

# SCIENTIFIC REPORTS



OPEN

## Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading

Received: 03 February 2016

Accepted: 13 May 2016

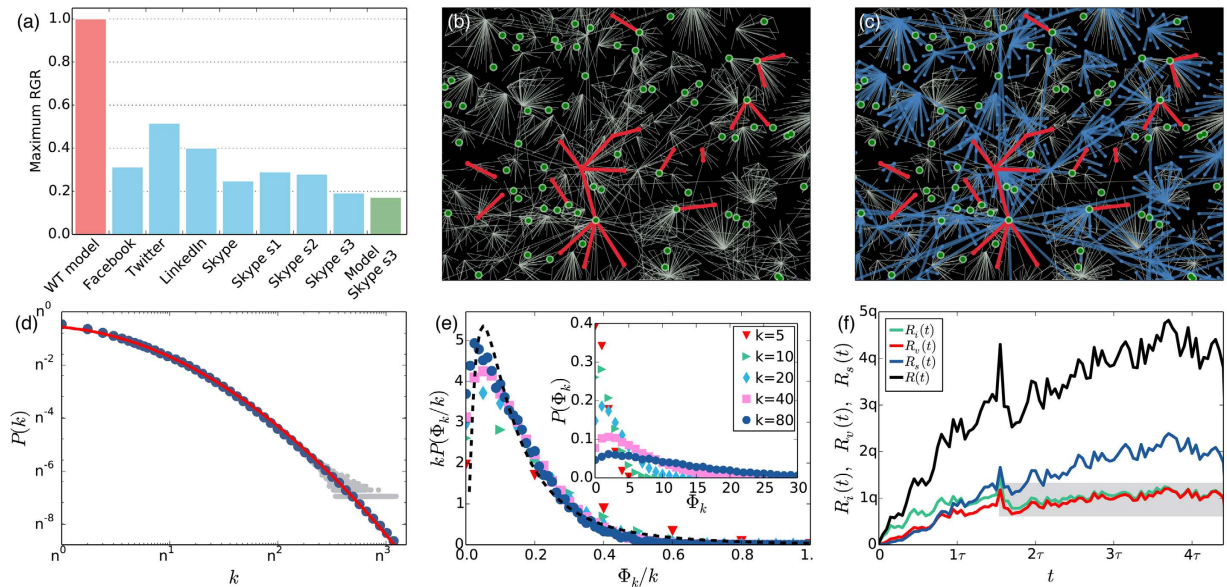
Published: 07 June 2016

Márton Karsai<sup>1</sup>, Gerardo Iñiguez<sup>2,3</sup>, Riivo Kikas<sup>4,5</sup>, Kimmo Kaski<sup>2</sup> & János Kertész<sup>6,7</sup>

Adoption of innovations, products or online services is commonly interpreted as a spreading process driven to large extent by social influence and conditioned by the needs and capacities of individuals. To model this process one usually introduces behavioural threshold mechanisms, which can give rise to the evolution of global cascades if the system satisfies a set of conditions. However, these models do not address temporal aspects of the emerging cascades, which in real systems may evolve through various pathways ranging from slow to rapid patterns. Here we fill this gap through the analysis and modelling of product adoption in the world's largest voice over internet service, the social network of Skype. We provide empirical evidence about the heterogeneous distribution of fractional behavioural thresholds, which appears to be independent of the degree of adopting egos. We show that the structure of real-world adoption clusters is radically different from previous theoretical expectations, since vulnerable adoptions—induced by a single adopting neighbour—appear to be important only locally, while spontaneous adopters arriving at a constant rate and the involvement of unconcerned individuals govern the global emergence of social spreading.

Spreading of opinions, frauds, behavioural patterns, and product adoptions are all examples of social contagion phenomena where collective patterns emerge due to correlated decisions of a large number of individuals. Although these choices are personal, they are not independent but potentially driven by several processes such as social influence<sup>1</sup>, homophily<sup>2</sup>, and information arriving from external sources like news or mass media<sup>3</sup>. Social contagion evolves over networks of interconnected individuals, where links associated with social ties transfer influence between peers<sup>4</sup>. Several earlier studies aimed to identify the dominant mechanisms at play in social contagion processes<sup>5–8</sup>. One key element, termed behavioural threshold by Granovetter<sup>6</sup>, is defined as “the number or proportion of others who must make one decision before a given actor does so”. Following this idea various network models have been introduced<sup>9–15</sup> to understand the threshold-driven spreading, commonly known as *complex contagion*<sup>16</sup>. Although these models are related to a larger set of collective dynamics, they are particularly different from *simple contagion* where the exposure of nodes is driven by independent contagion stimuli<sup>17,18</sup>. In addition, collective adoption patterns may appear as a consequence of homophilic structural correlations, where connected individuals adopt due to their similar interests and not due to direct social influence. Distinguishing between the effects of social influence and homophily at the individual level remains as a challenge<sup>19,20</sup>. Furthermore, in real social spreading phenomena all these mechanisms are arguably present. However, while in the case of homophily the adoption behaviour is only seemingly correlated, and for simple contagion only the number of exposures

<sup>1</sup>Univ de Lyon, ENS de Lyon, INRIA, CNRS, UMR 5668, IXXI, 69364 Lyon, France. <sup>2</sup>Department of Computer Science, School of Science, Aalto University, 00076, Finland. <sup>3</sup>Centro de Investigación y Docencia Económicas, CONACYT, 01210 México D.F., Mexico. <sup>4</sup>Institute of Computer Science, University of Tartu, 50409 Tartu, Estonia. <sup>5</sup>Software Technology and Applications Competence Center (STACC), 51003 Tartu, Estonia. <sup>6</sup>Center for Network Science, Central European University, 1051 Budapest, Hungary. <sup>7</sup>Institute of Physics, Budapest University of Technology and Economics, 1111 Budapest, Hungary. Correspondence and requests for materials should be addressed to M.K. (email: marton.karsai@ens-lyon.fr)



**Figure 1. Structure and dynamics of online service adoption.** (a) Yearly maximum relative growth rate (RGR) of cumulative adoption (see Appendix) for several online social-communication services<sup>48</sup>, including three Skype paid services (s1 - “subscription”, s2 - “voicemail”, and s3 - “buy credit”). The red bar corresponds to a rapid cascade of adoption suggested by the Watts threshold (WT) model, while the green bar is the model prediction for Skype s3. (b,c) Snowball sample of the Skype social network (gray links) with nodes and links coloured according to their adoption state: multiple innovators (green nodes), induced small vulnerable trees (red nodes and links), and the triggered connected stable cluster (blue nodes and links). Note that some vulnerable and stable clusters seemingly appear without an innovator seed due to the finite distance used in the snowball sampling method. (d) Degree distribution  $P(k)$  of the Skype network (gray/blue circles for raw/binned data) on double log-scale with arbitrary base  $n$ .  $P(k)$  is fitted by a lognormal distribution (see Appendix and SI) with parameters  $\mu_D = 1.2$  and  $\sigma_D = 1.39$ , and average  $z = 8.56$  (red line). (e) Distribution  $P(\Phi_k)$  of integer thresholds  $\Phi_k$  for several degree groups in Skype s3 (inset). By using  $P(\Phi_k, k) = kP(\Phi_k/k)$ , these curves collapse to a master curve approximated by a lognormal function (dashed line in main panel) with parameters  $\mu_T = -2$  and  $\sigma_T = 1$ , as constrained by the average threshold  $w = 0.19$  (see Appendix and SI). (f) Adoption rate of innovators [ $R_i(t)$ ], vulnerable nodes [ $R_v(t)$ ], and stable nodes [ $R_s(t)$ ], as well as net service adoption rate [ $R(t)$ ]. Rates are measured with a 1-month time window, while  $q$  and  $\tau$  are arbitrary constants. The shaded area indicates the regime where innovators adopt approximately with constant rate.

matters, in complex contagion the fraction of adopting neighbours relative to the total number of partners determines whether a node adopts or not, capturing the natural mechanisms involved in individuals’ decision makings<sup>21–23</sup>. Due to this additional complexity, threshold models are able to emulate system-wide adoption patterns known as global cascades. Regarding the effects of social influence several assumptions have been proposed about its functional dependency on the number of influencers. While Granovetter and others<sup>6,9</sup> suggest a simple linear correspondence, e.g. observed in large techno-social systems<sup>21</sup>, from the theory of Latane<sup>24</sup> one could derive a non-linear dependency as recently demonstrated by small-scale online experiments<sup>1</sup>.

Behavioural cascades are rare but potentially stupendous social spreading phenomena, where collective patterns of exposure emerge as a consequence of small initial perturbations. Some examples are the rapid emergence of political and grass-root movements<sup>25–27</sup>, fast spreading of information<sup>12,28–34</sup> or behavioural patterns<sup>35</sup>, etc. The characterisation<sup>34,36–40</sup> and modelling<sup>9,41–43</sup> of such processes have received plenty of attention and provide some basic understanding of the conditions and structure of empirical and synthetic cascades on various types of networks<sup>44–47</sup>. However, these studies commonly fail in addressing the temporal dynamics of the emerging cascades, which may vary considerably between different cases of social contagion. Moreover, they have not answered why real-world cascades can evolve through various dynamic pathways ranging from slow to rapid patterning, especially in systems where the threshold mechanisms play a role and social phenomena spread globally. Besides the case of rapid cascading mentioned above, an example of the other extreme is the propagation of products in social networks<sup>18</sup>, where adoption evolves gradually even if it is driven by threshold mechanisms and may cover a large fraction of the total population<sup>21</sup>. This behaviour characterises the adoption of online services such as Facebook, Twitter, LinkedIn and Skype (Fig. 1a), since their yearly maximum relative growth of cumulative adoption<sup>48</sup> (for definition see Appendix) is lower than in the case of rapid cascades as suggested e.g. by the Watts threshold (WT) model.

To fill this gap in the modelling of social diffusion, here we will analyse and model real-world examples of social contagion phenomena. Our aim is to identify the crucial mechanisms necessary to consider in models of complex contagion to match them better with reality, and define a model that incorporates these mechanisms and captures the possible dynamics leading to the emergence of real-world global cascades. We follow the adoption

dynamics of the Skype paid service “buy credit” for 89 months since 2004, which evolves over the social network of one of the largest voice over internet providers in the world. Data includes the time of first payment of each user, an individual and conscious action that tracks adoption behaviour. In addition we follow the “subscription” service over 42 months since 2008 (for results see Supplementary Information [SI]). In contrast to other empirical studies where incomplete knowledge about the underlying social network leads to unavoidable bias<sup>21</sup>, we use here the largest connected component of the aggregated free Skype service as the underlying structure, where nodes are Skype users and links confirmed contacts between them. This is a good approximation since it maps all connections in the Skype social network without sampling, and the paid service is only available for individuals already enrolled in the Skype network. Also note that the service adoption process evolves in a considerably faster time-scale than the underpinning social network. This way applying a time-scale separation, and considering the network to be static, may provide a good first approximation here. The underlying structure is an aggregate from September 2003 to November 2011 (i.e. over 99 months) and contains roughly 4.4 billion links and 510 million registered users worldwide<sup>49</sup>. The data is fully anonymised and considers only confirmed connections between users (for more data details see SI).

In what follows we first provide empirical evidence of the distribution of individual adoption thresholds and other structural and dynamical features of a worldwide adoption cluster. We incorporate the observed structural and threshold heterogeneities into a dynamical threshold model where multiple nodes adopting spontaneously (i.e. firstly among their neighbours) are allowed<sup>50</sup>. We find that if the fraction of users who reject to adopt the product is large, the system enters a quenched state where the evolution and structure of the global adoption cluster is very similar to our empirical observations. Model calculations and the analysis of the real social contagion process suggest that the evolving structure of an adoption cluster differs radically from what has been proposed earlier<sup>9</sup>, since it is triggered by several spontaneous adoptions arriving at a constant rate, while stable adopters who are initially resisting exposure, are actually responsible for the emergence of global social adoption (Fig. 1b,c).

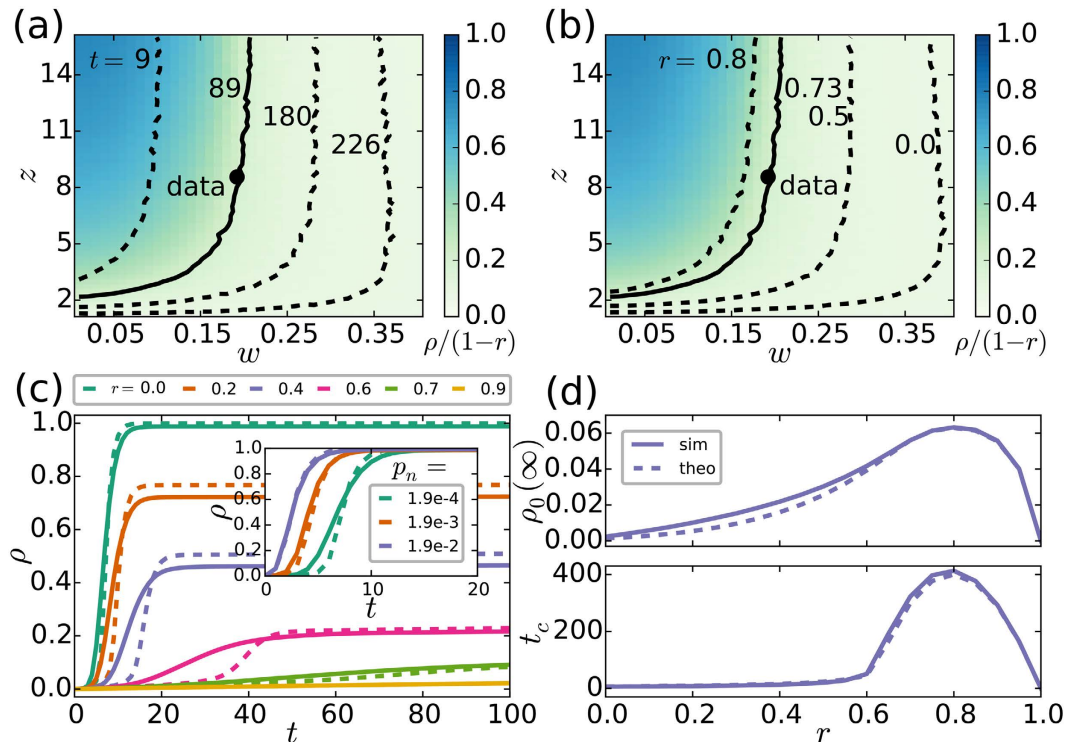
## Results

Social contagion phenomena can be modelled as binary-state processes evolving on networks and driven by threshold mechanisms. In these systems individuals are represented by nodes, each being either in a susceptible (0) or adopter (1) state and influencing each other by transferring information via social ties<sup>6</sup>. Nodes are connected in a network with degree distribution  $P(k)$  and average degree  $z = \langle k \rangle$ . In addition, each node has an individual threshold  $\phi \in [0, 1]$  drawn from a distribution  $P(\phi)$  with average  $w = \langle \phi \rangle$ . This threshold determines the minimum fraction of exposed neighbours that triggers adoption and captures the resistance of an individual against engaging in spreading behaviour. Once a node reaches its threshold, it switches state from 0 to 1 and keeps it until the end of the dynamics. In his seminal paper about threshold dynamics, Watts<sup>9</sup> classified nodes into three categories based on their threshold and degree. He identified *innovator* nodes that spontaneously change state to 1, thus starting the process. Such nodes have a trivial threshold  $\phi = 0$ . Then there are nodes with threshold  $0 < \phi \leq 1/k$ , called *vulnerable*, which need one adopting neighbour before their own adoption. Finally, there are more resilient nodes with threshold  $\phi > 1/k$ , denoted as *stable*, referring to individuals in need of strong social influence to follow the actions of their acquaintances.

In the WT model<sup>9</sup>, small perturbations (like the spontaneous adoption of a single seed node) can trigger global cascading patterns. However, their emergence is subject to the so-called *cascade condition*: the innovator seed has to be linked to a percolating vulnerable cluster, which adopts immediately afterwards and further triggers a global cascade (i.e. a set of adopters larger than a fixed fraction of the finite network). The cascade condition is satisfied if the network is inside a bounded regime in  $(w, z)$ -space<sup>9</sup>. This regime depends on degree and threshold heterogeneities<sup>9</sup> and may change its shape if several innovators start the process<sup>42</sup>. In addition, while models with more sophisticated social influence function can be introduced<sup>24,51</sup> the original linear-threshold assumption proposed by Watts and Granovetter seems to be sufficient to interpret our observations.

**Empirical observations.** Degree and threshold heterogeneities are indeed present in the social network of Skype. The degree distribution  $P(k)$  is well approximated by a lognormal function  $P(k) \propto k^{-1} e^{-(\ln k - \mu_D)^2 / (2\sigma_D^2)}$  ( $k \geq k_{\min}$ ) with parameters  $\mu_D = 1.2$ ,  $\sigma_D = 1.39$  and  $k_{\min} = 1$  (Fig. 1d), giving an average degree  $z = 8.56$  (for goodness of fit see SI). Moreover, at the time of adoption we can measure the threshold  $\phi = \Phi_k/k$  of a user by counting the number  $\Phi_k$  of its neighbours who have adopted the service earlier. We then group users by degree and calculate the distribution  $P(\Phi_k)$  of the integer threshold  $\Phi_k$ <sup>38</sup> (Fig. 1e). By using the scaling relation  $P(\Phi_k, k) = kP(\Phi_k/k)$  all distributions collapse to a master curve well approximated by a lognormal function  $P(\phi) \propto \phi^{-1} e^{-(\ln \phi - \mu_T)^2 / (2\sigma_T^2)}$ , with parameters  $\mu_T = -2$  and  $\sigma_T = 1$  as constrained by the average threshold  $w = 0.19$  (see Appendix and SI). Note that we observe qualitatively the same scaling and lognormal shape of the threshold distribution for another service (see SI). These empirical observations, in addition to the broad degree distribution, provide quantitative evidence about the heterogeneous nature of adoption thresholds.

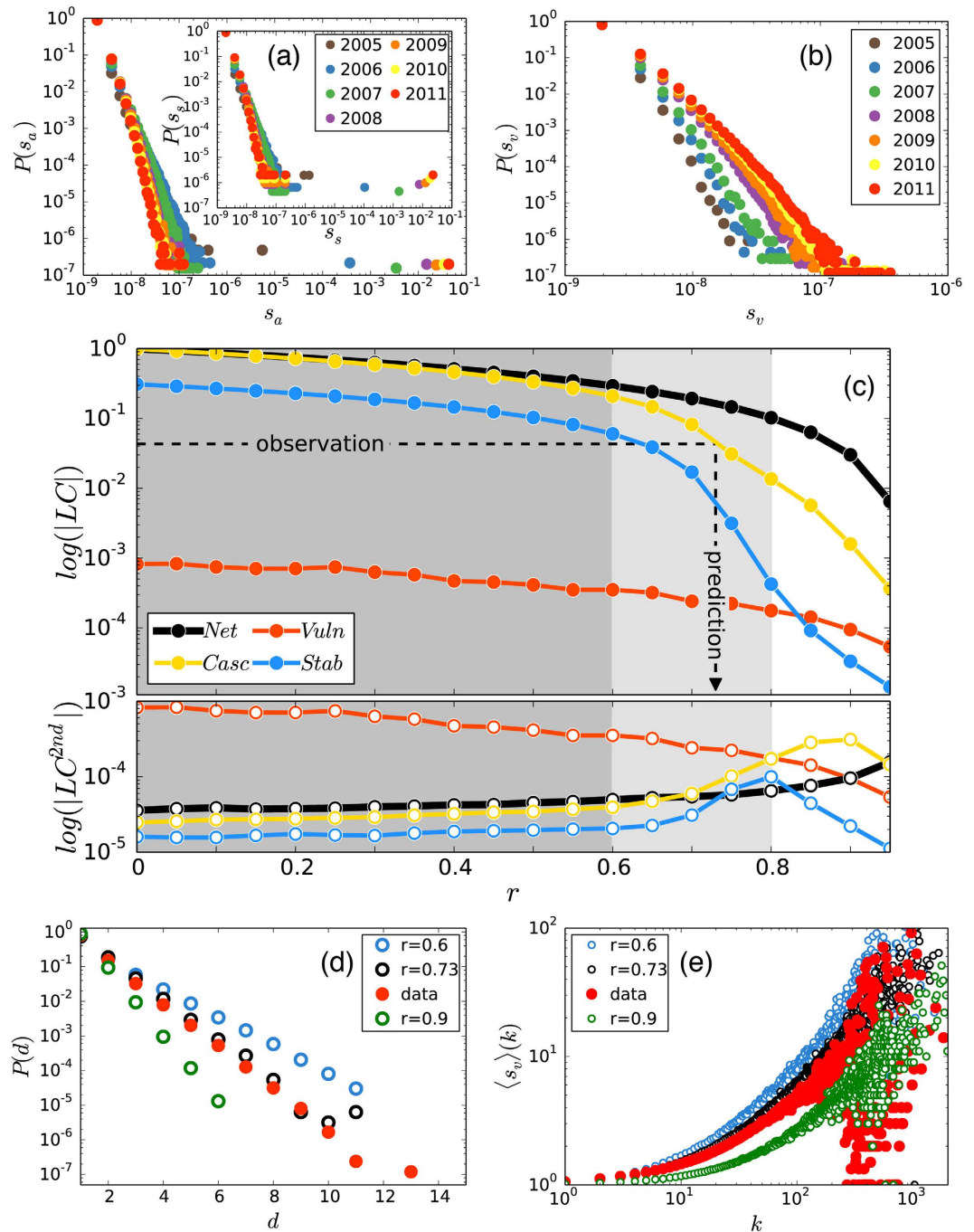
Since we know the complete structure of the online social network, as well as the first time of service usage for all adopters, we can follow the temporal evolution of the adoption dynamics. By counting the number of adopting neighbours of an ego, we identify innovators ( $\Phi_k = 0$ ), and vulnerable ( $\Phi_k = 1$ ) or stable ( $\Phi_k > 1$ ) nodes. The adoption rates for these categories behave rather differently from previous suggestions<sup>9</sup> (Fig. 1f). First, there is not only one seed but an increasing fraction of innovators in the system who, after an initial period, adopt approximately at a constant rate. Second, vulnerable nodes adopt approximately with the same rate as innovators suggesting a strong correlation between these types of adoption. This stationary behaviour is rather surprising as environmental effects, like competition or marketing campaigns, potentially influence the adoption dynamics. Nevertheless, this pattern appears to be consistent among the two investigated services (also see Fig. S5 in SI), where innovator



**Figure 2. Threshold model for the adoption of online services.** (a,b) Surface plot of the normalised fraction of adopters  $\rho/(1-r)$  in  $(w, z)$ -space, for  $r=0.73$  and  $t=89$ . Contour lines signal parameter values for which 20% of non-immune nodes have adopted, for fixed  $r$  and varying time (a), and for fixed time and varying  $r$ . (b) The continuous contour line and dot indicate parameter values in the last observation of Skype s3. A regime of maximal adoption ( $\rho \approx 1-r$ ) grows as time goes by, and shrinks for larger  $r$ . (c) Time series of the fraction of adopters  $\rho$  for fixed  $p_n = 0.00019$  and varying  $r$  (main), and for fixed  $r=0$  and varying  $p_n$  (inset). These curves are well approximated by the solution of Eq. (1) for  $k_0=3$ ,  $k_{M-1}=150$  and  $M=25$  (dashed lines). The dynamics is clearly faster for larger  $p_n$ . As  $r$  increases, the system enters a regime where the dynamics is slowed down and adopters are mostly innovators. (d) Final fraction of innovators  $\rho_0(\infty)$  and time  $t_c$  when 50% of non-immune nodes have adopted as a function of  $r$ , both simulated and theoretical. The crossover to a regime of slow adoption is characterised by a maximal fraction of innovators and time  $t_c$ . Unless otherwise stated,  $p_n = 0.00019$  and we use  $N=10^4$ ,  $\mu_D = 1.09$ ,  $\sigma_D = 1.39$ ,  $k_{\min} = 1$ ,  $\mu_T = -2$ , and  $\sigma_T = 1$  to obtain  $z = 8.56$  and  $w = 0.19$  as in Skype s3. The difference in  $\mu_D$  between data and model is due to finite-size effects (see Appendix). Numerical results are averages over  $10^2$  (a-b) and  $10^3$  (c-d) realisations.

and vulnerable adoptions evolve with approximately constant rates beyond statistical and periodic (circadian, monthly, yearly) fluctuations. On the other hand, the overall adoption process accelerates due to the increasing rate of stable adoptions induced by social influence. At the same time a giant adoption cluster grows and percolates through the whole network (Fig. 3a, main panel). Despite of this expansion dynamics and connected structure of the service adoption cluster, the service reaches less than 6% of the total number of active Skype users over a period of 7 years<sup>49</sup>. Therefore we ask whether one can refer to these adoption clusters as cascades. They are not triggered by a small perturbation but induced by several innovators; their evolution is not instantaneous but ranges through several years; and although they involve millions of individuals, they reach only a reduced fraction of the whole network. To answer we incorporate the above mentioned features into a dynamical threshold model<sup>50</sup> with a growing group of innovators and investigate their effect on the evolution of global social adoption. Note that although we cannot follow the direct pathways of social influence, we perform a null model study to demonstrate at the system level that social influence is present and dominates the contagion process, as compared to effects of homophily (see section S3 of the SI, together with another empirical spreading scenario in S7.1).

**Model.** Our modelling framework is an extension to conventional threshold dynamics on networks studied by Watts, Gleeson, Singh, and others, where all nodes are initially susceptible and innovators are only introduced as an initial seed of arbitrary size<sup>9,15,42,43</sup>. Apart from the threshold rule discussed above, our model considers two additional features: (i) a fraction  $r$  of ‘immune’ nodes that never adopt, indicating lack of interest in the service; (ii) due to external influence, susceptible nodes adopt the innovation spontaneously (i.e. become innovators) throughout time with constant rate  $p_n$ , rather than only at the beginning of the dynamics. In this way, the dynamical evolution of the system is completely defined by the online social network, the distribution  $P(\phi)$  and the parameters  $r, p_n$ . For the sake of simplicity we consider a configuration-model network, i.e., we ignore correlations in the social network and characterise it solely by its degree distribution  $P(k)$ . Furthermore, node degrees and



**Figure 3. Empirical cluster statistics and simulation results.** (a) Empirical connected-component size distribution at different times for the adoption  $P(s_a)$ , main panel] and stable adoption  $P(s_s)$ , inset] networks, with  $s_a$  and  $s_s$  relative to system size. (b) Empirical connected-component size distribution  $P(s_v)$  for the relative size of innovator-induced vulnerable trees at different times. (c) Average size of the largest ( $LC$ ) and 2nd largest ( $LC^{2nd}$ ) components of the model network ('Net'), adoption network ('Casc'), stable network ('Stab'), and induced vulnerable trees ('Vuln') as a function of  $r$ . Dashed lines show the observed relative size of the real  $LC$  of the adopter network in 2011 [see main panel in (a)] and the predicted  $r$  value. (d) Distribution  $P(d)$  of depths of induced vulnerable trees in both data and model for several  $r$  values, showing a good fit with the data for  $r = 0.73$ . The difference in the tail is due to finite-size effects. (e) Correlation  $\langle s_v \rangle(k)$  between innovator degree and average size of vulnerable trees in both data and model with the same  $r$  values as in (e). Model calculations for (d,e) correspond to networks of size  $N = 10^6$  and are averaged over  $10^2$  realisations.

thresholds are considered to be independent<sup>38,52,53</sup>. We remark that somewhat similar concepts called “stubborn nodes” mimicking individuals’ resistance against adoption<sup>54,55</sup>, and “global nodes” capturing adoptions driven by external effects<sup>56</sup> have been considered in threshold models with a rich variety of effects on cascading behaviour.

Our threshold model, which has also been introduced in<sup>50</sup>, can be studied analytically by extending the framework of approximate master equations (AMEs) for monotone binary-state dynamics recently developed by Gleeson<sup>38,52,53</sup>, where the transition rate between susceptible and adoption states only depends on the number  $m$  of network neighbours that have already adopted. We describe a node by the property vector  $\mathbf{k} = (k, c)$ , where  $k = k_0, k_1, \dots, k_{M-1}$  is its degree and  $c = 0, 1, \dots, M$  its type, i.e.  $c = 0$  is the type of the fraction  $r$  of immune nodes, while  $c \neq 0$  is the type of all non-immune nodes that have threshold  $\phi_c$ . In this way  $P(\phi)$  is substituted by the discrete distribution of types  $P(c)$  (for  $c > 0$ ). The integer  $M$  is the maximum number of degrees (or non-zero types) considered in the AME framework, which can be increased to improve the accuracy of the analytical approximation at the expense of speed in its numerical computation (see S4.2). Under these conditions, the AME system describing the dynamics of the threshold model is reduced to the pair of ordinary differential equations (see SI),

$$\dot{\rho} = h(\nu, t) - \rho, \quad (1a)$$

$$\dot{\nu} = g(\nu, t) - \nu, \quad (1b)$$

where  $\rho(t)$  is the fraction of adopters in the network,  $\nu(t)$  is the probability that a randomly chosen neighbour of a susceptible node is an adopter, and the initial conditions are  $\rho(0) = \nu(0) = 0$ . Here,

$$h = (1 - r)[f_t + (1 - f_t) \sum_{\mathbf{k}|c \neq 0} P(k)P(c) \sum_{m \geq k\phi_c} B_{k,m}(\nu)], \quad (2)$$

and,

$$g = (1 - r)[f_t + (1 - f_t) \sum_{\mathbf{k}|c \neq 0} \frac{k}{z} P(k)P(c) \sum_{m \geq k\phi_c} B_{k-1,m}(\nu)], \quad (3)$$

where  $f_t = 1 - (1 - p_r)e^{-p_r t}$ ,  $p_r = p_n/(1 - r)$ , and  $B_{k,m}(\nu) = \binom{k}{m} \nu^m (1 - \nu)^{k-m}$  is the binomial distribution. The fraction of adopters  $\rho$  is then obtained by solving Eq. (1) numerically. Since susceptible nodes adopt spontaneously with rate  $p_n$ , the fraction of innovators  $\rho_0(t)$  in the network is given by (see S4.3),

$$\rho_0 = p_r \int_0^t (1 - r - \rho) dt. \quad (4)$$

We also implement the threshold model numerically via a Monte Carlo simulation in a network of size  $N$ , with a lognormal degree distribution and a lognormal threshold distribution as observed empirically. Thus, we can explore the behaviour of  $\rho$  and  $\rho_0$  as a function of  $z$ ,  $w$ ,  $p_n$  and  $r$ , both in the numerical simulation and in the theoretical approximation given by Eqs (1) and (4). For  $p_n > 0$  some nodes adopt spontaneously as time passes by, leading to a frozen state characterised by a final fraction  $\rho(\infty) = 1 - r$  of adopters. However, the time needed to reach such state depends heavily on the distribution of degrees and thresholds, as signalled by a region of large adoption ( $\rho \approx 1 - r$ ) that grows in  $(w, z)$ -space with time (contour lines in Fig. 2a). If we fix a time in the dynamics and vary the fraction of immune nodes instead, this region shrinks as  $r$  increases (contour lines in Fig. 2b). In other words, the set of networks (defined by their average degree and threshold) that allow the spread of adoption is larger at later times in the dynamics, or when the fraction of immune nodes is small. When both  $t$  and  $r$  are fixed, the normalised fraction of adopters  $\rho/(1 - r)$  gradually decreases for less connected networks with larger thresholds (surface plot in Fig. 2a,b).

For  $r \approx 0$  the critical fraction of innovators necessary to trigger a cascade of fast adoption throughout all susceptible nodes may be identified as the inflection point in the time series of  $\rho$  (Fig. 2c, inset). The adoption cascade appears sooner for larger  $p_n$ , since this parameter regulates how quickly the critical fraction of innovators is reached. Yet as we increase  $r$  above a threshold  $r_c$ , the system enters a regime where rapid cascades disappear and adoption is slowed down. The crossover between these regimes is gradual, as seen in the shape of  $\rho$  for increasing  $r$  (Fig. 2c, main panel). We may identify  $r_c$  in various ways: by the maximum in both the final fraction of innovators  $\rho_0(\infty)$  and the critical time  $t_c$  when  $\rho = (1 - r)/2$  (Fig. 2d), or as the  $r$  value where the inflection point in  $\rho$  disappears. These measures indicate  $r_c \approx 0.8$  for the chosen parameters. All global properties of the dynamics (like the functional dependence of  $\rho$  and  $\rho_0$ ) are very well approximated by the solution of Eqs (1) and (4) (dashed lines in Fig. 2c,d). Indeed, the AME framework is able to capture the shape of the  $\rho$  time series, the crossover between regimes of fast and slow adoption, as well as the maximum in  $\rho_0(\infty)$  and  $t_c$ .

**Validation.** To better understand how innovation spreads throughout real social networks, we take a closer look at the internal structure of the service adoption process. By taking into account individual adoption times we construct an evolving adoption network with links between users who have adopted the service before time  $t$  and are connected in the social structure. In order to avoid the effect of instantaneous group adoptions (evidently not driven by social influence), we only consider links between nodes who are neighbours in the underlying social network and whose adoption did not happen at the same time. This way links in the adoption graph indicate ties where social influence among individuals could have existed. The size distribution  $P(s_n)$  of connected components in the adoption network shows the emergence of a giant percolating component over time (Fig. 3a), along with several other small clusters. Moreover, after decomposition we observe that the giant cluster does not consist of a single innovator seed and percolating vulnerable tree<sup>9</sup>, but builds up from several innovator seeds that induce small vulnerable trees locally (Fig. 3b), each with small depth (Fig. 3d)<sup>34,57</sup>. At the same time the stable adoption network (considering connections between all stable adopters at the time) has a giant connected

component, indicating the emergence of a percolating stable cluster with size comparable to the largest adoption cluster (Fig. 3a, inset). These observations suggest a scenario for the evolution of the global adoption component different from earlier threshold models<sup>9</sup>. It appears that here multiple innovators adopt at different times and trigger local vulnerable trees (Fig. 1b), which in turn induce a percolating component of connected stable nodes that holds the global adoption cluster together (Fig. 1c). Consequently, in the structure of the adoption network primary triggering effects are important only locally, while external and secondary triggering mechanisms seem to be responsible for the emergence of global-scale adoption.

To model the observed dynamics and explore the effect of immune nodes, we perform extensive numerical simulations of the threshold model with parameters determined directly from the data (see Appendix and SI). We use a network structure with empirical degree and threshold distributions and fix  $p_n = 0.00019$  as the constant rate of innovators, implying that the time scale of a Monte Carlo iteration in the model is 1 month. We measure the average size of the largest ( $LC$ ) and second largest ( $LC^{2nd}$ ) connected components of the background social network, and of the stable, vulnerable and global adoption networks, as a function of the fraction of immune nodes  $r$ . After  $T = 89$  iterations (matching the length of the real observation period) we identify three regimes of the dynamics (Fig. 3c): if  $0 < r < 0.6$  (dark-shaded area) the spreading process is very rapid and evolves in a global cascade, which reaches most of the nodes of the shrinking susceptible network in a few iteration steps. About 10% of adopters are connected in a percolating stable cluster, while vulnerable components remain very small in accordance with empirical observations. In the crossover regime  $0.6 < r < 0.8$  (light-shaded area), the adoption process slows down considerably (Fig. 2d, lower panel), as stable adoptions become less likely due to the quenching effect of immune nodes. The adoption process becomes the slowest at  $r_c = 0.8$  (Fig. 2d, lower panel) when the percolating stable cluster falls apart, as demonstrated by a peak in the corresponding  $LC^{2nd}$  curve (Fig. 3c, lower panel). Finally, around  $r = 0.9$  the adoption network becomes fragmented and no global diffusion takes place. We repeat the same calculations for another service and find qualitatively the same picture, but with the crossover regime shifted towards larger  $r$  values due to the different parametrisation of the model process. Note that another possible reason for the slow adoption could be the time users wait between their threshold has been reached and actual adoption. We test for the effect of this potential scenario on the empirical curves but find no qualitative change in the dynamics (see SI).

We can use these calculations to estimate the only unknown parameter (the fraction  $r$  of immune nodes in Skype) by matching the size of the largest component ( $LC_{Net}$ ) between real and model adoption networks at time  $T$ . Empirically, this value is the relative size corresponding to the last point on the right-hand side of the distribution for 2011 (Fig. 3a, main panel). The corresponding value in the model is  $r = 0.73$  (dashed lines in Fig. 3c; also Fig. 2a,b), suggesting that the real adoption process lies in the crossover regime. The other analysed service turns out to lie right of the crossover regime, which explains its large innovator adoption rate and reduced size of stable and vulnerable adoption clusters (see SI).

To test the validity of the prediction of  $r$  we perform three different calculations. First we measure the maximum relative growth rate of cumulative adoption and find a good match between model and data (Skype s3 and Model Skype s3 in Fig. 1a). In other words, the model correctly estimates the speed of the adoption process. Second, we measure the distribution  $P(d)$  of depths of induced vulnerable trees (Fig. 3d). Finally, in order to verify earlier theoretical suggestions<sup>42</sup>, we look at the correlation  $\langle s_i \rangle(k)$  between the degree of innovators and the average size of vulnerable trees induced by them (Fig. 3e). We perform the last two measurements on the real data and in the model process for  $r = 0.6$  and  $0.9$ , as well as for the predicted value  $r = 0.73$ . In the case of  $\langle s_i \rangle(k)$ , we find a strong positive correlation in the data, explained partially by degree heterogeneities in the underlying social network, but surprisingly well emulated by the model. More importantly, although both quantities appear to scale with  $r$ , measures for the estimated  $r$  value fit the empirical data remarkably well, confirming our earlier validation based on the matching of relative component sizes (for further discussion see SI).

## Discussion

Although some products and innovations diffusing in society may cover a large fraction of the population, their spreading tends to follow slow cascading patterns, the dynamics of which have been modelled before by simple diffusion models like that of Bass<sup>18</sup>. However, this approach neglects threshold mechanisms that arguably drive the decision making of single individuals. On the other hand, threshold models study the conditions for cascades in global diffusion but do not address their temporal evolution, which is clearly a relevant factor in real-world adoption processes. These models are commonly used to predict rapid cascading patterns of adoption, which is a more realistic scenario for the spreading of information, opinions, or behavioural patterns but are not observed in the case of product or innovation diffusion where adoption requires additional efforts, e.g., free or paid registration. Here we provide a solution for this conundrum by analysing and modelling the worldwide spread of an online service in the techno-social communication network of Skype. Beyond the novel empirical evidence about heterogeneous adoption thresholds and non-linear dynamics of the adoption process, we identify two additional components necessary to introduce in the modelling of product adoption, namely: (a) a constant flow of innovators, which may induce rapid adoption cascades even if the system is initially out of the cascading regime; and (b) a fraction of immune nodes that forces the system into a quenched state where adoption slows down. These features are responsible for a critical structure of empirical adoption components that radically differs from previous theoretical expectations. We incorporate these mechanisms into a threshold model controlled by the rate of innovators and the fraction of immune nodes. The model contains several simplifying assumptions, e.g., about the functional form of social influence, the uniformity of nodes with the same degree, or about ignoring homophily in the adoption process, however, it is able to reproduce several pathways ranging from cascading behaviour to more realistic dynamics of innovation adoption. By constraining the model with empirically determined

parameters, we provide an estimate for the real fraction of susceptible agents in the social network of Skype, and validate this prediction through correlated structural features matching empirical observations.

Our aim in this study was to provide empirical observations as well as methods and tools to model the dynamics of social contagion phenomena with the hope it will foster thoughts for future research. One possible direction would be the observation of the reported structure and evolution of the global adoption cluster in other systems similar to the ones studied in<sup>26,28,29,34,36,57</sup>. Other promising directions could be the consideration of homophilic or assortative structural correlations, or the evolving nature of the underpinning social network with timely created and dissolved social ties (as studied in<sup>21</sup>), or the effects of interpersonal influence or leader-follower mechanisms on the social contagion process. Finally, we hope that the reported results may improve efficiency in the strategies of enhancing the diffusion of products and innovations, by shifting attention from the creation of short-lived perturbations to the sustenance of external input.

## Material and Methods

**Data description.** We use a static representation of the Skype social network aggregated over 99 months between September 2003 and November 2011. We follow the adoption of the “buy credit” paid service for 89 months starting from 2004, and the paid service “subscription” for 42 months starting from 2008 (for further details about the network and service see SI). By considering the online social structure and adoption times, we identify users as innovator, vulnerable, or stable nodes based on the number  $\Phi_k$  of adopting neighbours at the time of exposure. Thresholds are calculated as  $\phi = \Phi_k/k$  for users with  $k$  contacts. The adoption network is constructed by considering confirmed social links between users who adopted the service earlier than  $t$ . In order to avoid the effect of instantaneous group adoptions (evidently not driven by social influence), we only consider links between nodes who are neighbours in the underlying social network and whose adoption did not happen at the same time. Note that for the categorisation of nodes we use only the adoption time and the state of their peers, and thus real categories may differ slightly. For example, an innovator may appear as a vulnerable or stable node, even if its decision was not driven by social influence but some of its peers adopted earlier. To consider this bias we measure effective rates of adoption for the model process as well, just like for the empirical case (Fig. 1) and section S3.

**Maximum relative growth rate.** This measure is obtained by taking the maximum of the yearly adoption rate (yearly count of adoptions) normalised by the final observed adoption number of a given service. It characterises the maximum speed of adoption a service experienced during its history and takes values between 0 (no cascade) and 1 (instantaneous cascade). We repeat this measurement for the estimated number of registered users of Facebook, Twitter, and LinkedIn<sup>48</sup>, as well as for the number of active users of Skype and three paid Skype services. Adoption rates for Facebook, Twitter, and LinkedIn correspond to the period between 2006 and 2012, and for Skype and its services to the interval from release date until 2011.

**Empirical parameter estimation.** We use the Skype data to directly determine all model parameters, apart from the fraction  $r$  of immune nodes. To best approximate the degree distribution of the real network, after testing different candidate functions (see SI) we select a lognormal function  $P(k) = e^{-(\ln k - \mu_D)^2 / (2\sigma_D^2)} / (k\sigma_D\sqrt{2\pi})$  with parameters  $\mu_D = 1.2$  and  $\sigma_D = 1.39$  and minimum degree  $k_{\min} = 1$ , leading to the average degree  $z = 8.56$ . To account for finite-size effects in the model results for low  $N$  (Fig. 2), we decrease  $\mu_D$  slightly to obtain the same value of  $z$  as in the real network.

The threshold distribution of each degree group collapses to a master curve after normalisation by using the scaling relation  $P(\Phi_k, k) = kP(\Phi_k/k)$ . This master curve can be well approximated by the lognormal distribution  $P(\phi) = e^{-(\ln \phi - \mu_T)^2 / (2\sigma_T^2)} / (\phi\sigma_T\sqrt{2\pi})$ , with parameters  $\mu_T = -2$  and  $\sigma_T = 1$  as determined by the empirical average threshold  $w = 0.19$  and standard deviation 0.233 (for further details see SI). We estimate a rate of innovators  $p_n = 0.00019$  by fitting a constant function to  $R_i(t)$  for  $t > 2\tau$  (Fig. 1f). The fit to  $p_n$  also matches the time scale of a Monte Carlo iteration in the model to 1 month. Model results (Fig. 3d,e) are calculated with  $r = 0.73$  and  $p_n = 0.00019$ . Simulation results in Fig. 3c–e are averaged over 100 configuration-model networks of size  $N = 10^5$  ( $10^6$ ) after  $T = 89$  iterations, matching the length of the observation period in Skype.

**Model description.** We characterise the static social network by the extended distribution  $P(\mathbf{k})$ , where  $P(\mathbf{k}) = rP(k)$  for  $c = 0$  and  $P(\mathbf{k}) = (1 - r)P(k)P(c)$  for  $c > 0$ . Non-immune, susceptible nodes with property vector  $\mathbf{k}$  adopt spontaneously at a constant rate  $p_n$ , else they adopt only if a fraction  $\phi_c$  of their  $k$  neighbours have adopted before. These rules are condensed in the probability  $F_{\mathbf{k},m}dt$  that a node will adopt in a small time interval  $dt$ , given that  $m$  of its neighbours are already adopters,

$$F_{\mathbf{k},m} = \begin{cases} p_r & \text{if } m < k\phi_c \\ 1 & \text{if } m \geq k\phi_c \end{cases}, \quad \forall m \text{ and } k, c \neq 0, \quad (5)$$

with  $F_{(k,0),m} = 0 \forall k, m$  and  $F_{(0,c),0} = p_r \forall c \neq 0$  (for immune and isolated nodes, respectively). The dynamics of adoption is well described by an AME for the fraction  $s_{\mathbf{k},m}(t)$  of  $\mathbf{k}$ -nodes that are susceptible at time  $t$  and have  $m = 0, \dots, k$  adopting neighbours<sup>52,53,58</sup>,

$$\dot{s}_{\mathbf{k},m} = -F_{\mathbf{k},m}s_{\mathbf{k},m} - \beta_s(k - m)s_{\mathbf{k},m} + \beta_s(k - m + 1)s_{\mathbf{k},m-1}, \quad (6)$$

where  $\beta_s(t) = \frac{\sum_{\mathbf{k}} P(\mathbf{k}) \sum_m (k - m) F_{\mathbf{k},m} s_{\mathbf{k},m}(t)}{\sum_{\mathbf{k}} P(\mathbf{k}) \sum_m (k - m) s_{\mathbf{k},m}(t)}$ . To reduce the dimensionality of Eq. (6) we consider the ansatz  $s_{\mathbf{k},m} = B_{\mathbf{k},m}(\nu) e^{-\rho r t}$  for  $m < k\phi_c$ , leading to the condition  $\nu = \beta_s(1 - \nu)$ . With  $\rho = 1 - \sum_{\mathbf{k}} P(\mathbf{k}) \sum_m s_{\mathbf{k},m}$  and some algebra, this condition is reduced to Eq. (1) (see SI).



## References

- Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
- McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a Feather: Homophily in Social Networks. *Ann. Rev. Sociol.* **27**, 415–444 (2001).
- Toole, J. L., Cha, M. & González, M. C. Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE* **7**, e29528 (2012).
- Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).
- Rogers, E. M. *Diffusion of Innovations*. (Simon & Schuster), 5th edition (2003).
- Granovetter, M. Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).
- Schelling, T. C. Models of segregation. *Am. Econ. Rev.* **59**, 488–493 (1969).
- Axelrod, R. The dissemination of culture. *J. Conflict Resolut.* **41**, 203–226 (1997).
- Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA* **99**, 5766–5771 (2002).
- Handjani, S. Survival of threshold contact processes. *J. Theo. Probab.* **10**, 737–746 (1997).
- Valente, T. W. Social network thresholds in the diffusion of innovations. *Social Networks* **18**, 69–89 (1996).
- Watts, D. J. & Dodds, P. S. Influentials, networks, and public opinion formation. *J. Consum. Res.* **34**, 441–458 (2007).
- Melnik, S., Ward, J. A., Gleeson, J. P. & Porter, M. A. Multi-stage complex contagions. *Chaos* **23**, 013124 (2013).
- Gómez, V., Kappen, H. J. & Kaltenbrunner, A. Modeling the structure and evolution of discussion cascades. (HT'11, ACM, New York, NY, USA), pp. 181–190 (2010).
- Karampourniotis, P. D., Sreenivasan, S., Szymanski, B. K. & Korniss, Gy. The Impact of Heterogeneous Thresholds on Social Contagion with Multiple Initiators. *PLoS ONE* **10**(11), e0143020 (2015).
- Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734 (2007).
- Barrat A., Barthélemy, M. & Vespignani, V. *Dynamical Processes on Complex Networks*. (Cambridge University Press, 2008).
- Bass, F. M. A new product growth for model consumer durables. *Manage. Sci.* **15**, 215–227 (1969).
- Aral, S., Muchnik, L. & Sundararajana, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* **106**, 21544–21549 (2009).
- Shalizi, C. R. & Thomas, A. C. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociol Methods Res.* **40**, 211–239 (2011).
- Karsai, M., Iñiguez, G., Kaski, K. & Kertész, J. Complex contagion process in spreading of online innovation. *J. Roy. Soc. Interface* **11**, 20140694 (2014).
- Holt, C. A. *Markets, Games, Strategic Behavior*. (Addison Wesley, 2006).
- Bikhchandani, S., Hirshleifer D. & Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100**, 992–1026 (1992).
- Latané, B. The psychology of social impact. *Am. Psychol.* **36** 4, 343–356 (1981).
- González-Bailón, S., Borge-Holthoef, J., Rivero, A. & Moreno, Y. The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 197 (2011).
- Borge-Holthoef, J. *et al.* Structural and dynamical patterns on online social networks: The Spanish May 15th movement as a case study. *PLoS ONE* **6**, e23883 (2011).
- Ellis, C. J. & Fender, J. Information cascades and revolutionary regime transitions. *Econ. J.* **121**, 763–792 (2011).
- Dow, P. A., Adamic, L. A. & Friggeri, A. The anatomy of large Facebook cascades. (ICWSM, AAAI, Boston, MA, USA), pp. 145–154 (2013).
- Gruhl, D., Guha, R., Nowell, D. L. & Tomkins, A. Information diffusion through blogspace. (WWW '04, ACM, New York, NY, USA), pp. 491–501 (2004).
- Baños, R. A., Borge-Holthoef, J. & Moreno, Y. The role of hidden influentials in the diffusion of online information cascades. *EPJ Data Sci.* **2**, 6 (2013).
- Hale, H. E. Regime change cascades: What we have learned from the 1848 revolutions to the 2011 Arab uprisings. *Annu. Rev. Polit. Sci.* **16**, 331–353 (2013).
- Leskovec, J., Singh, A. & Kleinberg, J. Patterns of influence in a recommendation network. (PAKDD '06, Singapore), pp. 380–389 (2006).
- Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. (TWEB, ACM, New York, NY, USA), vol. 1, pp. 5 (2007).
- Goel, S., Watts, D. J. & Goldstein, D. G. The structure of online diffusion networks. (EC '12, ACM, New York, NY, USA), pp. 623–638 (2012).
- Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proc. Natl. Acad. Sci. USA* **107**, 5334–5338 (2009).
- Borge-Holthoef, J., Baños, R. A., González-Bailón, S. & Moreno, Y. Cascading behaviour in complex socio-technical networks. *J. Complex Net.* **1**, 1–22 (2013).
- Hackett, A. & Gleeson, J. P. Cascades on clique-based graphs. *Phys. Rev. E* **87**, 062801 (2013).
- Gleeson, J. P. Cascades on correlated and modular random networks. *Phys. Rev. E* **77**, 046117 (2008).
- Brummitt, C. D., D'Souza, R. M. & Leicht, E. A. Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci. USA* **109**, E680–E689 (2011).
- Ghosh, R. & Lerman, K. A framework for quantitative analysis of cascades on networks, WSDM '11. (WSDM '11, ACM, New York, NY, USA), pp. 665–674 (2010).
- Hurd, T. R. & Gleeson, J. P. On Watts' cascade model with random link weights. *J. Complex Net.* **1**, 25–43 (2013).
- Singh, P., Sreenivasan, S., Szymanski, B. K. & Korniss, Gy. Threshold-limited spreading in social networks with multiple initiators. *Sci. Rep.* **3**, 2330 (2013).
- Gleeson, J. P. & Cahalane, D. J. Seed size strongly affects cascades on random networks. *Phys. Rev. E* **75**, 050101(R) (2007).
- Yağan, O. & Gligor, V. Analysis of complex contagions in random multiplex networks. *Phys. Rev. E* **86**, 036103 (2012).
- Brummitt, C. D. & Kobayashi, T. Cascades in multiplex financial networks with debts of different seniority. *Phys. Rev. E* **91**, 062813 (2015).
- Karimi, F. & Holme, P. Threshold model of cascades in empirical temporal networks. *Physica A* **392**, 16 (2013).
- Backlund, V.-P., Saramäki, J. & Pan, R. K. Effects of temporal correlations on cascades: Threshold models on temporal networks. *Phys. Rev. E* **89**, 062815 (2014).
- White, D. S. Social Media Growth 2006 to 2012 (2013). Date of access: 2015.01.29.
- Morrissey, R. C., Goldman, N. D. & Kennedy, K. P. Skype S.A. United States Security Registration Statement, Amendment 3, Reg. No. 333-168646 (2011). Date of access: 2014.10.14.
- Ruan, Z., Iñiguez, G., Karsai, M. & Kertész, J. Kinetics of social contagion. *Phys. Rev. Lett.* **115**, 218702 (2015).
- Dodds, P. S. & Watts, D. J. Universal Behavior in a Generalized Model of Contagion. *Phys. Rev. Lett.* **92**, 218701 (2004).
- Gleeson, J. P. Binary-state dynamics on complex networks: Pair approximation and beyond. *Phys. Rev. X* **3**, 021004 (2013).
- Gleeson, J. P. High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.* **107**, 068701 (2011).
- Brummitt, C. D., Lee, K.-M. & Goh, K.-I. Multiplexity-facilitated cascades in networks. *Phys. Rev. E* **85**, 045102(R) (2012).

55. Lee, K.-M., Brummitt, C. D. & Goh, K.-I. Threshold cascades with response heterogeneity in multiplex networks. *Phys. Rev. E* **90**, 062816 (2014).
56. Kobayashi, T. Trend-driven information cascades on random networks. *Phys. Rev. E* **92**, 062823 (2015).
57. Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J. Everyone's an influencer: Quantifying influence on Twitter. (WSDM '11, ACM, New York, NY, USA), pp. 65–74 (2011).
58. Porter, M. A. & Gleeson, J. P. Dynamical systems on networks: A tutorial. Eprint arXiv 1403.7663 (2014).

### Acknowledgements

The authors gratefully acknowledge the support of M. Dumas, A. Saabas, and A. Dumitras from STACC and Microsoft/Skype Labs as well as constructive comments by J. Saramäki and T. Näsi. G.I. acknowledges the Academy of Finland, J.K. the CIMPLEX FET Open H2020 EU project, and M.K. the SoSweet (ANR-15-CE38-0011-01) and CODDDE (ANR-13-CORD-0017-01) ANR projects for support. This research was partly funded by Microsoft/Skype Labs.

### Author Contributions

M.K., G.I., R.K., K.K. and J.K. designed the research and participated in writing the manuscript. R.K. and M.K. analysed the empirical data, G.I. made the analytical calculations, and M.K. and G.I. performed the numerical simulations.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Karsai, M. *et al.* Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading. *Sci. Rep.* **6**, 27178; doi: 10.1038/srep27178 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>