



Research

Cite this article: Karsai M, Iñiguez G, Kaski K, Kertész J. 2014 Complex contagion process in spreading of online innovation. *J. R. Soc. Interface* **11**: 20140694.
<http://dx.doi.org/10.1098/rsif.2014.0694>

Received: 30 June 2014

Accepted: 30 September 2014

Subject Areas:

mathematical physics, environmental science, biocomplexity

Keywords:

complex contagion phenomena, mean-field approximation, data-driven modelling

Author for correspondence:

Márton Karsai
e-mail: marton.karsai@ens-lyon.fr

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2014.0694> or via <http://rsif.royalsocietypublishing.org>.

Complex contagion process in spreading of online innovation

Márton Karsai^{1,2,3,4}, Gerardo Iñiguez², Kimmo Kaski^{2,5} and János Kertész^{2,6,7}

¹Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA 02115, USA

²Department of Biomedical Engineering and Computational Science (BECS), Aalto University School of Science, 00076 Aalto, Finland

³Software Technology and Applications Competence Center (STACC), 51003 Tartu, Estonia

⁴Laboratoire de l'Informatique du Parallélisme, INRIA-UMR 5668, IXXI, ENS de Lyon, 69364 Lyon, France

⁵CABDyN Complexity Centre, Säid Business School, University of Oxford, Oxford OX1 1HP, UK

⁶Center for Network Science, Central European University, 1051 Budapest, Hungary

⁷Institute of Physics, Budapest University of Technology and Economics, 1111 Budapest, Hungary

Diffusion of innovation can be interpreted as a social spreading phenomenon governed by the impact of media and social interactions. Although these mechanisms have been identified by quantitative theories, their role and relative importance are not entirely understood, as empirical verification has so far been hindered by the lack of appropriate data. Here we analyse a dataset recording the spreading dynamics of the world's largest Voice over Internet Protocol service to empirically support the assumptions behind models of social contagion. We show that the rate of spontaneous service adoption is constant, the probability of adoption via social influence is linearly proportional to the fraction of adopting neighbours, and the rate of service termination is time-invariant and independent of the behaviour of peers. By implementing the detected diffusion mechanisms into a dynamical agent-based model, we are able to emulate the adoption dynamics of the service in several countries worldwide. This approach enables us to make medium-term predictions of service adoption and disclose dependencies between the dynamics of innovation spreading and the socio-economic development of a country.

1. Introduction

Diffusion of news, ideas and innovations as well as the distribution of services and products are all examples of social spreading phenomena that have become an integral part of our everyday life, strongly accelerated by novel, Web-based interaction channels. These innovations serve as an engine of economic development [1], but only their diffusion throughout society brings them to success. The processes involved in innovation spreading have been the focus of research for decades [1–4], yet their dynamics and modelling have remained as challenges to our scientific understanding.

The propagation of innovations takes place in a social network [2–5] and is driven by the entanglement of individuals' decision-making processes [6] as well as by the influence of media and social interactions [7,8]. Although the effects of network structure on contagion processes have recently been shown to be important [9], knowledge about the social network itself is rather limited as its structure and dynamics usually remain hidden. In this respect, the digital age has opened up unprecedented opportunities, as online social networks and Voice over Internet Protocol services record detailed information of the connections and activities of their users. These services partially decode the underlying social structure by acting as proxies for the network of real social ties between individuals, and also provide accurate records of the users' adoption behaviour. In this way, the different sources of influence on the decisions of an individual immersed in a perpetually changing environment of social interactions become traceable. We are therefore encouraged to devise dynamic agent-based models

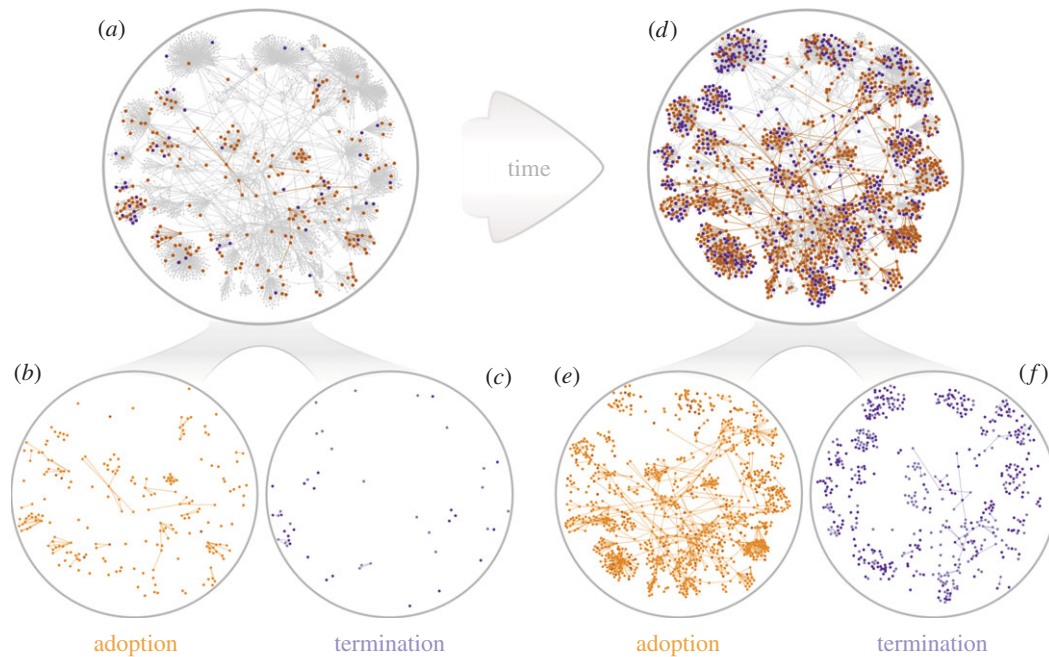


Figure 1. Sample of the aggregated Skype network of Switzerland. (a,d) Snowball sample (maximum distance from seed $d = 4$) where user accounts and confirmed links between them are shown for two intermediate times in the adoption process. Nodes are coloured according to their adoption state: grey for future users, orange for current adopters and purple for terminated accounts (corresponding to states S , A and R in our compartmental model). A link has one of these colours if the states of the connected nodes are equal and is grey otherwise. (b,c,e,f) Decomposed networks for adoption (b,e) and termination (c,f). Nodes have the same colours as before but are shaded depending on their action (dark for spontaneous action and light for peer-pressure action). Only links connecting nodes with equal states are shown. The termination network has a larger fraction of dark nodes than its adoption counterpart, meaning that social interactions affect adoption more than termination.

to describe, simulate and even predict emergent behaviour of such social contagion phenomena [10–12].

These phenomena are identified as *complex contagion processes* when the exposure of an individual is conditional on the decision of a fraction of its peers [13]. This is particularly different from *simple spreading processes*, where a rate determines the transmission of infection between nodes and one infected neighbour is always sufficient to expose a susceptible node [9,14]. Complex contagion phenomena are commonly modelled by processes where the fractions of adopting neighbours necessary for exposure are set as individual thresholds. This idea was first introduced by Granovetter [15] who discussed the ideal network structure and threshold distribution to allow for the evolution of riots or other collective movements. Subsequently, Watts [16] proposed a simplistic model to explore sufficient structural and threshold conditions for the evolution of global adoption cascades. During the last 10 years, several studies contributed to the foundations of complex contagion [16–21], and in addition online experiments were carried out to provide empirical evidence about the effect of social influence [22,23]. Beyond the conventional threshold mechanism, the effect of homophily [5,20,24] and the role of external media influence [3] were also investigated recently.

Here, we study one of today's largest online communication services, the Skype network, with over 300 million monthly connected users [25]. Data cover the history of individuals that adopted Skype from September 2003 until March 2011 (i.e. 2738 days), including registration events and contact network evolution for every registered user around the world. For our investigation, we select user accounts with an identified country of registration and consider only their mutually confirmed connections, both within the country and abroad. To receive the best estimation of node degrees

in the underlying social network, we integrate the evolving Skype network for the whole available period and count the number of confirmed relationships per node (including international ties). The adoption dynamics of a given country can be directly observed by assigning times of adoption (t_a) and termination (t_t) to all the accounts. These are, respectively, defined as the dates of registration and last activity (as regards to any of the services) in Skype. Explicitly, we identify any account as terminated if its last activity happened earlier than 1 year prior to the end of the observation period. In this way, we are able to build a complete adoption and termination history of Skype for 2373 days. As an illustration of the adoption process, in figure 1a,d we show a sample of the contact network of Switzerland for two intermediate times (for further details of the dataset, see the electronic supplementary material, S51).

Taking advantage of this large digital dataset, our goal is to fill the persisting gap between real observations and the assumptions made in models of product adoption spreading in techno-social networks. We empirically study the assumptions borrowed from conventional models of complex contagion and analyse the crucial effect of social influence. Finally, we introduce an agent-based model that combines the detected diffusion mechanisms and provides plausible medium-term predictions for the spreading of online innovations in several countries worldwide.

2. Results

2.1. The adoption dynamics

The spreading of the online service is determined by competing processes of adoption and termination as described by

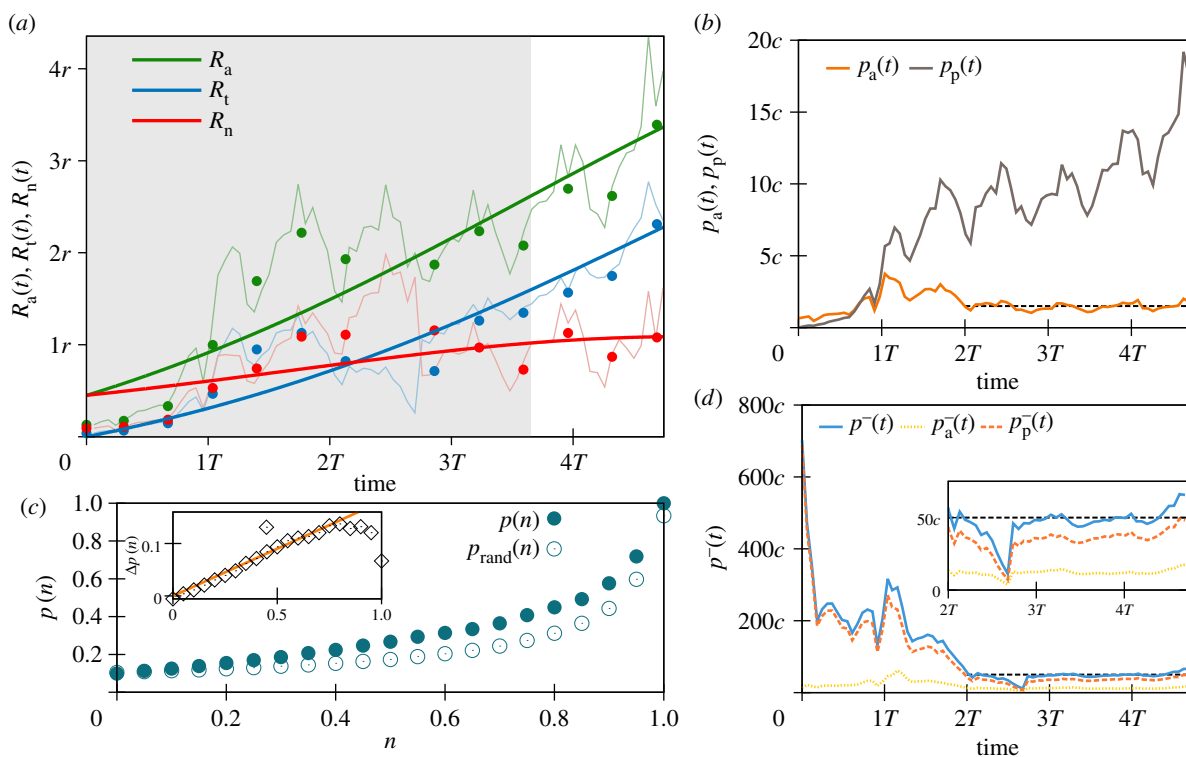


Figure 2. Empirical rates and probabilities for Switzerland. (a) Thin curves denote empirical rates of adoption [$R_a(t)$], termination [$R_t(t)$] and net adoption [$R_n(t)$], while symbols are their corresponding binned values. A binned data point in $[2T, 3T]$ has been removed due to systematic bias in $R_t(t)$ caused by a major software update during this period. A shaded (white) area indicates the training (predicted) period for the theoretical fit of our model, drawn as thick lines with the same colours as the empirical rates. (b) Probabilities of spontaneous [$p_a(t)$] and peer-pressure [$p_p(t)$] adoption per unit time. (c) Average conditional probability of adoption as a function of the fraction of adopting neighbours n , measured in the original data [$p(n)$, solid circles] and in the shuffled data corresponding to the null model [$p_{\text{rand}}(n)$, open circles]. Inset shows the unbiased difference $\Delta p(n) = p(n) - p_{\text{rand}}(n)$ (symbols) and a fitted linear function (continuous line). (d) Probabilities of overall termination [$p^-(t)$], and of spontaneous [$p_a^-(t)$] and peer-pressure [$p_p^-(t)$] termination per unit time. The inset depicts a zoom from time $2T$ onwards. T , r and c are arbitrary linear scaling constants, with time dimensions for T . Black lines in panels (b,d) are fitted constants.

the evolution of the corresponding rates $R_a(t)$ and $R_t(t)$, which measure the fraction of all users that adopt or terminate the service in a given time window Δt (figure 2a). These simple rate functions already disclose interesting features of the adoption dynamics, as their overall growth signals continuously accelerating processes of adoption and termination. Yet the actual time evolution of spreading service is better characterized by the net adoption rate $R_n(t) = R_a(t) - R_t(t)$ (for an overview of all empirical quantities, see table 1).

Opening a user account constitutes a single event in the decision-making process of an individual that is triggered by spontaneous decisions, by the influence of media or by the social environment [5,16]. On the other hand, users may terminate their accounts for several reasons including vanishing demand or dissatisfaction, by switching to another product permanently or by simply abandoning the service with a chance of re-adoption (e.g. due to loss of password or intention for lower monitoring). Some of these processes are observable by investigating the data. An example is shown in figure 1, where the contact network of Switzerland is further decomposed into sub-networks of adopted and terminated users. In the former, some nodes appear disconnected, which indicates individuals that have adopted Skype prior to their friends. This so-called *spontaneous* adoption, where individual factors and external media play a role, is a typical adoption pattern in the beginning of the process (figure 1b). Alternatively, at the time of adoption many nodes have neighbours who are already existing users, a common pattern for

later times (figure 1e). This second scenario of *peer-pressure* adoption indicates the possible influence of the social environment. By contrast, the termination network consists mostly of single nodes at all times (figure 1c,f), meaning that these users, although they are surrounded by adopters, decide individually to terminate. This observation suggests a negligible effect of social influence on termination.

2.2. Mechanisms of adoption

An analysis of the evolving network structure around a given user can help us to detect whether an ego adopted or terminated the product before any of its neighbours did; or else followed the decisions previously made by a fraction of them. In this way, we can label the performed action as either spontaneous or driven by peer pressure. To define the related measures, we consider the underlying social network as static, meaning that its evolution requires a much larger temporal scale than the adoption process itself. This static structure is defined as the aggregated social network of Skype at the end of the recorded period and provides a lower estimate for the total number of friends of each individual. Moreover, we assume that the maximum size of the static social network is the number I of Internet users in a given country at the end of the observation period [26], and thus define $I - N_a(t)$ as the population that has not yet adopted Skype at time t .

Under these assumptions, the probabilities per unit time that a user adopts either spontaneously or due to peer

Table 1. Data and model quantities, including notation and a brief description of their meaning. Groups correspond to: quantities measured in both the data and model; empirical quantities; parameters used by the model; and parameters estimated from the data.

quantity	description
data/model quantities	
$R_a(t), R_t(t)$	fraction of all users that adopt/terminate the service in a time window
$R_n(t) = R_a(t) - R_t(t)$	net rate of adoption
empirical quantities	
t_a, t_t	time of adoption/termination of service for a user
$t_l = t_t - t_a$	lifetime of a user account in the service
Δt	time window used to define rates
I	number of Internet users in a country at the end of observation period
$N_a(t)$	number of service users at time t
$p_a(t), p_p(t)$	rate of spontaneous/peer-pressure adoption at time t
$p_a^-(t), p_p^-(t)$	rate of spontaneous/peer-pressure termination at time t
$p^-(t) = p_a^-(t) + p_p^-(t)$	rate of termination at time t
$p(n)$	probability that a user adopts, if a fraction n of its neighbours are adopters
$p_{\text{rand}}(n)$	value of $p(n)$ after shuffling adoption times
$\Delta p(n) = p(n) - p_{\text{rand}}(n)$	unbiased measure of social influence in adoption
τ	inverse speed of innovation diffusion
model parameters	
p_a, p_p	constant probability of spontaneous/peer-pressure adoption
p_s, p_r	constant probability of temporary/permanent termination
$\langle k \rangle$	average degree of network
$p_{pk} = p_p(\langle k \rangle - 1) / \langle k \rangle$	peer-pressure adoption probability, weighted by degree
estimated parameters	
$\widetilde{\langle k \rangle}, \widetilde{p}^-$	estimates of average degree and asymptotic rate of termination
$\widetilde{p}_s = (\widetilde{p}^- - p_r) / (1 - p_r)$	estimate of temporary termination probability

pressure are defined as

$$\text{and } \left. \begin{aligned} p_a(t) &= \frac{\#\text{ad}(t + \Delta t | \text{SF} = 0)}{I - N_a(t)} \\ p_p(t) &= \frac{\#\text{ad}(t + \Delta t | \text{SF} \neq 0)}{I - N_a(t)} \end{aligned} \right\} \quad (2.1)$$

where $\#\text{ad}(t + \Delta t | \text{SF} = 0)$ [$\#\text{ad}(t + \Delta t | \text{SF} \neq 0)$] is the number of users who adopt the service in a time window Δt , under the condition that their number of adopting neighbours at time t is $\text{SF} = 0$ ($\text{SF} \neq 0$). In a similar fashion, the probabilities per unit time that a user terminates the service either spontaneously or due to peer pressure are

$$\text{and } \left. \begin{aligned} p_a^-(t) &= \frac{\#\text{tr}(t + \Delta t, \text{TF} = 0)}{N_a(t)} \\ p_p^-(t) &= \frac{\#\text{tr}(t + \Delta t, \text{TF} \neq 0)}{N_a(t)} \end{aligned} \right\} \quad (2.2)$$

where TF stands for the number of neighbours of a user that have terminated usage up to time t (for a discussion on the restrictions of these empirical quantities, see the electronic supplementary material, §§S2.1–S2.4).

The data show that after an initial, transient period, the rate of spontaneous adoption $p_a(t)$ (figure 2b) and the rate of termination $p^-(t) = p_a^-(t) + p_p^-(t)$ (figure 2d) become constant apart from small fluctuations. The same holds separately for

the rates of spontaneous [$p_a^-(t)$] and peer-pressure [$p_p^-(t)$] termination. The time invariance of these rates is an obvious assumption for most biological epidemics, which, however, has never been empirically shown before in the case of social contagion phenomena, despite its wide use [27,28]. Our results provide the first validation of this quite fundamental assumption used in the conventional modelling of social spreading processes, where probabilities analogous to the ones described here are treated like constants at the outset.

When the ego is not the first adopter among neighbours, the rate $p_p(t)$ of adoption via peer pressure is not constant but increases with time (figure 2b). This is mainly due to social influence arising from the user's social circle. An appropriate way to quantify such effects is to measure the conditional probability $p(n)$ of adoption provided that a fraction n of the ego's neighbours have adopted the product before as

$$p(n) = \frac{\#\text{ad}(n)}{N - \sum_{m=0}^{m < n} \#\text{ad}(m)}. \quad (2.3)$$

Here, the numerator counts the number of users with a fraction n of adopter friends at the time of adoption, while the denominator is the number of people with a larger or equal fraction $m \geq n$, i.e. all individuals who had the chance to adopt Skype while having a fraction n of adopter neighbours (for further details, see the electronic supplementary

material, §S2.3). We observe that the probability $p(n)$ is monotonically increasing (figure 2c), an empirical finding in agreement with the assumptions of several threshold models for epidemic spreading and social dynamics [16,29–31]. However, as we cannot see the entire social network (only the part uncovered by the Skype graph), this probability is biased as $n \rightarrow 1$. To estimate such bias, we build a reference null model by shuffling the adoption times of all accounts and measuring the corresponding conditional probability $p_{\text{rand}}(n)$ for this system. The shuffling procedure removes the effect of social influence but conserves the adoption rates and keeps the social structure unchanged. In other words, the reference probability is biased in the same way as the original measurement, but is not driven by social influence as all such correlations have been removed by the shuffling. Consequently, the difference $\Delta p(n) = p(n) - p_{\text{rand}}(n)$ quantifies the effect of social influence in the adoption process (inset of figure 2c): $\Delta p(n)$ increases approximately in a linear fashion with the fraction of adopting neighbours. This observation is in agreement with previous studies where a similar scaling of social influence has been recognized through small scale experiments [22], data-driven observations [20] and modelling [29,32].

2.3. The model process

The analogy between epidemic spreading and social contagion has been widely used to model various societal diffusion processes [11,14,33,34]. Here, we take this approach to build a compartmental model based on the identified mechanisms in Skype usage, aimed at a generic description of the large-scale adoption dynamics of technological innovations. We depict individuals as agents in one of three non-overlapping states, susceptible (S), adopter (A) and removed (R), describing people who may adopt the product later, are users already, and will never use it again. In accordance with our observations, the behaviour of an agent can be characterized by four elementary processes. (a) *Spontaneous adoption*, influenced by individual factors or external media independently of the social network. This is certainly the dominant mechanism for agents with no user neighbours at the time of adoption. (b) *Peer-pressure adoption*, an intrinsic social effect implemented here by making use of the observed linear scaling of the probability $p(n)$. (c) *Temporary termination*, describing the case in which agents stop usage with a chance of re-adoption. (d) *Permanent termination*, when users abandon the service altogether. The flow $S \rightarrow A$ is regulated by processes (a) and (b), $A \rightarrow S$ by (c), and $A \rightarrow R$ by (d). Finally, we assume that the underlying social network evolves with a much longer time scale than the ongoing adoption process, so that its structure may be considered static with fixed size.

For large systems, the modelled adoption process can be well characterized by a rate equation formalism using the heterogeneous mean-field approximation [9,35] (see appendix A). This approach takes agents with identical degree to be statistically equivalent and ignores fluctuations in their dynamical properties. Thus, assuming no degree–degree correlations in the network, the adoption dynamics is reduced to the following system of nonlinear ordinary differential equations:

$$\frac{da}{dt} = [p_a + p_{pk}(1 - p_a)a]s - [p_r + p_s(1 - p_r)]a, \quad (2.4)$$

$$\frac{ds}{dt} = -[p_a + p_{pk}(1 - p_a)a]s + p_s(1 - p_r)a \quad (2.5)$$

$$\text{and} \quad \frac{dr}{dt} = p_r a, \quad (2.6)$$

where $s(t)$, $a(t)$ and $r(t)$ are the average probabilities that an agent is in state S , A or R , respectively, and satisfy the normalization condition $s(t) + a(t) + r(t) = 1$. The elementary mechanisms (a–d) are parametrized through the constant probabilities of spontaneous (p_a) and peer-pressure (p_p) adoption, and of temporary (p_s) and permanent (p_r) termination (for an overview of all model parameters, see table 1). Under the above conditions, the model does not depend on the degree distribution of the social network, as p_p appears only in the weighted form $p_{pk} = p_p(\langle k \rangle - 1)/\langle k \rangle$, with $\langle k \rangle$ the average degree of the network. Moreover, for large $\langle k \rangle$, the model becomes independent of this quantity as $p_{pk} \sim p_p$. The system (2.4)–(2.6) finally allows us to write the theoretical rates of adoption and termination as

$$R_a(t) = [p_a + p_{pk}(1 - p_a)a]s \quad (2.7)$$

and

$$R_t(t) = [p_r + p_s(1 - p_r)]a, \quad (2.8)$$

that is, the gain and loss terms in equation (2.4) (detailed derivation in the electronic supplementary material, §S3).

In order to measure the effect of degree–degree correlations on the spreading dynamics, we perform data-driven simulations by evaluating the model process over the integrated Skype network of a country (figure 3a). While this empirical network retains its full topological complexity (in terms of real community structure, assortativity, etc.), we consider a model scale-free network [36] of the same size and average degree as a control case. We then run the model process over the empirical and control networks and compare their corresponding rates of adoption and termination with the mean-field prediction of equations (2.7) and (2.8). As the average degree of the Skype network is not too small, deviations of the simulated rates from the theoretical values are not large, resulting in rates with the same qualitative behaviour. More interestingly, there is only a small discrepancy between the rates of the empirical and control networks. This suggests that topological correlations have a minor slowing-down effect on the spreading dynamics, and thus play a negligible role in the overall rates of the adoption process. Note that a similar independence of the population structure has been observed earlier in controlled experiments of networked public goods games [24] and a spatial Prisoner's Dilemma [37]. These observations validate retrospectively the theoretical considerations mentioned above, where we have assumed the background social network to be uncorrelated.

2.4. Validation and socio-economic correlations

In order to validate our model process, we compare the theoretical rate functions of equations (2.7) and (2.8) with the empirical data by estimating some of the model parameters. An estimate $\widetilde{\langle k \rangle}$ for the average degree of the social network can be obtained from the fully aggregated Skype network of a country, if we consider the ego-network of each user with international links included. The estimated rate of termination, $\widetilde{p_r}$, can be measured from the long-time behaviour of the spreading process (figure 2d) and then used to fix p_s to the value $\widetilde{p_s} = (\widetilde{p_r} - p_r)/(1 - p_r)$, where p_r is a free parameter. While p_a could also be measured directly (by counting adoption events where no user neighbours are

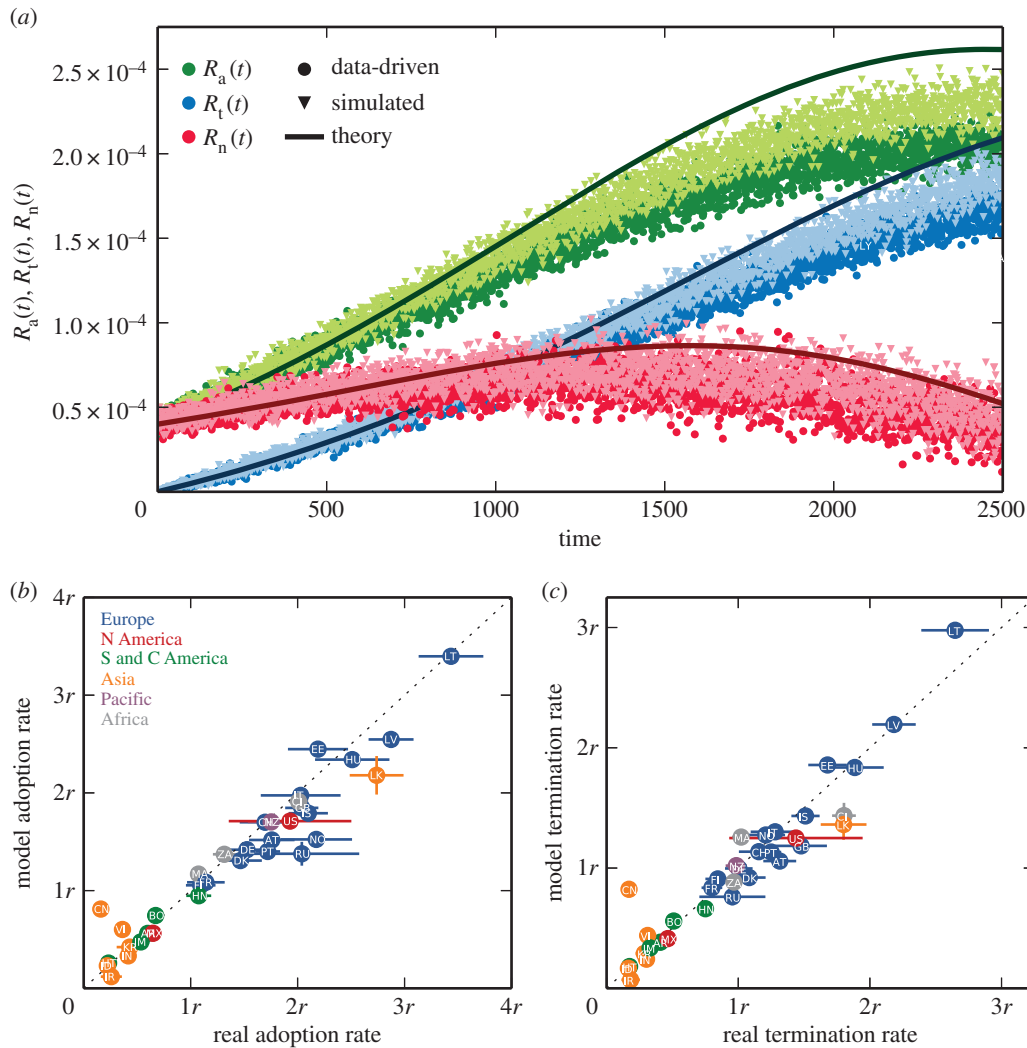


Figure 3. Data-driven simulations and validation of model predictions. (a) Rates of the model process over the integrated Skype network of Switzerland (dark symbols) and over an uncorrelated scale-free network (light symbols) with similar size and average degree. Parameter values for the numerical simulations are arbitrarily set to $p_a = 0.00004$, $p_p = 0.0024$, $p_s = 0.0004$ and $p_r = 0.0008$. Solid lines indicate the theoretical rates calculated from equations (2.7) and (2.8). Similar scaling of the data-driven, simulated and theoretical curves suggests a minor effect of structural correlations. (b,c) Comparison between empirical and theoretical values of the rates of adoption and termination for 34 different countries. Symbols depict rates averaged over the last six months of the observation period, with their corresponding standard deviations as error bars. In-symbol letters are country abbreviations, colours denote continental territories, r is an arbitrary linear scaling constant and the dashed line is a linear function with unit slope.

present at the time), the observation time of the dataset is not sufficiently long to estimate a constant value \tilde{p}_a in all countries. Therefore, we leave p_a , p_p and p_r as free quantities to be fitted (estimation of p_a for selected countries in the electronic supplementary material, S54).

Overall, the model dynamics is characterized by $\{p_a, p_p, p_r, \tilde{p}_s, \langle k \rangle\}$, a set of three free parameters and two estimated quantities. The free parameters are used to simultaneously fit the model rates on the binned empirical rates $R_a(t)$, $R_t(t)$ and $R_n(t)$, by means of a bounded nonlinear least-squares method. To ascertain the predictive power of our model, we fit over a 5-year training period and look for predictions in the last 1.5 years (figure 2a). Such prognosis can be quantified by comparing the average rates provided by the model with their corresponding empirical values during the final six months of observation. After repeating the calculations for 34 different countries (with diverse levels of technological development), the related values of the final empirical and modelled rates all collapse close to a line with unit slope (figure 3b,c), thus validating our model for the studied adoption process.

Our model may also be used to disclose relevant differences between the adoption dynamics of countries at various levels of societal and economical development. One characteristic indicator is the inverse speed of innovation diffusion, defined as the time τ when the theoretical $R_n(t)$ is maximal (see the electronic supplementary material, S53.4). If we relate τ with one of the standard measures of economical development, gross domestic product (GDP) *per capita* (the GDP dollar estimates used in our study are derived from purchasing power parity calculations using World Bank data, 2011; www.worldbank.org), large differences emerge between countries (figure 4a). Specifically, the larger the GDP of a country, the faster the adoption process is in its society. Another way to characterize the adoption dynamics is through the average account lifetime, $\langle t_i \rangle = \langle t_t - t_a \rangle$, where t_a and t_t are the corresponding registration and termination times. We relate this empirical measure to its theoretical analogue, the inverse probability of termination $1/\tilde{p}$ obtained from the fitted model process (figure 4b). Their correlation indicates that our model captures this dynamical property correctly. Moreover, the typical duration of user engagement uncovers clusters of countries at

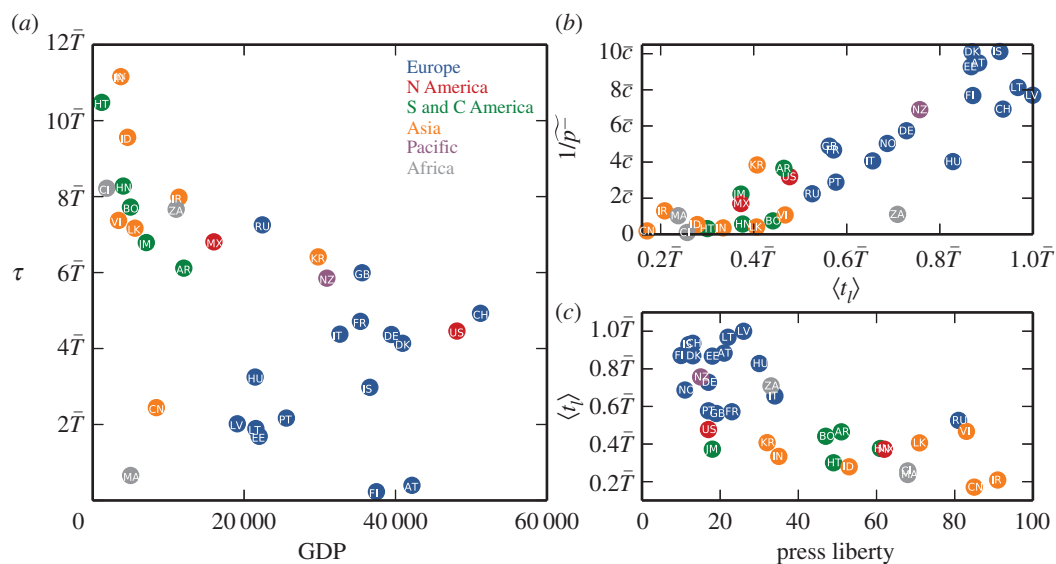


Figure 4. Measures of innovation spreading versus socio-economic development in a country. (a) Predicted inverse speed of adoption spreading τ as a function of GDP *per capita* (the GDP dollar estimates used in our study are derived from purchasing power parity calculations using World Bank data, 2011; www.worldbank.org) for various countries. (b) Theoretical inverse termination probability $1/\bar{p}$ as a function of the empirical average account lifetime $\langle t_i \rangle$. (c) $\langle t_i \rangle$ as a function of a press liberty measure [38] (large scores imply weak liberties). \bar{T} and \bar{c} are arbitrary linear scaling constants with time dimensions for \bar{T} .

different levels of socio-economic development. This can be better understood by linking $\langle t_i \rangle$ with general civil liberty measures [38] (figure 4c). We observe that the weaker the press liberty is in a country, the shorter the time online accounts are used there (other liberty measures in the electronic supplementary material, §S4.3). Such observations indicate a quantifiable dependence between the dynamics of innovation spreading and the socio-economic status of a country.

3. Discussion

Our analysis of one of the largest online communication services worldwide aimed at clarifying several long-standing questions about the spreading mechanisms of novel technologies. We have shown that innovation diffusion can be interpreted as a competition between service adoption and termination; a process characterized, after a transition time, by constant rates and by a linearly increasing influence of user neighbours on service adoption. In addition, we have integrated the identified mechanisms into a minimal modelling framework that provides accurate medium-term predictions for the spreading of an online service.

It should be pointed out that this study has some limitations. First, the complete structure of society cannot be mapped by using online interactions only, as observations taken from any online social network underestimate the real number of contact peers of an ego. This incompleteness allows us only to estimate effective degrees and adoption thresholds. Second, the observation of correlated adoption does not necessarily imply the presence of actual social influence, only its possibility; even more so since other mechanisms like homophily cannot be synthesized from this dataset. Despite these limitations, the presented results provide strong evidence of key mechanisms driving the complex contagion of online technologies, up to a level of detail and scale that has not been possible before.

These results may help fill an enduring gap between the theoretical understanding and the empirical observation of social contagion phenomena, validating several earlier studies based on similar assumptions, like constant adoption

rates and the effect of social influence. In addition, we have shown how the adoption of novel technologies is related to the societal and economical development of a country. Beyond the clear advantage of these observations for the design of marketing and business plans, they also provide further insight into the differences in the development of modern online societies.

Acknowledgements. The authors gratefully acknowledge the support of M. Dumas and A. Saabas from STACC and Microsoft/Skype Labs, and from the ICTeCollective EU FP7 project. M.K. thanks R. Kikas for the data preparation and P. Gonçalves for useful comments.

Funding statement. G.I. acknowledges the Academy of Finland for funding. J.K. thanks FiDiPro (TEKES) and the DATASIM EU FP7 project for support, and S. Fortunato for discussions. This research was partly funded by Microsoft/Skype Labs.

Appendix A. Model description

For a static social network G with degree distribution ρ_k , the probability that individual i becomes a user is $p_i^+ = p_a + p_p(1 - p_a)n_i$ with $n_i = N_i/k_i$. Here, N_i is the number of neighbours of i that have already adopted the product and k_i its degree. Furthermore, the probability that i stops being a user is $p_i^- = p_r + p_s(1 - p_r)$. In the thermodynamic limit, we assume that all agents with the same degree are statistically equivalent, allowing us to group individuals and write rate equations for each degree class k . We denote by s_k , a_k and r_k the average probabilities that a randomly chosen agent with degree k is susceptible, adopter and removed, respectively. A first-order moment closure method leads to the rate equation $da_k/dt = \langle p_i^+ \rangle s_k - \langle p_i^- \rangle a_k$. In other words, the average probability that an adopting agent becomes either removed or susceptible is $\langle p_i^- \rangle a_k = [p_r + p_s(1 - p_r)]a_k$, while the average probability that a susceptible individual adopts the product is $\langle p_i^+ \rangle s_k = [p_a + p_p(1 - p_a)n_a]s_k$ with $n_a = \langle n_i \rangle$. This approximation ignores higher moments of the dynamical quantities s_k , a_k and r_k , as well as any correlations between them. In the presence of degree-degree correlations in G , we have $n_a = \sum_{k'} (k' - 1)\rho_{k',k}a_{k'}/k'$, where $\rho_{k',k}$ is the conditional

probability that an edge departing from an agent with degree k arrives at an agent with degree k' . Similar rate equations can be written for s_k and r_k , leading to a system of nonlinear ordinary

differential equations that determines adoption at the degree class level (for further details, see the electronic supplementary material, S3).

References

- Solow RM. 2007 The last 50 years in growth theory and the next 10. *Oxford Rev. Econ. Policy* **23**, 3–14. (doi:10.1093/oxrep/grm004)
- Bass FM. 1969 A new product growth for model consumer durables. *Manage. Sci.* **15**, 215–227. (doi:10.1287/mnsc.15.5.215)
- Toole JL, Cha M, González MC. 2012 Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE* **7**, e29528. (doi:10.1371/journal.pone.0029528)
- Onnela JP, Reed-Tsochas F. 2010 Spontaneous emergence of social influence in online systems. *Proc. Natl Acad. Sci. USA* **107**, 18 375–18 380. (doi:10.1073/pnas.0914572107)
- Aral S, Muchnik L, Sundararajan A. 2009 Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl Acad. Sci. USA* **106**, 21 544–21 549. (doi:10.1073/pnas.0908800106)
- Qudrat-Ullah H, Spector JM, Davidsen PI (eds). 2007 *Complex decision making*. New York, NY: Springer.
- Rogers EM. 2003 *Diffusion of innovations*, 5 edn. New York, NY: Free Press.
- Wejnert B. 2002 Integrating models of diffusion of innovations: a conceptual framework. *Annu. Rev. Sociol.* **28**, 297–326. (doi:10.1146/annurev.soc.28.1.10601.141051)
- Barrat A, Barthelemy M, Vespignani A. 2008 *Dynamical processes on complex networks*. Cambridge, UK: Cambridge University Press.
- Vespignani A. 2009 Predicting the behavior of techno-social systems. *Science* **325**, 425–428. (doi:10.1126/science.1171990)
- Jackson MO, Leeat Y. 2005 Diffusion on social networks. *Econ. Pub.* **16**, 3–16.
- López-Pintado D. 2008 Diffusion in complex social networks. *Games Econ. Behav.* **62**, 573–590. (doi:10.1016/j.geb.2007.08.001)
- Centola D, Macy M. 2007 Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734. (doi:10.1086/521848)
- Hill AL, Rand DG, Nowak MA, Christakis NA. 2010 Infectious disease modeling of social contagion in networks. *PLoS Comput. Biol.* **6**, e1000968. (doi:10.1371/journal.pcbi.1000968)
- Granovetter M. 1978 Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443. (doi:10.1086/226707)
- Watts DJ. 2002 A simple model of global cascades on random networks. *Proc. Natl Acad. Sci. USA* **99**, 5766–5771. (doi:10.1073/pnas.082090499)
- Backstrom L, Huttenlocher D, Kleinberg J, Lan X. 2006 Group formation in large social networks: membership, growth, and evolution. In *Proc. 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD'06*, pp. 44–54. New York, NY: ACM. (doi:10.1145/1150402.1150412)
- Gleeson JP, Cahalane DJ. 2007 Seed size strongly affects cascades on random networks. *Phys. Rev. E* **75**, 056103. (doi:10.1103/PhysRevE.75.056103)
- Romero DM, Meeder B, Kleinberg J. 2011 Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proc. 20th Int. Conf. on World Wide Web, WWW'11*, pp. 695–704. New York, NY: ACM. (doi:10.1145/1963405.1963503)
- Bakshy E, Rosenn I, Marlow C, Adamic L. 2012 The role of social networks in information diffusion. In *Proc. 21st Int. Conf. on World Wide Web, WWW'12*, pp. 519–528. New York, NY: ACM. (doi:10.1145/2187836.2187907)
- Singh P, Sreenivasan S, Szymanski BK, Korniss G. 2013 Threshold-limited spreading in social networks with multiple initiators. *Sci. Rep.* **3**, 2330. (doi:10.1038/srep02330)
- Centola D. 2010 The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197. (doi:10.1126/science.1185231)
- Centola D. 2011 An experimental study of homophily in the adoption of health behavior. *Science* **334**, 1269–1272. (doi:10.1126/science.1207055)
- Suri S, Watts DJ. 2011 Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE* **6**, e16836. (doi:10.1371/journal.pone.0016836)
- Morrissey RC, Goldman ND, Kennedy KP. 2011 Skype SA, US SEC registration statement, amendment 3, registration no. 333-168646.
- Internet World Stats. 2011 www.internetworldstats.com (accessed 15 July 2013).
- Vespignani A. 2012 Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* **8**, 32–39. (doi:10.1038/nphys2160)
- Castellano C, Fortunato F, Loreto V. 2009 Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–645. (doi:10.1103/RevModPhys.81.591)
- Dodds PS, Watts DJ. 2004 Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.* **92**, 218701. (doi:10.1103/PhysRevLett.92.218701)
- Klimek P, Lambiotte R, Thurner S. 2008 Opinion formation in laggard societies. *Europhys. Lett.* **82**, 28008. (doi:10.1209/0295-5075/82/28008)
- Takaguchi T, Masuda N, Holme P. 2013 Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics. *PLoS ONE* **8**, e68629. (doi:10.1371/journal.pone.0068629)
- Dodds PS, Harris KD, Danforth CM. 2013 Limited imitation contagion on random networks: chaos, universality, and unpredictability. *Phys. Rev. Lett.* **110**, 158701. (doi:10.1103/PhysRevLett.110.158701)
- Goffman W, Newill VA. 1964 Generalization of epidemic theory. *Nature* **204**, 225–228. (doi:10.1038/204225a0)
- Daley DJ, Kendall DG. 1964 Epidemics and rumours. *Nature* **204**, 1118. (doi:10.1038/2041118a0)
- Pastor-Satorras R, Vespignani A. 2001 Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)
- Barabási AL, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
- Gracia-Lázaro C, Ferrer A, Ruiz G, Tarancón A, Cuesta JA, Sánchez A, Moreno Y. 2012 Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma. *Proc. Natl Acad. Sci. USA* **109**, 12 922–12 926. (doi:10.1073/pnas.1206681109)
- Freedom House. 2011 *Freedom in the world*. www.freedomhouse.org (accessed 6 March 2013).