1 1HP, United Kingdom; and [f]Institute for Science, Innovation and Society, Saïd Business School, University of Oxford, Oxford OX1 1HP, United Kingdom

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved September 9, 2010 (received for review December 17, 2009)

Social influence drives both offline and online human behavior. It pervades cultural markets, and manifests itself in the adoption of scientific and technical innovations as well as the spread of social practices. Prior empirical work on the diffusion of innovations in spatial regions or social networks has largely focused on the spread of one particular technology among a subset of all potential adopters. Here we choose an online context that allows us to study social influence processes by tracking the popularity of a complete set of applications installed by the user population of a social networking site, thus capturing the behavior of all individuals who can influence each other in this context. By extending standard fluctuation scaling methods, we analyze the collective behavior induced by 100 million application installations, and show that two distinct regimes of behavior emerge in the system. Once applications cross a particular threshold of popularity, social influence processes induce highly correlated adoption behavior among the users, which propels some of the applications to extraordinary levels of popularity. Below this threshold, the collective effect of social influence appears to vanish almost entirely, in a manner that has not been observed in the offline world. Our results demonstrate that even when external signals are absent, social influence can spontaneously assume an on–off nature in a digital environment. It remains to be seen whether a similar outcome could be observed in the offline world if equivalent experimental conditions could be replicated.

collective behavior | social networks | fluctuation scaling

Social influence captures the ways in which people affect each others' beliefs, feelings, and behaviors. It has traditionally been within the domain of social psychology with a particular focus on microlevel processes among individuals (1), but it also plays a prominent role across the social sciences, for example in the study of contagion in sociology (2), herding behavior in economics (3), speculative bubbles in financial markets (4), voting behavior (5), and interpersonal health (6). Social influence plays an especially important role in cultural markets (7), for products such as books and music, and generally pervades any arena of life where the attitudes and tastes of individuals are influenced by others.

It is often useful to distinguish between local and global sources of influence, which typically are identified with an individual's interpersonal environment and the mass media, respectively (8). The overall social influence arises from a mixture of local and global influences, which themselves emerge from different signals. The fact that these two processes operate at very different scales poses considerable challenges for the empirical study of social influence. For the purposes of our study, we define (*i*) *local signal* as information on the behavior of individuals who are friends or acquaintances of ego, the person whose behavior is being analyzed, and (*ii*) *global signal* as information on the aggregate behavior of the population. Note that these definitions rely on the potentially observable behaviors of others as opposed to the nonobservable ones, such as their intentions or feelings. This framework incorporating local and global signals is very generic, and possible behaviors range from the consumption of cultural products to making lifestyle choices.

The structures of social influence are most naturally addressed from the perspective of social network analysis (9). The notion of local influence presupposes that individuals are embedded in a social network that channels and directs how behaviors spread. Examples of such behaviors include innovation adoption among physicians (10), as well as other empirical and theoretical studies of diffusion (11–15). The notion of global influence, on the other hand, presupposes that individuals have information on the aggregate popularity of products and behaviors. Although a given social network can be used as a proxy for communicating any behavioral signals, one should ideally have access to a network that accurately represents the potential communication channels for a specific local signal as these channels may vary across behaviors. In addition, individuals are often selective as to what information they choose to disclose to their friends, resulting in the local signal being necessarily incomplete, biased, or misrepresented (16). Similarly, whereas accurate population-level statistics exist for popular items, it is much harder to find statistics for more marginal products and behaviors.

A novel opportunity to study human behavior in a setting that overcomes these methodological limitations is provided by certain online environments. These systems have the advantage of allowing access to complete subpopulations of agents. When combined with appropriate tools of analysis, they enable the direct study of collective macrolevel social behavior in very large social systems without sampling. We study a complete online social system with well-defined local and global signals by harnessing data from Facebook, a hugely popular social networking site, which at the time of data collection had ≈50 million active users worldwide. In addition to the current popular interest in social networks, scholars have recognized the potential of these and other social websites for research (17–22), reflecting the current move to using rich large-scale datasets on human behavior and communication (23, 24). Facebook users, in line with other social networking sites, can construct a public or semipublic profile within a bounded system, articulate a list of other users, "Facebook friends," with whom they share a connection, and view and traverse their own connections and those made by others within the system (25).

Facebook users can also install (and uninstall) applications (Fig. 1*A*) that enable them, for instance, to play poker and compare their taste in movies with their friends. Users who are friends can see all of each other's applications simply by visiting the respective profile. In addition, whenever a user adopts a new application, her friends are automatically notified by the system. (This applied at the time the data were collected, but Facebook

SOCIAL SCIENCES

Author contributions: J.-P.O. and F.R.-T. designed research; J.-P.O. coded and ran the simulations; J.-P.O. and F.R.-T. analyzed data; and J.-P.O. and F.R.-T. 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.

[1]To whom correspondence should be addressed. E-mail: felix.reed-tsochas@sbs.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0914572107/-/DCSupplemental.

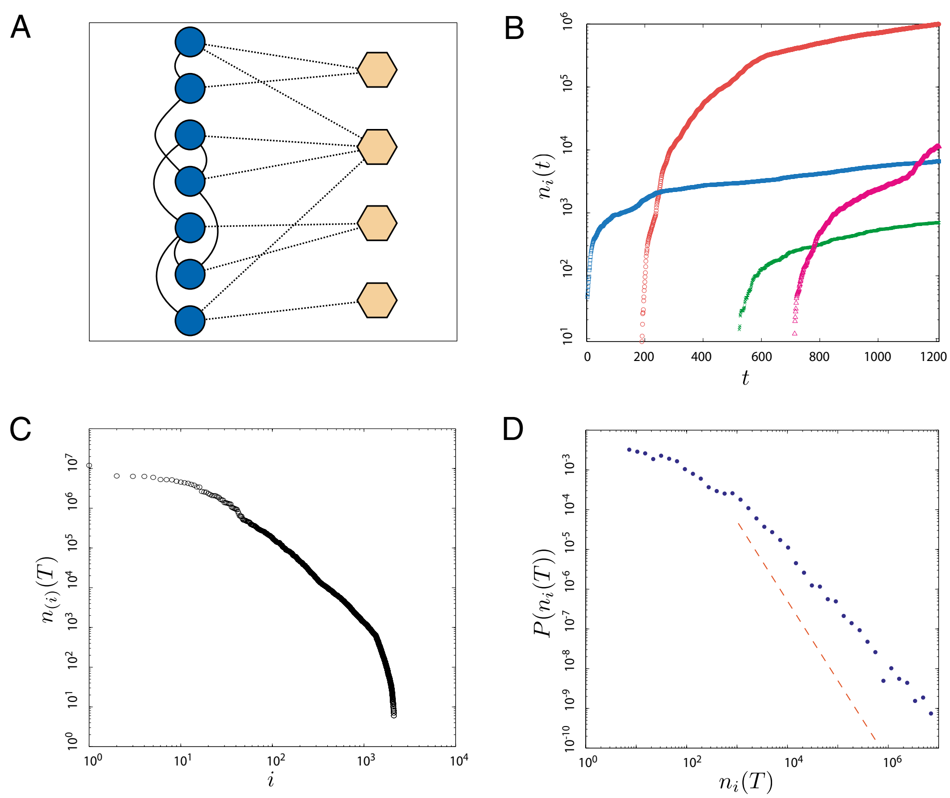

Fig. 1. Facebook users and applications. (*A*) The users (round nodes) form a social network (solid lines) which influences their behavior in adopting applications (hexagons). The dashed lines connecting users to applications indicate which applications each user has installed. (*B*) Number of users $n_i(t)$ as a function of time t for four applications of which "Texas HoldEm Poker" is the most popular one at the end when $t = T$. (*C*) Number of users $n_i(T)$ sorted in descending order for the 2,123 applications that have $n_i(T) > 0$ (Zipf plot). (*D*) Probability density distribution $P(n(T))$ versus $n(T)$ is fat-tailed. The dashed line $\sim n(T)^{-2}$ corresponds to the limit where the mean of the distribution diverges.

subsequently discontinued this practice.) Consequently, users with many Facebook friends are then, at least in principle, in a position to influence a larger number of other users. In addition, everyone has access at all times to an all-inclusive listing of applications ranked by their current global popularity, which acts as an effective "best-seller" list. Although applications are free of charge, popular applications have the advantage of being readily discoverable (low search cost), and are more likely to be of higher quality both with respect to reliability (exhaustively tested) and functionality (superior features). The applications provide recreational value and can be seen as cultural goods, and the different ways users process local and global signals in choosing applications reflect their personal preferences, that is, the underlying heterogeneity of the population.

In addition to the distinction between local and global signals, it is important to classify systems into two separate categories based on whether their dynamics are *endogenous* without external drivers, or *exogenous* and driven externally, or both. These distinctions are useful because they identify the fundamental characteristics of the system, and hence enable systematic comparisons with other systems. Epidemic spreading in a closed system is an example of an endogenous process with local transmission, because the pathogens need to be passed from one person to another in close physical proximity. Similarly, it is possible to model the spread of innovations such as the uptake of new hybrid crops by farmers as an endogenous social contagion process, and to try to distinguish between different types of local processes that may underlie the observed rate at which the innovation is adopted (26). However, studies of social influence which focus on local and endogenous processes such as word-of-mouth transmission are almost always open to the challenge that they neglect equally important exogenous effects such as marketing or mass advertising, and typically trying to separate these two confounding factors is highly problematic. For instance, a reanalysis (27) of the classic diffusion studies on how prescriptions for an antibiotic drug spread among physicians in different communities (10, 11) suggests that marketing efforts, in this context corresponding to external drivers, can account for most of the observed behavior. Although in general both endogenous and exogenous effects may be present in both online and offline systems, as part of our research design we have identified a system that does not have an exogenous component. Instead, both local and global signals are generated endogenously within the system, that is, there is no exogenous driver. This is in contrast to classic innovation diffusion models (e.g., 28), which feature one rate of contagion from within the group (local signal) and another externally imposed (as opposed to endogenously generated) rate of contagion from outside the group (global signal). Another important feature is that here the user has an active role in deciding whether or not to adopt an application.

We downloaded the data from Facebook for all existing 2,720 applications between June 25 and August 14, 2007, shortly after applications were introduced. These data consist of time series $n_i(t)$ with $i = 1,2,\ldots,M = 2{,}720$ and $t = 1,2,\ldots,T = 1{,}208$ corresponding to the aggregate number of users who have application i installed at time t (Fig. 1*B*). Data for 15 applications were partly corrupted and were consequently omitted from the analysis, leaving us with 2,705 applications, or 99% of the data. These data cover 100% of the population of 50 million potential adopters and 99% of all applications that may be adopted, in practice giving us a complete view of system-wide adoptions. Importantly, studying all of the applications avoids a selection bias, which is generated by examining the trajectories of those applications that spread successfully, as tends to be done in most studies on social influence (29). Successful products in cultural markets have been found to be orders of magnitude more popular than the average cultural product (7). This finding is also manifest in the case of Facebook applications. The number of users at the end of the time horizon, $n_i(T)$, sorted in descending order is shown in Fig. 1*C*. For the 10 most popular applications, these numbers vary between $n_{(1)}(T) \approx 12$ million and $n_{(10)}(T) \approx 4.6$ million, whereas $n_{(100)}(T) \approx 180{,}000$ and $n_{(1{,}000)}(T) \approx 1{,}300$. The probability density distribution for the number of application installations (Fig. 1*D*) has a very fat tail and decays so slowly that even its mean value diverges in the limit of infinite system size.

Each new installation, in addition to increasing the overall user base of the application and thus its global signal, also generates a local signal, through which the adopter may in turn influence the future behavior of his friends. Each installation thus acts as a microscopic social stimulus and creates a form of positive feedback in the system. Note that the observable behavior which generates patterns of social influence in this case is restricted to the adoption of an application, rather than its use. Given that the users are part of a very large social network, the consequences of adopting an application are not limited to a user's immediate neighborhood, but may percolate further in the network. This underlines the importance of having data that reflect the behavior of the entire system even if the underlying microscopic data are not available. Whereas the impact of a single installation is admittedly minute, the superposition of the observed 104 million application installations leaves behind a detectable footprint.

To study the effect of social influence, that is, the extent to which the behavior of an individual (his installing an application) is related to the behavior of others (their installing the same application), we turn to the method of fluctuation scaling (FS). This allows us to extract a key signature of the system's behavior purely on the basis of the above aggregate data. FS has been applied successfully to a number of complex systems whose interacting elements participate in some dynamic process. Examples of application domains range from fluctuations in population sizes in ecology to fluctuations in stock-trading activity in financial markets (30–32). Here we outline how FS can be used in the current problem, and refer the reader to *SI Text* for details. For a given application i, the act of individual j regarding installation of the application is encoded by the random variable $S_{i,j}(t)$, where $S_{i,j}(t) = 1$ corresponds to him installing the application at time t, and $S_{i,j}(t) = 0$ corresponds to him doing nothing. From the stochastic process point of view, one can think of each individual tossing coins at every time step, one per application, to decide whether he will install the given application. In terms of this analogy, each individual has several coins, one per application. The probability of individual j installing application i, that is, the probability to obtain $S_{i,j} = 1$ per time step, incorporates many sources of uncertainty, including his disposition and the properties of the application. The probability of obtaining $S_{i,j} = 1$ therefore varies not only from person to person but also from application to application. Social influence, operating through local and global signals, is likely to render the coin tosses dependent for any given application (Fig. 2 *A* and *B*). To measure the strength of social influence, we define *net activity* $f_i(t)$ of application i at time t as

$$f_i(t) \equiv n_i(t) - n_i(t-1) = \sum_{j=1}^{N} S_{i,j}(t) = \sum_{k=1}^{N-n_i(t)} S_{i,j_k}(t), \qquad [1]$$

which corresponds to the net increase in the number of installations for application i between times $t-1$ and t. It can be expressed in terms of the individual constituent variables as shown, where the first sum is taken over all N individuals whereas the latter sum is taken over potential new installers, with the subset of indices $j_1, j_2, \ldots, j_{N-n_i(t)} \in \{1,2,\ldots,N\}$ such that $S_{i,j_k}(\tau) = 0$ for $\tau < t$. In terms of the above analogy, once a user has installed a given application, he stops tossing the particular coin corresponding to that application.

According to FS, the temporal average and SD of $f_i(t)$ are related through the relationship $\sigma_i \sim \mu_i^{\alpha}$. This motivates us to identify a region in which the relationship between $\log \mu_k$ and $\log \sigma_k$ for different values of k is linear. The value of the *fluctuation scaling exponent* α is given by the slope of the line. Although α lies in the rather narrow range [1/2, 1], its value is crucial as an indicator of statistical coupling in the system (Fig. 2 *A* and *B*). If the behavior of a user is independent of the behavior of others, one would expect $\alpha = 1/2$, whereas if her behavior is fully correlated with others one would expect $\alpha = 1$ for all applications. We estimate the mean and SD of the entire activity time series using $\langle f_i \rangle \equiv \mu_i = \frac{1}{T_i}\sum_{t=1}^{T_i} f_i(t)$ and $\sigma_i = \left(\frac{1}{T_i - 1}\sum_{t=1}^{T_i}[f_i(t) - \langle f_i \rangle]^2\right)^{1/2}$, where T_i is the application-specific time series length reflecting the fact that different applications were introduced at different times.

Results

As shown in Fig. 2*C*, applications with $\log(\mu_i) > \log(\mu_x)$ define the *collective regime* governed by $\alpha_C \approx 0.85$, which indicates strong correlations among the constituent variables, that is, the underlying "coin tosses." Application installations above this point are influenced by the behavior of others. Unexpectedly and contrary to previous empirical studies of other systems (32), breakpoint analysis (*SI Text*) shows that the system exhibits another qualitatively different regime for the less popular applications. This *individual regime* with $\log(\mu_i) < \log(\mu_x)$ has $\alpha_I \approx 0.55$, which is very close to the limiting case of $\alpha = 1/2$, meaning that application installations are nearly uncorrelated and social influence is negligible. The transition between the two regimes takes place at approximately $\log(\mu_x) = 0.36$, which translates into an average daily activity of $24 \times 10^{0.36} \approx 55$ new installations a day. We emphasize that theoretical considerations guided our choice to fit a linear function to the data in Fig. 2*C* as opposed to, say, trying to find the best fit among a class of curvilinear functions. Although it would be interesting to also resolve the precise location and nature of the transition (sharp or continuous), we are unable to make this distinction on the basis of the empirical data. However, the central finding on the existence of two different regimes remains unaffected.

The interpretation of FS exponents in terms of correlations assumes that the underlying stochastic process is stationary (32). However, the fact that $n_i(t)$ increases over time demonstrates that the system cannot be stationary. Phrased in terms of the earlier analogy with coin tossing, the number of coins being tossed per round decreases as those who have adopted an application stop tossing the coin. The question then becomes whether the system is sufficiently close to stationarity so that the fluctuation scaling exponents can be given the above interpretation, that is, whether the number of coins that are being tossed remains approximately constant. Let us impose the stringent condition that the system is sufficiently close to stationarity when at most 1% of users have the application installed (meaning that 99% of users continue tossing the coins, leaving the stochastic process almost unaltered). We show in *SI Text* that even under this strict condition, 98% of the time series are stationary. This also means that the scaling in Fig. 2*C* holds for over two orders of magnitude *above* the crossover point. We conclude that the system is sufficiently stationary so that the temporal fluctuations may indeed be given the above interpretation.

As a simple explanatory hypothesis for the observed behavior, one might suggest that the different scaling properties result from applications having different lifetimes. To test this, we divide the applications into three distinct groups based on their time of introduction such that each group covers an equally long

SOCIAL SCIENCES

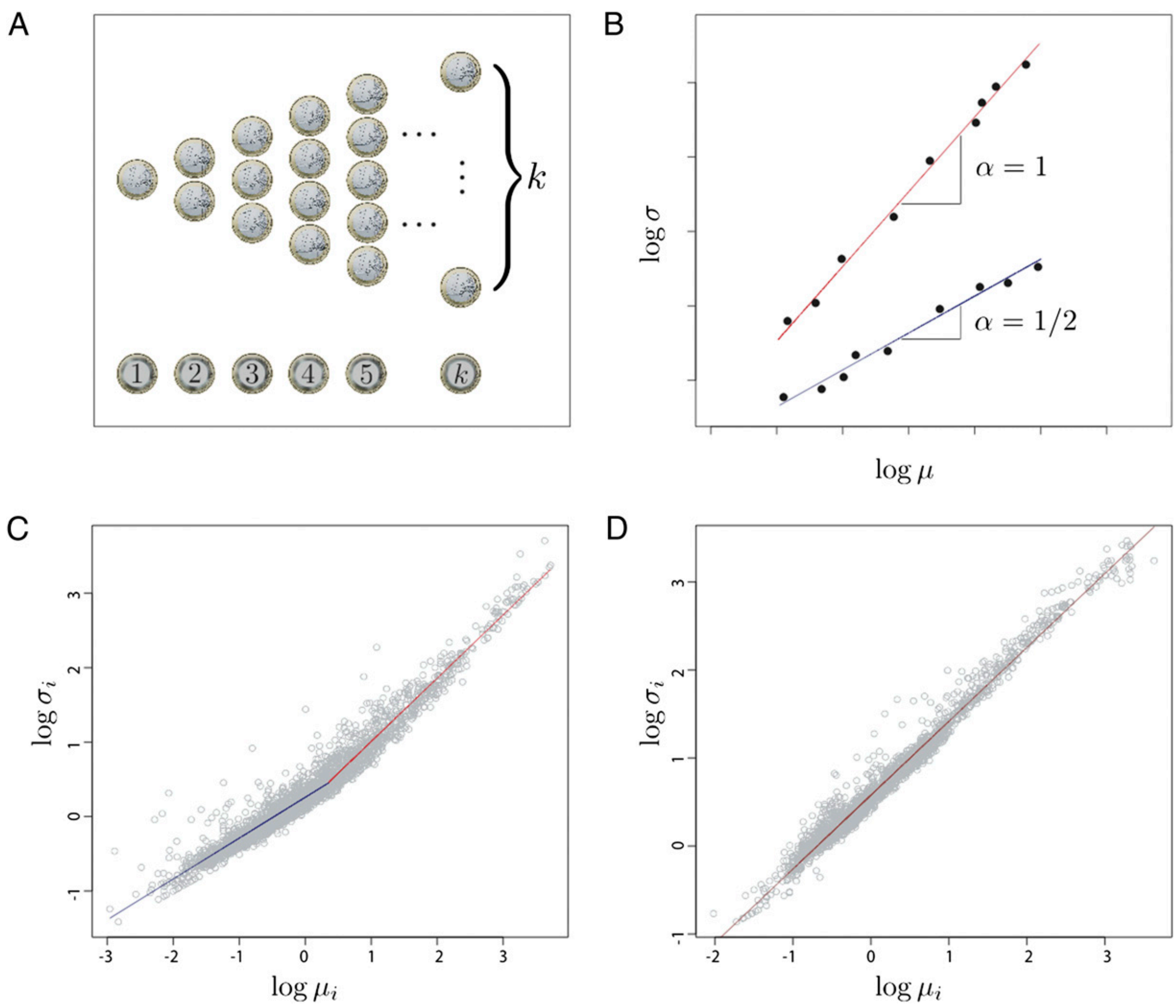

Fig. 2. Fluctuation scaling. (*A*) The concept of FS can be illustrated by considering tossing coins in two ways (32). (*i*) We toss a group of k coins independently with sides corresponding to 0 and 1 and let f_k equal their sum. (*ii*) We toss a single coin with sides 0 and k, which corresponds to tossing k fully coupled coins. (*B*) We perform the experiment several times and calculate the average $\langle f_k \rangle$ and SD σ_k of f_k as shown in the schematic. In both cases $\langle f_k \rangle \sim k$, whereas $\sigma_k \sim \sqrt{k}$ in *i* but $\sigma_k \sim k$ in *ii*. Varying the value of k produces a series of points in the $\log \mu_k$, $\log \sigma_k$ plane. From the FS point of view, this simple example resembles Facebook users making decisions on application adoption; the "coins" are now biased, reflecting individual heterogeneity, and the tosses are not independent but coupled via local and global signals (*SI Text*). (*C*) Of the 2,705 Facebook applications in the empirical dataset, 2,562 with $\mu_i > 0$ and $\sigma_i > 0$ are plotted here (*SI Text*). Two qualitatively different regimes emerge, and are separated by a cross-over point located at $\log \mu_x = 0.36$. The first, *individual regime* is characterized by the exponent $\alpha_I \approx 0.55$, and the second, *collective regime* by $\alpha_C \approx 0.85$. (*D*) The synthetic dataset consists of 2,705 time series, of which 2,163 have $\mu_i > 0$ and $\sigma_i > 0$. We now obtain a single regime characterized by the exponent $\alpha_S \approx 0.84$. Note that in *C* and *D* the exponents lie between 1/2 and 1, corresponding to the extremes of completely uncorrelated and correlated decisions of users to adopt applications.

time interval. We repeat the scaling plot by choosing randomly 300 applications from each group, with the red, green, and blue colors indicating whether the application was introduced in the first, second, or third interval (Fig. 3). Because any interval of x values contains an approximately equal number of markers of different colors, the time of introduction and, hence, application lifetime do not explain its scaling properties.

It is also possible that network externalities are present for some applications, meaning that the utility of having a particular application increases with its user base. We identified 402 applications with $\log(\mu) < -1$ and 495 applications with $\log(\mu) > 1$ excluding applications close to the transition. From each subset, we chose 50 applications at random and classified them manually based on whether significant network externalities were present or not, where they were deemed significant if the application was used for repeated social interactions with friends. For example, an application enabling one to play poker against friends clearly has network externalities, whereas an application that places a virtual lava lamp on the user's profile does not. Only 10% of the sampled applications with $\log(\mu) < -1$ had significant network externalities associated with them, whereas 28% of the sampled applications with $\log(\mu) > 1$ did. Because we find both types of applications in both regimes, network externalities alone cannot be the distinguishing factor, and in general are present less frequently than one might expect.

Should the transition from one regime to the other be attributed to the popularity of applications reaching a certain threshold value, or should it be attributed to the structural properties of the system and the dynamic behaviors it sustains? If the former is true, then one might think that the transition corresponds to a phase transition or to the crossing of an epidemic threshold, essentially a density threshold effect, resulting in an epidemic of popularity. To isolate the effects of popularity, we construct rank-order-preserving *synthetic time series* from the empirical ones. This deterministic process (apart from ties) cuts the empirical time series into pieces and then recombines the pieces using a rank-based rule

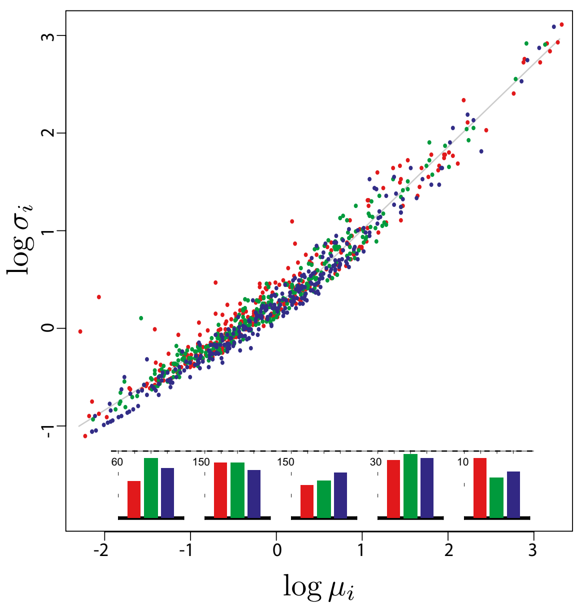

Fig. 3. Effect of application lifetime on scaling. Visual inspection shows that any interval of log-μ values contains a roughly equal number of red, green, and blue markers, indicating that the time of introduction and, hence, application lifetime are not related to its scaling properties. The histograms at the bottom of the panel show exactly how many applications from each of the three periods (red, green, blue) fall in the [−2, −1), [−1,0), [0, 1), [1, 2), and [2, 3) intervals, demonstrating clearly that there is no age trend in the scaling plot.

(*Materials and Methods*). As shown in Fig. 2*D*, the transition disappears for the synthetic data. Statistical tests also support the existence of a single regime (*SI Text*) and, in addition, the Pearson's linear correlation coefficient between log σ and log μ is 0.99. The consequences of this are threefold. First, the lack of two regimes for the synthetic data demonstrates that the transition from one regime to the other for the empirical data cannot be attributed solely to the popularity of an application exceeding a certain threshold. Hence, the phenomenon is not analogous to crossing an epidemic threshold. Second, it demonstrates that the collective (correlated) regime does not result from the system becoming saturated with users of a given application that would then induce correlations between the behaviors of the individuals. This is because all of the synthetic time series obey the same scaling relation also for small values of log(μ) (corresponding to a dilute limit), where the system is far from being saturated. Third, the synthetic regime has an exponent $\alpha_C \approx 0.84$, which is very close to $\alpha_C \approx 0.85$ that characterizes the collective regime for empirical data. This shows that we can recover the exponent of the collective regime by assuming that the future popularity of an application is driven by its current popularity, a finding that has also been used to predict popularity of online content (33).

Discussion

We have harnessed data on Facebook applications to study the role of social influence on the dynamics of popularity in an endogenous online system. The way the platform, Facebook, and the cultural products, Facebook applications, have been set up in this self-contained system guarantees that the agents are subject to both local and global signals of influence. Although our analysis cannot separate the contribution of local and global signals to the resulting behavior, it is nevertheless useful to characterize the fundamental structure of the information that individuals can access because this enables comparisons with other systems (34). We have shown that the studied online system exhibits a collective and individual regime, and argued that the emergence of the two regimes is an inherent property of the system. Because each regime is characterized by a single fluctuation scaling exponent, the strength of social influence is approximately constant across each regime. Consequently, the extent of social influence becomes discretized: Either there is virtually no influence or, alternatively, the strength of influence is that given by the exponent of the collective regime. This suggests that social influence assumes a binary, on–off nature in the system. Had we only monitored the more successful (high-μ) applications, as one is usually constrained to do in the offline world, we would have been able to observe only (part of) the collective regime. However, it is unclear what would happen in the offline world if equivalent experimental conditions could be replicated.

Our ability to identify the two distinct regimes exhibited by the system does not, however, allow us to infer or rule out specific microlevel social mechanisms. This would require us to analyze individual-level data, rather than aggregate time-series data where distinct Facebook applications are the units of analysis. There is strong empirical evidence for diffusion in social systems (35–37), but the precise underlying mechanism can vary from case to case. Recent work by Young (26) has made an important theoretical contribution by showing how the shape of cumulative adoption curves can be used to differentiate between social contagion, social influence, and social learning processes. However, to keep the mathematical treatment tractable, Young's approach assumes that there is perfect mixing of past and potential future adopters, and hence no underlying social network. This clearly does not apply in our case, as well as in many other offline and online social systems. Further theoretical and empirical advances that take into account network characteristics are required before distinct behaviors at the aggregate level can be mapped into different processes at the individual level.

Web-based interactive systems have the potential to transform our understanding of collective human behavior (38). We believe that our finding on the existence of the two regimes may well generalize to other online systems. The move of an increasing number of our activities to the online world has endowed users with the power of participation. Familiar examples include the online book retailer Amazon and the online DVD rental service Netflix, both of which allow their users to rate products and, consequently, influence their future popularity. Although some books and films in these systems are certainly highly advertised by their producers, they arguably stand for only a small fraction

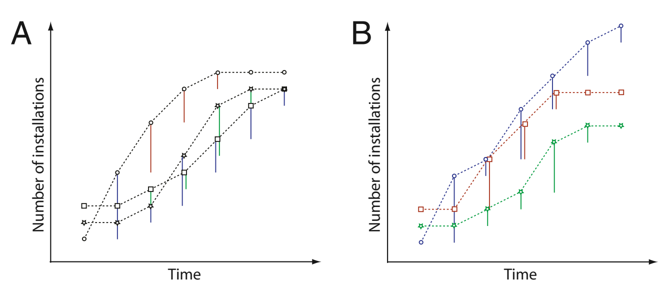

Fig. 4. Schematic of the construction of the synthetic time series $\tilde{n}_i(t)$. (*A*) The empirical data consist of $t = 1,\ldots,7$ observations for three applications. The data points have been connected with dashed black lines to guide the eye. For the most popular application at time $t-1$, the change in the number of users between $t-1$ and t is indicated by the height of the vertical red bar at time t, which corresponds to $\tilde{f}_1(t)$ in the text. Similarly, $\tilde{f}_2(t)$ and $\tilde{f}_3(t)$ are indicated by the green and blue bars, respectively. An easy way to understand the process is first to compute the difference in the number of users for all applications given by $f_i(t) = n_i(t) - n_i(t-1)$ and then color the difference based on $r_i(t-1)$, the rank of the application at time $t-1$. (*B*) The synthetic time series are seeded by the initial values taken from the empirical data such that $\tilde{n}_1(1) = n_{\square}(1), \tilde{n}_2(1) = n_{\star}(1)$, and $\tilde{n}_3(1) = n_{\circ}(1)$ of the empirical data and they are constructed by adding together the difference bars of the same color. Overlapping bars have been shifted slightly horizontally for clarity of presentation.

of the choices available, leaving a large majority of books and films exposed to endogenously generated social influence. Social influence may then emerge spontaneously in a wide range of online environments over and above purely endogenous systems. Whether it becomes discretized in these systems as well remains to be seen.

Materials and Methods

Synthetic Time Series. We construct rank-order-preserving synthetic time series from the empirical time series to isolate the effects of popularity from other factors in the log σ, log μ plots. This process is deterministic (apart from ties), and essentially it cuts the empirical time series into pieces and then recombines the pieces using a rank-based rule (Fig. 4). Let us denote the global ranking of application k at time t with $r_k(t) \in 1,\ldots,M$ such that $n_{(k-1)}(t) \geq n_{(k)}(t) \geq n_{(k+1)}(t)$. We define $\tilde{n}_i(t) = \tilde{n}_i(t-1) + \tilde{f}_i(t)$ analogously to what we had before, but now $\tilde{f}_i(t) = n_k(t) - n_k(t-1)$ such that $r_k(t-1) = i$. Here $\tilde{f}_i(t)$ is the number of new installations over a single time step for an application that at time $t - 1$ had ranking i. The series are seeded by setting $\tilde{n}_i(1) = n_{(i)}(1)$ for all $i = 1,\ldots,M$ and are constructed using the above recursive relation for $t > 1$.

The synthetic time series $\tilde{n}_i(t)$ by construction has a constant relative popularity as measured by the global rank order of $\tilde{n}_i(t)$ and, consequently, the future popularity of the synthetic time series is systematically driven only by its current popularity (rank). In the absence of rank crossings of applications, the synthetic data would behave like empirical data. The increments $\tilde{f}_i(t)$ of the synthetic data result from a combined effect of both local and global signals. The impact of the global signal remains constant (in terms of rank) because the synthetic time series $\tilde{n}_i(t)$ always holds rank i on the global best-seller list. A single synthetic time series $\tilde{n}_i(t)$ is typically a combination of several empirical time series and, therefore, the local signal in the synthetic time series corresponds to a mean-field approximation of the local signals of the applications that make up the synthetic time series $\tilde{n}_i(t)$.

ACKNOWLEDGMENTS. We thank S. Gourley for help with data acquisition; D. Barron, N. A. Christakis, J. Denrell, Z. Eisler, J. Fowler, P. Hedström, N. Jones, G. Kossinets, E. Lopez, A. Motter, M. A. Porter, S. Saavedra, M. Salganik, G. Szabo, and B. Uzzi for their comments on earlier versions of the manuscript; and the workshop participants at the University of North Carolina at Chapel Hill, the University of Oxford, Northwestern, Northeastern, and Harvard University. J.-P.O. is supported by Wolfson College, University of Oxford, the Fulbright Program, and acknowledges the support of grant P01 AG-031093 from the National Institute on Aging. F.R.-T. acknowledges the support of the UK Engineering and Physical Sciences Research Council Grant EP/E056997/1 and the European Commission FP7 Future and Emerging Technologies Open Scheme Project ICTeCollective (Contract 238597).

1. Mason WA, Conrey FR, Smith ER (2007) Situating social influence processes: Dynamic, multidirectional flows of influence within social networks. *Pers Soc Psychol Rev* 11:279–300.
2. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83:1420–1443.
3. Avery C, Zemsky P (1998) Multidimensional uncertainty and herd behavior in financial markets. *Am Econ Rev* 88:724–748.
4. Shiller RJ (2000) *Irrational Exuberance* (Princeton Univ Press, Princeton, NJ).
5. Lazarsfeld PF, Berelson B, Gaudet H (1994) *The People's Choice* (Columbia Univ Press, New York).
6. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357:370–379.
7. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311:854–856.
8. Katz E, Lazarsfeld PF (1955) *Personal Influence* (Free Press, New York).
9. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge Univ Press, Cambridge, UK).
10. Coleman J, Katz E, Menzel H (1957) The diffusion of an innovation among physicians. *Sociometry* 20:253–270.
11. Griliches Z (1957) Hybrid corn: An exploration in the economics of technological change. *Econometrica* 25:501–522.
12. Rogers EM (2003) *Diffusion of Innovations* (Free Press, New York).
13. Valente TW (1995) *Network Models of the Diffusion of Innovations* (Hampton, Cresskill, NJ).
14. Young HP (2005) *The Economy as a Complex Evolving System, III*, eds Blume LE, Durlauf SN (Oxford Univ Press, New York).
15. Dodds PS, Watts DJ (2004) Universal behavior in a generalized model of contagion. *Phys Rev Lett* 92:218701.
16. Goel S, Mason W, Watts DJ (2010) Real and perceived attitude agreement in social networks. *J Pers Soc Psychol*, 10.1037/a0020697.
17. Mayer A, Puller SL (2008) The old boy (and girl) network: Social network formation on university campuses. *J Public Econ* 92:329–347.
18. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis NA (2008) Tastes, ties, and time: A new (cultural, multiplex, and longitudinal) social network dataset using Facebook.com. *Soc Networks* 30:330–342.
19. Golder S, Wilkinson DM, Huberman BA Rhythms of social interaction: Messaging within a massive online network. arXiv.org/abs/cs/0611137.
20. Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proc Natl Acad Sci USA* 105:15649–15653.
21. Traud AL, Kelsic ED, Mucha PJ Porter MA Community structure in online collegiate social networks. arXiv:0809.0690v1.
22. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2009) Scaling laws of human interaction activity. *Proc Natl Acad Sci USA* 106:12640–12645.
23. Onnela J-P, et al. (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA* 104:7332–7336.
24. Lazer D, et al. (2009) Computational social science. *Science* 323:721–723.
25. Boyd DM, Ellison NB (2008) Social network sites: Definition, history, and scholarship. *J Comput Mediat Commun* 13:210–230.
26. Young HP (2009) Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *Am Econ Rev* 99:1899–1924.
27. Van den Bulte C, Lilien GL (2001) Medical innovation revisited: Social contagion versus marketing effort. *Am J Sociol* 106:1409–1435.
28. Bass FM (1969) A new product growth model for consumer durables. *Manage Sci* 15:215–227.
29. Denrell J, Kovacs B (2008) Selective sampling of empirical settings in organizational studies. *Adm Sci Q* 53:109–144.
30. Smith HF (1938) An empirical law describing heterogeneity in the yields of agricultural crops. *J Agric Sci* 28:1–23.
31. Taylor L (1961) Aggregation, variance, and the mean. *Nature* 189:732–735.
32. Eisler Z, Bartos I, Kertesz J (2008) Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv Phys* 57:89–142.
33. Szabo G, Hubermann BA (2010) Predicting the popularity of online content. *Commun ACM* 53:80–88.
34. López-Pintado D, Watts DJ (2008) Social influence, binary decisions and collective dynamics. *Rationality Soc* 20:399–443.
35. Strang D, Soule SA (1998) Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annu Rev Sociol* 24:265–290.
36. Hedström P (1994) Contagious collectives: On the spatial diffusion of Swedish trade unions, 1890–1940. *Am J Sociol* 99:1157–1179.
37. Biggs M (2005) Strikes as forest fires: Chicago and Paris in the late nineteenth century. *Am J Sociol* 110:1684–1714.
38. Watts DJ (2007) A twenty-first century science. *Nature* 445:489.