SCIENTIFIC REPORTS

Competition between global and local online social networks

Kaj-Kolja Kleineberg & Marián Boguñá

The overwhelming success of online social networks, the key actors in the Web 2.0 cosmos, has reshaped human interactions globally. To help understand the fundamental mechanisms which determine the fate of online social networks at the system level, we describe the digital world as a complex ecosystem of interacting networks. In this paper, we study the impact of heterogeneity in network fitnesses on the competition between an international network, such as Facebook, and local services. The higher fitness of international networks is induced by their ability to attract users from all over the world, which can then establish social interactions without the limitations of local networks. In other words, inter-country social ties lead to increased fitness of the international network. To study the competition between an international network and local ones, we construct a 1:1000 scale model of the digital world, consisting of the 80 countries with the most Internet users. Under certain conditions, this leads to the extinction of local networks; whereas under different conditions, local networks can persist and even dominate completely. In particular, our model suggests that, with the parameters that best reproduce the empirical overtake of Facebook, this overtake could have not taken place with a significant probability.

The rapid growth of online social networks (OSNs), such as Twitter or Facebook, led to over two billion accounts being active in 2014[1], connecting over one quarter of the world's population and 72% of online U.S. adults[2]. Bridging the gap between social sciences, and information and communication technologies, OSNs constitute a crucial building block in the development of innovative approaches to the challenges society faces today. However, technological progress over the last decade has dramatically outpaced our understanding of the new systems and their impact on society.

This lack of understanding of the complex dynamics of the digital world reveals the pressing need for a comprehensive and concise theory to describe and model the online world as a set of interacting networks. In this context, the activity of users has become a resource to be fought over and which drives competition in the digital world. Digital services only persist if they can attract and maintain users' attention. Here, we describe the online world as a complex, digital ecosystem in which interacting networks constitute species in competition for survival. Within this ecological context, the extinction of a network corresponds to the complete absence of activity, as an entirely passive network cannot function and, more importantly, will not attract new users. In a recent study[3], we demonstrated that a moderate number of identical networks can coexist in the digital ecosystem, in contrast to the principle of competitive exclusion[4].

In contrast to that previous work, in this paper we address the heterogeneity of networks. Networks can differ in functionality, features, and –most importantly– they can address different peer groups. Here, we show how the effect of different overlapping peer groups can be described in terms of different degrees of network fitness. We find that under certain conditions, the heterogeneity of degrees of fitness can impede coexistence which would indeed be possible for identical networks. This effect is particularly important for the competition between local networks and an international network. Unlike users of local networks, users of the international network have the possibility to interact with people in other countries, providing this network an advantage over local ones, similar to a higher fitness of a certain species. A proper modeling of this effect requires taking into account the network of interactions among countries in the world, which results in a highly complex and non-linear dynamical system made of interconnected multilayer networked entities – we are hence dealing with networks of multiplex networks[5–8], in contrast to ref. 3. Besides, inter-country interactions induce a different type of bifurcation as the symmetry of the previous model is broken, which constitutes a fundamentally new behavior not observed in[3].

Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain. Correspondence and requests for materials should be addressed to K.-K.K. (email: kkl@ffn.ub.edu) or M.B. (email: marian.boguna@ub.edu)
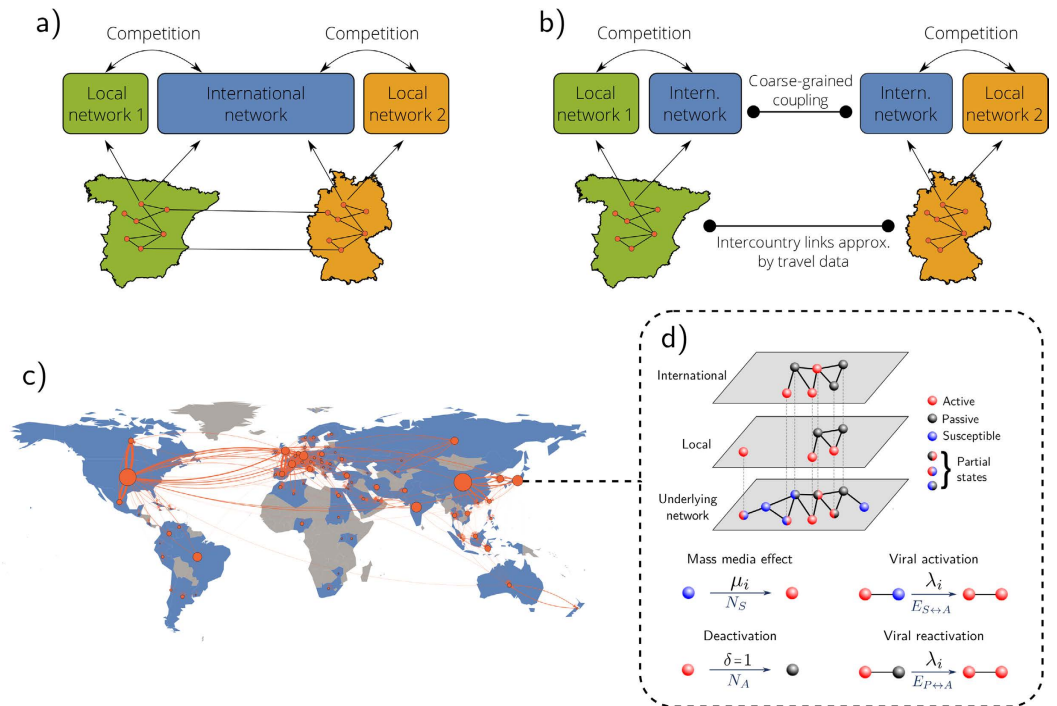
Figure 1. Constituents of our model. (**a**) Design of the international network and local networks. (**b**) Sketch of our model using coarse-grained coupling. (**c**) Visualization of the flight network. The size of the nodes is proportional to the number of users in the respective countries with Internet access. The transparency and thickness of the links represents the density of passengers between the countries concerned. (**d**) Illustration of the competition between the international network and the local network within one country. *Maps created with Mathematica, Version 9,* https://www.wolfram.com/mathematica/.

Empirical observations have shown that Facebook expanded massively in the middle of the first decade of this century, starting in the US, when local networks were the most popular services in most countries. Only a few years later, Facebook had become the most popular network in most countries. So, is the fate of the digital world to become dominated by a single "big brother" as it takes over all our digital interactions? Alternatively, is digital diversity possible from a system-level perspective? In this paper, we show that due to the nonlinear character of our model, the answer to both questions can be positive or negative depending on a range of parameters and, quite surprisingly, depending on chance. As we will show, our model, despite its simplicity and the limited number of parameters, is able to describe surprisingly well the complex behavior of globally interacting online social networks.

Results

Complex organization of the digital world.
The digital world consists of highly connected and strongly coupled interacting subsystems. These basic building blocks are single networks, each of which obeys specific dynamics in the absence of coupling to the whole system. So the complexity of the digital world is a consequence of both the dynamics of networks in isolated environments and the interactions between many such networks. Finally, not all of these building blocks are identical. Instead, different networks address different peer groups or have different functionalities. Hence, to reveal the fundamental mechanisms that determine the fate of the digital world, it is necessary to understand the interaction of heterogeneous networks, each driven by intrinsic dynamics.

Isolated dynamics of online social networks.
The key actors in the digital world are OSNs; loosely defined as web-based platforms that enable digital social interactions over the Internet. However, societies were organized as networks long before OSNs were even thought of. From this point of view, the growth of OSNs can be described through the dynamical processes by which people in the traditional offline social structure come to engage in OSNs. The topology of the OSN is now the digital counterpart of the underlying offline social network[9,10].

In isolation, this process of formation can be described by a set of simple dynamical mechanisms[9]. The system is initially given by an empty OSN and the underlying social structure. Individuals can be in three different states: active, passive or susceptible. While active and passive nodes exist in the online as well as the offline networks, susceptible nodes are only present in the latter. A susceptible node can join the OSN via two different mechanisms: a *viral activation* effect which means that a susceptible node becomes active due to the influence of an active neighbor in the offline network and a *mass media* effect, which represents the spontaneous activation of a susceptible node. In addition, an active node can become passive spontaneously (*deactivation*) and a passive node can become active again due to the influence of an active neighbor (*viral reactivation*). See Fig. 1d for a

visualization of these mechanisms. Notice that, in the long time limit, when the number of susceptible individuals is basically zero, this dynamics is equivalent to the susceptible-infected-susceptible model widely used in epidemiology[11]. As found in[9,12], this implies that online social networks can either exhibit a sustained activity (similar to the endemic phase), or they can become entirely passive (similar to the healthy phase).

The evolution of OSNs rarely takes place in isolation. Nevertheless, we found a perfect case study in the Slovakian social network "Pokec", which, due to the particularities of the country, has been growing in quasi-isolation for more than ten years. By analyzing the evolution of the topology of the social contact graph of "Pokec", we were able to rigorously validate our model. Quite remarkably, with only two parameters, the model reproduced the entire topological evolution with astonishing precision.

Viral activation and *viral reactivation* occur at the same rate, λ, and the ratio between this rate and the rate of *mass media* influence, μ, governs the topological evolution. In particular, we observed that the real system underwent a dynamical percolation transition; that is, a phase transition between a disconnected phase and a phase in which a macroscopic fraction of the system is connected. The position of this transition is controlled by λ/μ, due to the complementary roles that the *viral activation* and *mass media* effects play in the topological evolution of the network (the former tends to connect components; whereas the latter tends to create new components). Finally, without loss of generality, we set the *deactivation* rate δ to 1 which is equivalent of fixing the timescale of the model.

To sum up, in our previous work[9], we were able to rigorously validate the dynamics ruling OSNs in isolation; the fundamental building blocks of the digital world. These findings constitute the foundation for the development of a more comprehensive theory of interacting heterogeneous networks.

Competitive interaction between multiple networks. The simultaneous existence of multiple digital services in competition for the attention of users suggests an ecological perspective from which to explain the prevalence of a given network or the coexistence of multiple networks. In ecology theory, the principle of competitive exclusion[4] states that multiple species in competition for the same resource cannot coexist, as even the slightest advantage of one species over the other is successively amplified; a mechanism referred to as "rich get richer" or preferential attachment[13–21]. This eventually leads to the extinction of the inferior species.

The key principle that drives the competition between OSNs is the fact that, due to the physical and cognitive limitations of users, the time they devote to online activities is limited. As a consequence, the viral parameter, λ, constitutes a conserved quantity that is nevertheless distributed between the competing networks as $\lambda_i = \omega_i(\boldsymbol{\rho}^a)\lambda$, where $\omega_i(\boldsymbol{\rho}^a)$ represents a normalized set of weights, that is $\sum_i \omega_i(\boldsymbol{\rho}^a) = 1$, and $\boldsymbol{\rho}^a \equiv (\rho_1^a, \rho_2^a, \ldots)$ is a vector denoting the fraction of active nodes (activities) in the different networks. In general, users are more likely to subscribe to and engage in networks that are more active. Therefore, the viral activity of each network must be a function of the activity of the network itself. In particular, we model this by assuming that the weighting $\omega_i(\boldsymbol{\rho}^a)$ is a function, such that $\partial \omega_i(\boldsymbol{\rho}^a)/\partial \rho_i^a > 0$. In[3] we proposed the particular form:

$$\omega_i(\boldsymbol{\rho}^a) = \frac{[\rho_i^a]^\sigma}{\sum_{j=1}^{n_l} [\rho_j^a]^\sigma},$$

(1)

were n_l denotes the number of networks. This choice allows us to interpolate between a set of independent networks ($\sigma \ll 1$) and highly coupled ones ($\sigma \gg 1$). The activity affinity parameter, σ, then quantifies the tendency of users to subscribe to or engage in more active networks. Interestingly, in contrast to the principle of competitive exclusion, multiple networks can coexist because the "rich get richer" mechanism is damped by the diminishing returns of the dynamics of network evolution. For details, we refer the reader to ref. 3. In the following, we take into account the heterogeneity of networks induced by different groups of individuals that can subscribe to the different networks. These aspects are not discussed in ref. 3 and have important implications and applications, as we will show.

Network heterogeneity leads to effective activity. As mentioned above, since its official launch in 2004, Facebook has become the most popular OSN in most countries; even in countries where there was already a popular OSN before Facebook was launched. To mimic the real evolution of the digital ecosystem at the worldwide scale, we assume that one local network exists in each country in addition to a globally operating, international network (see Fig. 1a). In the US, both networks are launched at the same time; whereas the international network is launched with a delay Δt in the remaining countries, to take into account the initial prevalence of local networks.

Once launched, the international network provides the user with the possibility to connect to individuals in different countries, in contrast to local networks, making it more attractive to users. For a given country, the advantage of the international network is directly related to the abundance of social ties between that country and the rest of the world. We use passenger air travel data as a proxy for the abundance of such ties. This choice is justified by the strong correlation between air travel flows and further measures of inter-country exchange, for instance email communication[22] or Twitter activity[23].

Users in country i experience the greater attractiveness of the international network as they perceive its activity with respect to the population of their own country and also with respect to their contacts in other countries. To account for this on a coarse grained level, in Eq. (1), we replace the activity of the international network by an effective activity as follows

$$\tilde{\rho}_{i,\text{int}}^a = \rho_{i,\text{int}}^a + \alpha \sum_j \Omega_{ij} \rho_{j,\text{int}}^a,$$

(2)

where

$$\Omega_{ij} = \frac{W_{ij}/N_i}{\max\ [W_{ij}/N_i]} \tag{3}$$

denotes the fraction of the number of air travel passengers between countries i and j, W_{ij}; and N_i, the population of country i. Notice that, in an ecological context, this corresponds to increased fitness of the international network. In Eq. (2), we have implicitly assumed proportionality between the number of passengers and the number of contacts in the respective countries, namely $N_{ij} \propto W_{ij}$. Finally, note that the arbitrary normalization in Eq. (3) serves the sole purpose of ensuring that reasonable values for the parameter α are of the order of unity.

Hereafter, we decompose the international network into a set of disjunct coupled subnetworks operating in each country and in competition with the respective local network (see Fig. 1b). These subnetworks are nevertheless not independent, as they are globally coupled via the effective activity defined in Eq. (2) and ultimately by the network representing the inter-country social ties. Hence, our model forms a network of networks[5–8], where each node in Fig. 1c represents a three-layer multiplex network[24,25] in which the bottom layer corresponds to the underlying social structure and the two upper layers denote the local and international networks operating in the respective country (see Fig. 1d).

Double meanfield approximation reveals complex role of the activity affinity.
To understand the qualitative behavior of the system, in this section we present a double meanfield approximation of the system. This reduces the system given by a network of networks to a set of evolution equations of the average activity in the international network and in local networks. As we show in the following section, the results of the full model with heterogeneous topologies exhibits similar behavior to that encountered by the double meanfield approximation.

The first meanfield approximation consists of assuming a fully mixed homogeneous population in each country. Let $\rho_{i,l}^{\rm a}$ denote the fraction of active users in network $l \in (\mathrm{loc},\ \mathrm{int})$ in country i and $\rho_{i,l}^{\rm s}$ the fraction of nodes susceptible to joining this network. Then, the fraction of passive users is given by $1 - \rho_{i,l}^{\rm s} - \rho_{i,l}^{\rm a}$. As explained above, in each country the virality is distributed between the local and international network via the weight functions $\omega_{\rm loc}(\rho_{i,\mathrm{loc}}^{\rm a}, \tilde{\rho}_{i,\mathrm{int}}^{\rm a})$ and $\omega_{\rm int}(\rho_{i,\mathrm{loc}}^{\rm a}, \tilde{\rho}_{i,\mathrm{int}}^{\rm a}) = 1 - \omega_{\rm loc}(\rho_{i,\mathrm{loc}}^{\rm a}, \tilde{\rho}_{i,\mathrm{int}}^{\rm a})$, as introduced in Eq. (1). Here, $\tilde{\rho}_{i,\mathrm{int}}^{\rm a}$ denotes the effective activity of the international network as defined in Eq. (2). The evolution equations of the resulting system represent a generalization of the evolution equations for identical networks which we derived in[3], where one replaces the activity of the international network with the effective activity from Eq. (2). This procedure yields:

$$\begin{aligned}
\dot{\rho}_{i,l}^{\rm a} &= \rho_{i,l}^{\rm a}\{\lambda\langle k\rangle\omega_l(\rho_{i,\mathrm{loc}}^{\rm a}, \tilde{\rho}_{i,\mathrm{int}}^{\rm a})[1 - \rho_{i,l}^{\rm a}] - 1\} + \mu_{i,l}\rho_{i,l}^{\rm s} \\
\dot{\rho}_{i,l}^{\rm s} &= -\mu_{i,l}\rho_{i,l}^{\rm s}\{1 + \nu\langle k\rangle\rho_{i,l}^{\rm a}\}.
\end{aligned} \tag{4}$$

As shown in[3,9], we further assume the the same linear relationship between virality and media influence in each country, that is:

$$\begin{aligned}
\mu_{i,\mathrm{loc}} &= \frac{\lambda\omega_{\rm loc}(\rho_{i,\mathrm{loc}}^{\rm a}, \tilde{\rho}_{i,\mathrm{int}}^{\rm a})}{\nu} \\
\mu_{i,\mathrm{int}} &= \frac{\lambda\omega_{\rm int}(\rho_{i,\mathrm{loc}}^{\rm a}, \tilde{\rho}_{i,\mathrm{int}}^{\rm a})}{\nu}.
\end{aligned} \tag{5}$$

As shown in[3], the value of ν does not affect the stability of the system. In what follows, we perform the stability analysis in the limit $\nu \to \infty$. This decouples the evolution of $\rho_{i,l}^{\rm a}$ from $\rho_{i,l}^{\rm s}$, so that we only have to consider $\rho_{i,l}^{\rm a}$. Plugging in the weights function defined in Eq. (1) and the effective activity from Eq. (2) yields the evolution equations for the activities of the local and international networks in country i:

$$\begin{aligned}
\dot{\rho}_{i,\mathrm{loc}}^{\rm a} &= \rho_{i,\mathrm{loc}}^{\rm a}\left[\frac{\lambda\langle k\rangle\left[\rho_{i,\mathrm{loc}}^{\rm a}\right]^\sigma}{\left[\rho_{i,\mathrm{loc}}^{\rm a}\right]^\sigma + (\rho_{i,\mathrm{int}}^{\rm a} + \delta_i)^\sigma}\left[1 - \rho_{i,\mathrm{loc}}^{\rm a}\right] - 1\right] \\
\dot{\rho}_{i,\mathrm{int}}^{\rm a} &= \rho_{i,\mathrm{int}}^{\rm a}\left[\frac{\lambda\langle k\rangle(\rho_{i,\mathrm{int}}^{\rm a} + \delta_i)^\sigma}{\left[\rho_{i,\mathrm{loc}}^{\rm a}\right]^\sigma + (\rho_{i,\mathrm{int}}^{\rm a} + \delta_i)^\sigma}\left[1 - \rho_{i,\mathrm{int}}^{\rm a}\right] - 1\right],
\end{aligned} \tag{6}$$

where $\delta_i = \alpha\sum_j \Omega_{ij}\rho_{j,\mathrm{int}}^{\rm a}$.

The second meanfield approximation consists of applying the hypothesis of a fully mixed homogeneous network for the inter-country social ties. We use $\overline{\Omega} = \alpha\langle\Omega_{ij}\rangle$ and define the mean activity of the local networks as $x \equiv \langle\rho_{i,\mathrm{loc}}^{\rm a}\rangle$ and the mean activity of the international network as $y \equiv \langle\rho_{i,\mathrm{int}}^{\rm a}\rangle$. Finally, our double meanfield approximation leads to the following system of coupled differential equations
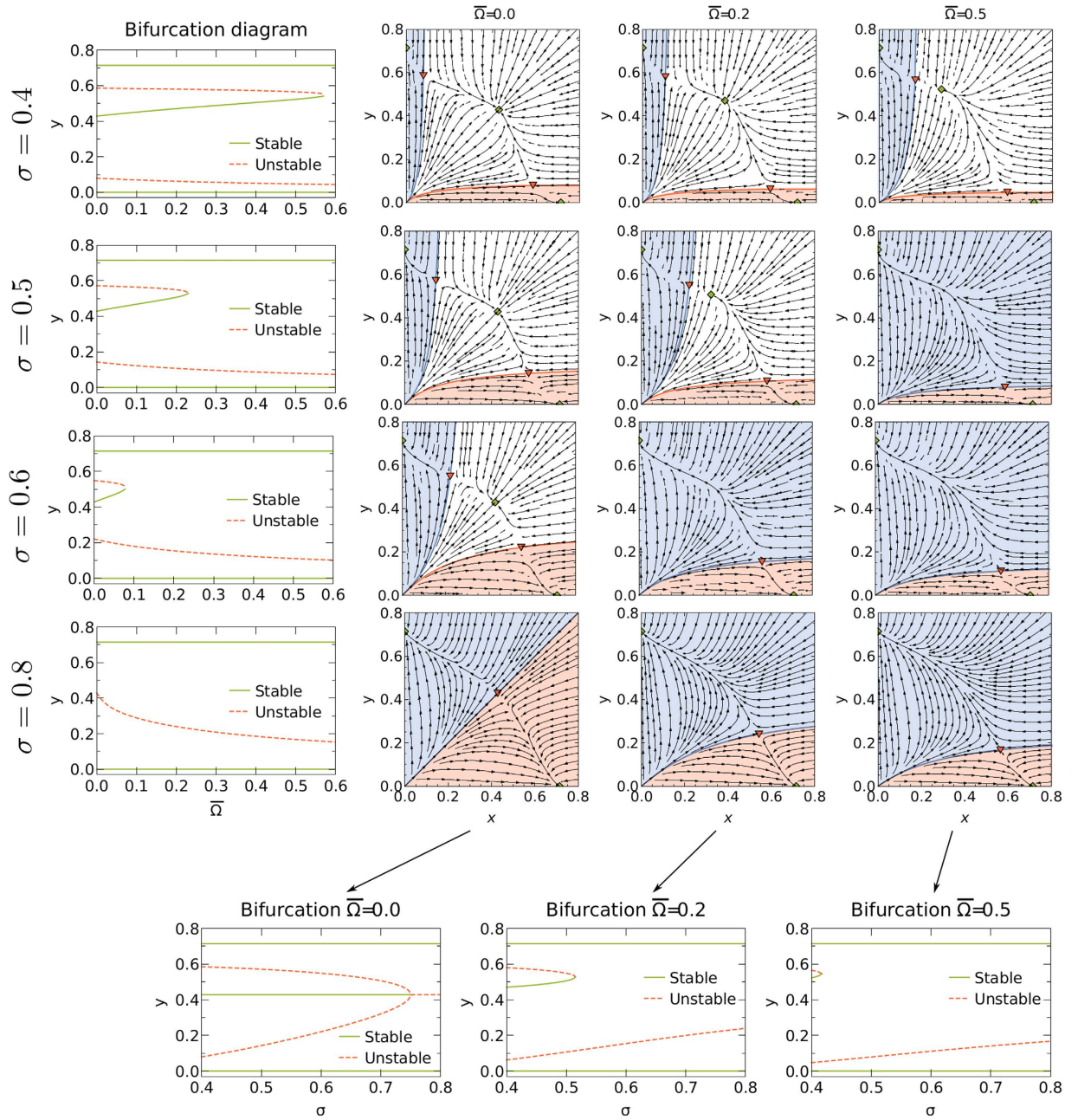
Figure 2. Bifurcation diagram and stream plots for the double meanfield approximation (7) for $\lambda\langle\kappa\rangle = 3.5$ and $\nu \to \infty$. The basins of attraction for the domination of the international network are in blue, the basins of attraction for the domination of local networks are in red; the white areas correspond to the basins of attraction of the coexistence solution (if it exists).

$$\dot{x} = x\left[\lambda\langle k\rangle\frac{x^{\sigma}}{x^{\sigma} + (y(1 + \overline{\Omega}))^{\sigma}}[1 - x] - 1\right]$$

$$\dot{y} = y\left[\lambda\langle k\rangle\frac{(y(1 + \overline{\Omega}))^{\sigma}}{x^{\sigma} + (y(1 + \overline{\Omega}))^{\sigma}}[1 - y] - 1\right],$$

(7)

which has three relevant parameters: $\lambda\langle k\rangle$, σ, and $\overline{\Omega}$. Note that by setting $\overline{\Omega}$ to zero, we recover the equations for identical networks presented in[3].

In what follows, we discuss the dynamical properties of the system given by Eq. (7). For constant σ, the system exhibits a saddle-node bifurcation at a critical value of the global connectivity $\overline{\Omega}_c(\sigma)$ (see Fig. 2). Above this point, coexistence is not possible and the only stable solutions correspond to the domination of either local networks or the international one. Both above and below the critical value $\overline{\Omega}_c(\sigma)$, the basin of attraction of the solution corre-
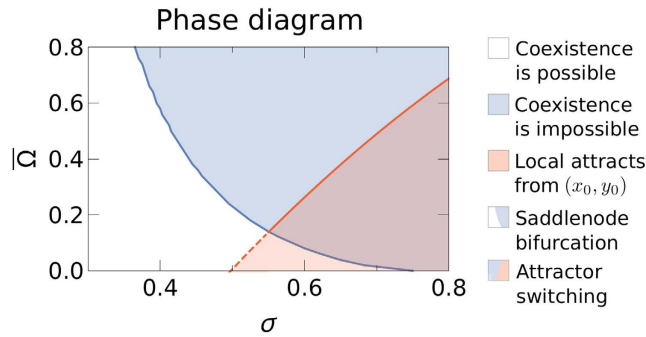
Figure 3. Phase diagram of the double meanfield approximation for $\lambda \langle k \rangle = 3.5$. The white area denotes the parameters for which coexistence is possible. The blue area denotes the parameters for which only domination of either local or the international network can occur. At the blue line, the system undergoes a saddle-node bifurcation in which the stable coexistence solution disappears. The red line denotes the combination of parameters for which the system switches attractors for the initial conditions given by Eq. (8), where we use $\beta = 0.2$ to reflect that the US contributes about 20% of the population. The red region shows the parameters for which the solution of domination of local networks is reached for these initial conditions. Above the red line, the system approaches the domination of the international network (solid red line) or the coexistence solution (dashed red line).

sponding to the domination of local networks decreases with $\overline{\Omega}$, whereas that of the international network increases (see the rows of Fig. 2). Furthermore, at the critical point, the basin of attraction of the international network is amplified discontinuously as the region of coexistence in the subcritical regime is now merged with the basin of attraction of the domination of the international network.

For constant $\overline{\Omega} > 0$, the system also exhibits a saddlenode bifurcation at a critical value of the activity affinity $\sigma_c(\overline{\Omega})$. In[3] we showed that the system undergoes a subcritical pitchfork bifurcation with respect to the control parameter σ, above which no stable coexistence is possible. $\overline{\Omega} > 0$ breaks the symmetry of the pitchfork bifurcation and in this case the system undergoes a saddlenode bifurcation with respect to σ instead (see bottom of Fig. 2). This behavior is well known in bifurcation theory and results from adding a small error term to the normal form of the pitchfork bifurcation (see Supplementary Materials). The evolution of the basins of attraction is more complex compared to the previous case. Below the critical point, both basins of attraction increase with σ. Above the critical point $\sigma_c(\overline{\Omega})$, the basin of attraction of the local network increases whereas the basin of attraction of the international network decreases with σ (see the columns of Fig. 2). This is particularly interesting as it implies that an intermediate value of the activity affinity just slightly above the critical point $\sigma \gtrsim \sigma_c(\overline{\Omega})$ represents the worst scenario for the survival of local networks, since at this point the size of the basin of attraction of the domination of the international network is maximum.

In Fig. 3, the blue line indicates the critical line $\overline{\Omega}_c(\sigma)$ in the σ-$\overline{\Omega}$ plane, which separates a phase in the parameter space where coexistence is possible (white region) and one in which only domination can occur (blue region). However, the increasing size of the basin of attraction of the domination of local networks above the critical point with respect to $\sigma > \sigma_c(\overline{\Omega})$ can dramatically alter the fate of the system for a given set of initial conditions. Assume, for instance, that the international network dominates in the US and starts with a significant delay in each other country, which causes the local networks to dominate in those countries. At the time when the international network is launched globally, the state of the system can be approximated as follows

$$
\begin{aligned}
x_0 &= (1 - \beta)\left[1 - \frac{1}{\lambda \langle k \rangle}\right] \\
y_0 &= \beta\left[1 - \frac{1}{\lambda \langle k \rangle}\right],
\end{aligned}
\tag{8}
$$

which we now use as initial conditions to study the further evolution. Notice that if one network dominates in country i, its activity is given by $\rho_{i,l}^{a} = 1 - \frac{1}{\lambda \langle k \rangle}$. Hence, the initial conditions given by Eq. (8) reflect the fact that local networks dominate in the fraction $(1 - \beta)$ of the system and the international one dominates in the remainder. The evolution of the basins of attraction makes the system approach different stationary solutions from these initial conditions for different parameters. Below the red line in Fig. 3, the system approaches the domination of local networks starting from the initial conditions given in Eq. (8). Above this line, the system either approaches coexistence (white area; crossing dashed red line) or domination of the international network (blue area; crossing solid red line). This means that in the red region, when the international network is launched globally, it is not able to overcome the initial advantage of the local networks due to its earlier launch.

To conclude, the double meanfield approximation predicts that intermediate values of the activity affinity most favor the international network; whereas the local networks can dominate for a high activity affinity and low global connectivity. We confirm these findings by numerical simulations in the following section.

China	253	UnitedStates	231	Japan	90.9	India	81	Brazil	64.9
Germany	62	UnitedKingdom	48.8	Russia	45.2	France	42.9	SouthKorea	37.5
Indonesia	30	Spain	25.2	Canada	25.1	Italy	25	Turkey	24.5
Mexico	23.3	Iran	23	Vietnam	20.8	Poland	18.7	Pakistan	18.5
Colombia	17.1	Malaysia	16.9	Thailand	16.1	Australia	15.2	Taiwan	15.1
Netherlands	14.3	Egypt	11.4	Argentina	11.2	Nigeria	11	Ukraine	10.4
Morocco	10.3	Sweden	8.1	SaudiArabia	7.7	Belgium	7.3	Venezuela	7.2
Peru	7.1	Romania	6.1	CzechRepublic	6	Austria	5.9	Hungary	5.9
Switzerland	5.7	Philippines	5.6	Chile	5.5	Denmark	4.6	Portugal	4.5
Finland	4.4	Greece	4.3	Sudan	4.2	SouthAfrica	4.2	HongKong	4.1
Algeria	4.1	Norway	3.9	Slovakia	3.6	Syria	3.6	Singapore	3.4
Kenya	3.4	Belarus	3.1	NewZealand	3	Serbia	2.9	UnitedArabEmirates	2.9
Ireland	2.8	Tunisia	2.8	Bulgaria	2.6	Uganda	2.5	Uzbekistan	2.5
Kazakhstan	2.3	Lebanon	2.2	DominicanRepublic	2.1	Israel	2.1	Guatemala	2
Croatia	1.9	Lithuania	1.8	Jamaica	1.5	Jordan	1.5	Azerbaijan	1.5
CostaRica	1.5	Cuba	1.4	Zimbabwe	1.4	Uruguay	1.3	Ecuador	1.3

Table 1. List of countries and estimated number of Internet users ($\times 10^6$) according to Wolfram Alpha database.

Numerical simulations and synthetic networks. In this section, we go beyond the meanfield approximation and study, by means of numerical simulations, the effects of the real topology of inter-country social ties and of underlying social structures. To this end, we use the air travel network (see Fig. 1c and Materials and Methods) as a proxy for inter-country social ties and construct 1:1000 scaled synthetic networks to model the structure of the 80 countries with most Internet users (see Table 1). To generate these networks, we make use of a model introduced in[26–28], which produces realistic topologies of the traditional offline social networks, including heterogeneous node degrees and a high level of clustering (see Materials and Methods).

Figure 4 shows results from our model for the set of parameters that best matches empirical observations, as explained in the following section. The international network starts with a delay in all countries except the US; so that initially in these countries the respective local network dominates. After some time, the international network obtains a significant advantage and quickly takes over in most countries.

To further study the properties of the model presented here, we define the relative prevalence of the international network compared to local networks as:

$$\Phi = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{\rho_{i,\text{int}}^{\text{a}}\big|_{\text{st}}}{\rho_{i,\text{int}}^{\text{a}}\big|_{\text{st}} + \rho_{i,\text{loc}}^{\text{a}}\big|_{\text{st}}}$$

(9)

where n_c denotes the number of countries, and $\rho_{i,\text{int}}^{\text{a}}\big|_{\text{st}}$ and $\rho_{i,\text{loc}}^{\text{a}}\big|_{\text{st}}$ are the activities of the international and local networks in country i in the stationary state. With this definition, a value of $\Phi \approx 0$ implies that local networks dominate in most countries, whereas $\Phi \approx 1$ corresponds to the domination of the international network. The relative prevalence of the international network averaged over many realizations is shown in Fig. 5 for different values of α as a function of the activity affinity, σ, and the launch time delay, Δt. For small values of Δt, we observe that when σ is small, the international and local networks coexist and we observe values around $\Phi \approx 0.5$ for the relative prevalence; then, increasing σ favors the international network, which dominates for values of $\sigma \gtrsim 0.5$ (see Fig. 6a,b). For larger values of Δt, this behavior smoothly translates into a more complex case, which we discuss below.

We observe in Fig. 5 that for launch time delays $\Delta t \geq 2$, the actual length of the delay becomes irrelevant. This behavior corresponds to the limit of saturation of the evolution of local networks before the international OSN is launched; as discussed in the previous section. We consider this limit by averaging over regions with $\Delta t \geq 2$ in Fig. 6c, which yields a two dimensional parameter space σ-α. Indeed, numerical simulations of the full model confirm the results from the meanfield analysis; in particular the complex role of the activity affinity σ. For small α and σ, local networks and the international OSN can coexist. Increasing σ or α favors the domination of the international network, which gives rise to the blue "V"-shaped region around $\sigma = 0.5$. This corroborates the saddlenode bifurcation predicted by the double meanfield approximation. See supplementary video for an explicit realization. For high values of σ and small values of α (red region in the bottom right-hand corner of Fig. 6), local networks dominate. Note that partial states are also possible, in which the international network dominates in some countries and local networks dominate in the remaining countries. See supplementary video for an explicit realization of this case.

Between the regions of domination of the international network and of local networks, there is a region in which the final fate of the system varies significantly between different realizations of the model ("coinflip region"). In this region, if the international network wins initially in the US, it will become dominant globally; otherwise, local networks maintain their initial prevalence. Although in this region the prevalence of the international network averaged over many realizations is about 0.5, as in the coexistence region in the bottom left-hand
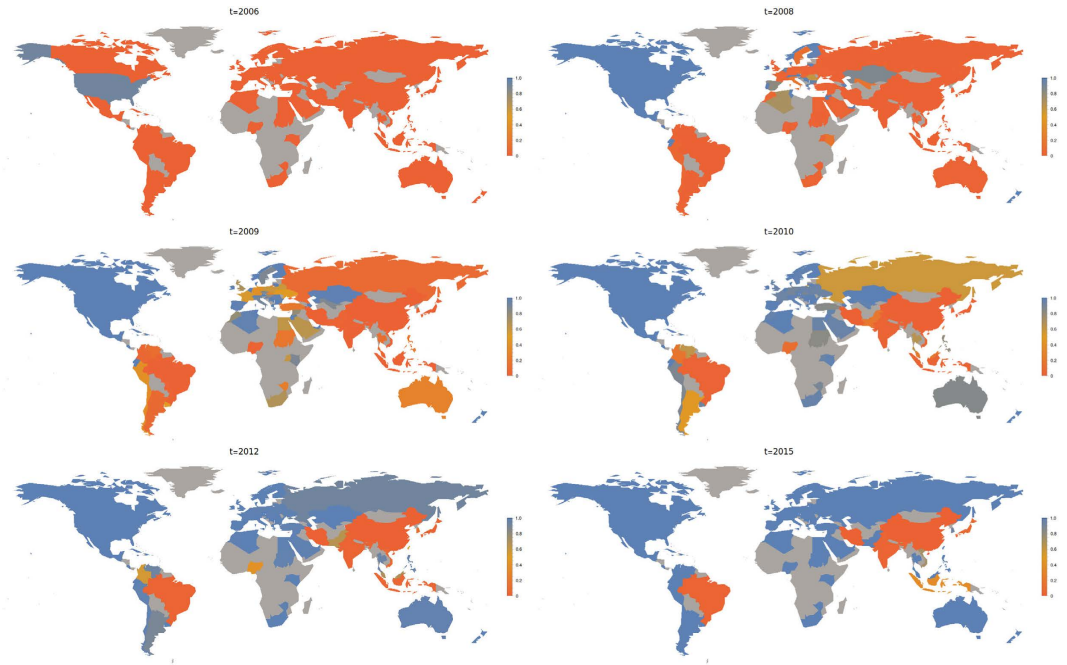
Figure 4. Evolution of our model averaged over several realizations for the parameters $\sigma = 1.25$, $\Delta t = 3$, and $\alpha = 2$ (here, we excluded realizations in which local networks dominate, which occurs with approximately 30% probability for these parameters). The mapping from model time to real time is explained in the following section. The relative importance of the influence of mass media compared to the viral spreading mechanism in governed by the parameter ν introduced in Eq. (5). In all numerical simulations, we set $\nu = 4$: the value found empirically in[9]; and $\lambda = 0.2$ (this corresponds to $\lambda / \lambda_c^1 \approx 4.3$ in[3]). The relative prevalence of the international network, given by $\rho_{i,\text{int}}^a / (\rho_{i,\text{int}}^a + \rho_{i,\text{loc}}^a)$, is color coded. We consider the international network to be banned in China and Iran. To model this, we set the values of $\Omega_{ij} = 0$ for each entry which involves one of these countries. This is equivalent to assuming that in these countries two local networks compete without any coupling to the rest of the world. *Maps created with Mathematica, Version 9,* https://www.wolfram.com/mathematica/.
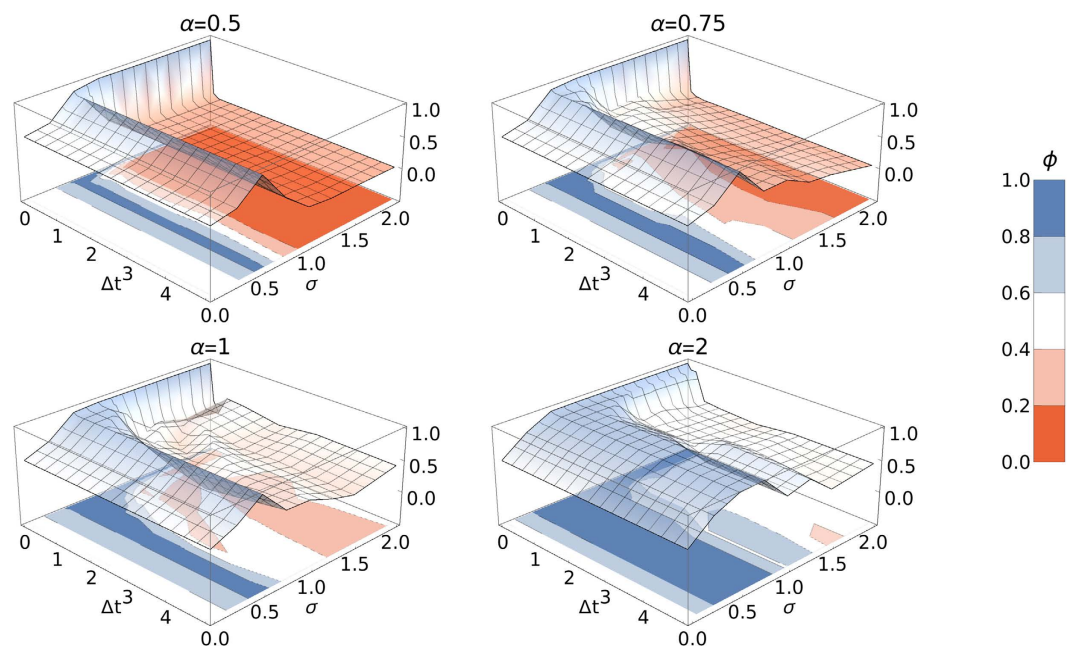

Figure 5. Relative prevalence Φ of the international network is plotted on the z axis as a function of the launch time delay Δt and the coupling strength σ. Averaged over 30 realizations.
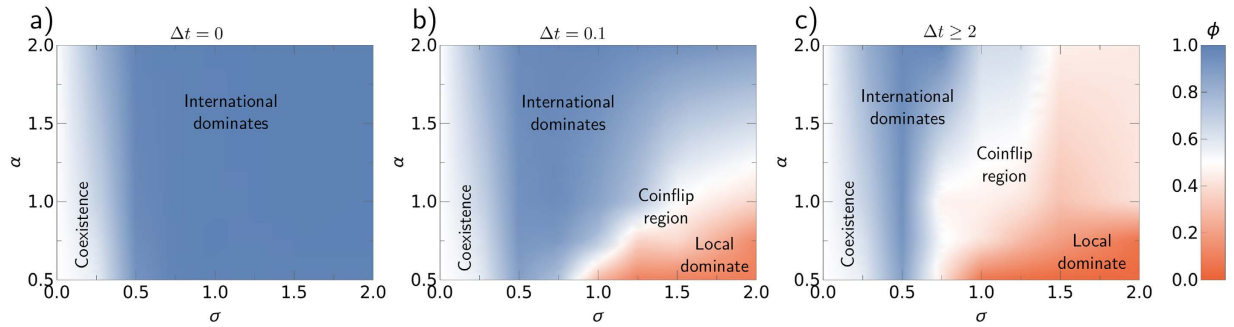
Figure 6. The prevalence of the international network for (**a**) $\Delta t = 0$, (**b**) $\Delta t = 0.1$, and (**c**) averaged over time delays $\Delta t \geq 2$ as a function of the activity affinity (σ) and the global connectivity (α). Averaged over 30 realizations.

corner of Fig. 6, the behavior of the system differs dramatically from one to another. In the coexistence region, each realization of the model leads to the same final state: coexistence of local networks and the international OSN. In contrast, in the coinflip region, coexistence is not possible, as this region of the parameter space corresponds to the supercritical regime (the blue area in Fig. 3). In the coinflip region, about 50% of the realizations end up with domination of the international network, whereas the remaining 50% lead to the domination of local OSNs. As a consequence, even if we know the exact parameters, it is impossible to predict the fate of the system beforehand.

We can summarize these findings as follows. A higher value of α, which is a measure of the global connectivity of society, favors the prevalence of the international network and hinders the survival of the local ones. The role of the tendency of individuals to participate in more active networks (activity affinity), σ, is particularly interesting. Low values allow the networks to coexist, whereas intermediate values always lead to the prevalence of the international network and the extinction of local OSNs. A high activity affinity, however, enables the prevalence of local networks and thus can even lead to the extinction of the international network.

Comparison with empirical data. In this section, we compare the results of our model with empirical data on the recent expansion of Facebook at the cost of many local networks. In particular, we consider the evolution of the number of countries in which local networks (i.e. networks that are not Facebook) are the most popular ones, as measured in[29] using Alexa traffic data (see Fig. 7b). We observe a significant decline of this number, which rules out the possibility that the empiric case corresponds to the domination of local networks. Because the past can be considered a single realization of a stochastic process[30], the empiric case can still be within the coinflip region of our model where –by chance– the international network was more successful. Hence, we will perform the following comparison only for realizations of our model in which local networks do not dominate.

The intrinsic timescale of the model is arbitrary and hence has to be mapped to real time. The optimal mapping is given such that it produces the best agreement with the empirical data. We quantify the agreement between model results and empirical data using the sum of the squared distances between the data points and model results. In particular, we use the χ^2 statistic defined as:

$$\chi^2 = \frac{1}{\sigma_N^2} \sum_i [N_i - N_i^{\text{model}}]^2,$$
(10)

where N_i denotes the number of countries where the local network is more popular and N_i^{model} is the corresponding result from the model. The index i denotes the individual datapoints and σ_N^2 is the estimated variance of the data (Supplementary Materials). The real time, t_R, is a linear function of the model time t_M given by $t_R = a_1 + a_2 t_M$, where a_1 is the starting year and a_2 represents the time stretch: how many years of real time correspond one model time step. For a given set of parameters, σ, α, and Δt, the optimal values for a_1 and a_2 are those that minimize χ^2, as shown in Fig. 7a.

We can also use the χ^2 statistic to estimate the parameters α, σ, and Δt which best reproduce the empirical observations. In Fig. 7c, we plot the values of χ^2 as a function of α, σ, and Δt, where –at each point– we applied the respective best time mapping, as described above. These results are averaged over several realizations of the model; however, in the coinflip region we exclude realizations where the local networks dominate, to mimic the empirical case. Interestingly, the overall best fit is achieved for $\alpha = 2$ at $\sigma = 1.25$ and $\Delta t = 3$, which lies in the coinflip region (the optimal value of χ^2 is statistically consistent with the model, given the number of degrees of freedom in the data) with a probability for domination of the international network of 70%. This scenario corresponds to the time mapping $a_1 = 2006$ and $a_2 = 0.6$, meaning the system started at the beginning of 2006; while the launch time delay of $\Delta t = 3$ in the model translates to 1.8 years in real time. In Fig. 7b, we show the evolution of the number of countries where local networks are more popular for the optimal fit from the model.

Discussion

Understanding the complex dynamics of the digital world constitutes an important challenge for interdisciplinary science. To meet this challenge, here we describe the the worldwide web as a complex, digital ecosystem in which
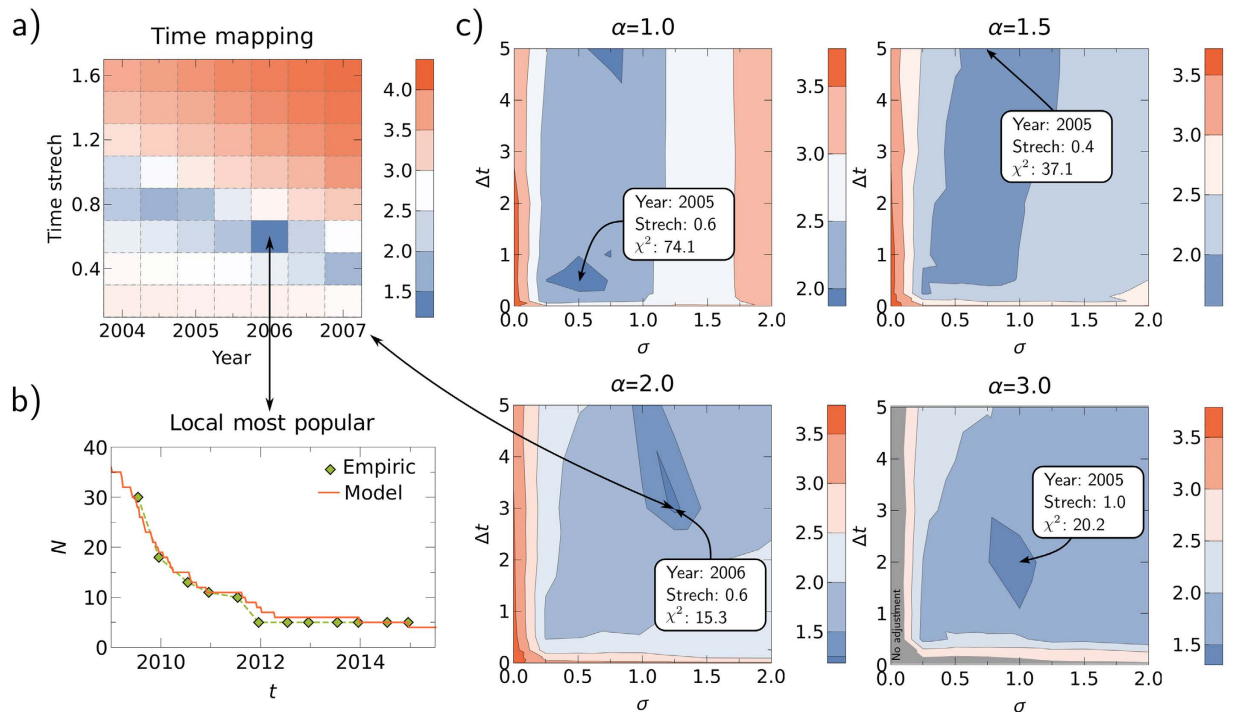
Figure 7. Comparison between model results and empirical data. (a) χ^2 for different values of a_1 (year) and a_2 (time stretch) for $\sigma = 1.25$, $\alpha = 2$, and $\triangle t = 3$. **(b)** Model results and empirical data. Here, the parameters are $\sigma = 1.25$, $\alpha = 2$, and $\triangle t = 3$ and the optimal mapping is given by $a_1 = 2006$ and $a_2 = 0.6$ (see **(a)**). **(c)** χ^2 for the respective best time mapping as a function of σ and $\triangle t$ for different values of α. In each plot, the minimal value of χ^2 and the respective time mapping are shown in the boxes. The color coding in all plots represents the logarithm of χ^2.

interacting networks play the role of species in competition for survival. In particular, we study the competition between local networks operating in single countries and an international network that operates in all countries. Therefore, a proper description of this system must necessarily involve the network of worldwide social interactions between different countries.

We show that the effect of inter-country social ties can be mapped to the increased fitness of the international network by means of an effective activity. Interestingly, there is a critical global coupling strength below which networks can coexist. However, above that threshold, only domination is possible: in general, local networks become extinct with a high probability. Yet, we find that if local networks are launched earlier they can persist and dominate the international network, which happens only if local networks have accumulated a sufficiently large active userbase when the global launch of the international network takes place. The accumulation of a sufficient base depends on the parameters; and for certain parameters on chance. For these parameters the final state of the system –whether local networks dominate or become extinct– can be completely unpredictable, as it varies randomly between different realizations of the model.

Quite remarkably, a thorough comparison of our model with empirical data from the recent takeover of Facebook indicates that the most probable launch date of Facebook was at the beginning of 2006 and its global launch was in late 2007. Facebook was in fact started in 2004, but opened to the public in 2006; in good agreement with the estimate from our model. Moreover, according to Google trend data (see Supplementary Materials), 2007 was the year when the global search volume for Facebook started to increase rapidly. Last but not least, our best estimation of the model parameters corresponds to the "coinflip" region, which means that the observed takeover of Facebook only had a probability of around 70%. With a 30% probability, we would have been living in a world where each country had its own successful local network and a network like Facebook would not exist[30].

Our findings suggest interesting future lines of research. On the one hand, even without adjusting the parameters on a country-by-country level, our model reproduces the main features empirically observed in the takeover of Facebook and the extinction of local networks in most countries for a certain parameter region. It remains an interesting task for future research to further increase the precision of the model. This could be done by improving the proxy for the similarity between countries or by adjusting parameters on a country-by-country basis. On the other hand, the model could be extended to account for several international networks and to study their global competition. For a second international network to overcome the first, a certain minimal difference of fitness is needed; which could be the result of different properties of the networks, such as features or functionalities. Finally, random fluctuations of fitness could be incorporated to describe Darwinian selection in the digital ecosystem.
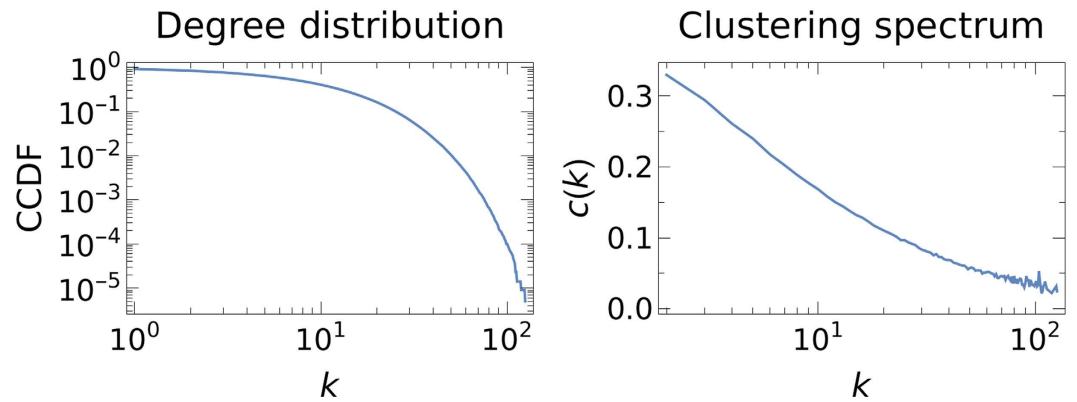
Degree distribution

Clustering spectrum

Figure 8. Degree distribution and clustering spectrum for networks generated for the example of the US ($\approx$230.000 nodes).

Methods and Material

S1 model. We use the $\mathbb{S}1$ model[26–28] to generate the synthetic networks for the underlying societies in each country. The model allows us to specify the degree distribution and the level of clustering. The model is based on a circle as a hidden metric space and works as follows:

1. All nodes are placed on the circle with a randomly assigned variable, θ, which represents the polar coordinate. θ is uniformly distributed in $[0, 2\pi)$. To keep the average node density on the circle constant, its radius grows linearly with the number of nodes, to satisfy $N = 2\pi R$.
2. We assign each node a second hidden variable, κ, which represents its expected degree. κ is drawn from an arbitrary distribution $\rho(\kappa)$.
3. A pair of nodes are connected with a probability, r, that depends on their hidden variables (θ, κ) and (θ', κ'):

$$r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu\kappa\kappa'}\right)^{-\alpha},$$

(11)

with $\mu = \frac{\alpha - 1}{2\langle k \rangle}$. Here, $d(\theta, \theta')$ denotes the geodesic distance between the two nodes on the circle and $\langle \kappa \rangle$ the mean degree. Then, the expected degree, $\overline{k}(\kappa)$, of a node with hidden variable κ can be shown to be proportional to κ[28]. As a consequence, the degree distribution, $p(k)$, of the network follows the shape of the distribution $\rho(\kappa)$.

Here, we use an exponential distribution $\rho_\xi(\kappa) = \xi e^{-\xi\kappa}$ with $\xi = 10$. We set the parameters $\alpha = 1.5$ and $\mu = 0.02$. After generating the networks, we remove nodes with zero degree. Figure 8 shows the degree distribution and the clustering spectrum for the synthetic network created for the US.

Air travel data. Air travel data aggregated on a country basis were taken from http://visualising.org/datasets/global-flights-network (date of access July 2014). The original data can be accessed at http://openflights.org/data.html. The network on a country basis incorporates 230 nodes and 4600 weighted edges which correspond to the number of routes between countries, i.e. the number of total flights. The dataset contains around 60.000 of such flights. We extract the subnetwork (see Fig. 1c) by constraining to the countries listed in Table 1.

References

1. WeAreSocial. Sdmw/we are social (year published: 2015, date of access: 31/03/2015). URL http://wearesocial.net/tag/sdmw/.
2. Brenner, J. & Smith, A. 72% of online adults are social networking site users. *Pew Research Internet Project* (2013, Date of access: 31/03/2015). URL http://www.pewinternet.org/2013/08/05/72-of-online-adults-are-social-networking-site-users.
3. Kleineberg, K.-K. & Boguñá, M. Digital ecology: Coexistence and domination among interacting networks. *Sci. Rep.* **5,** 10268 (2015).
4. Hardin, G. The competitive exclusion principle. *Science* **131,** 1292–1297 (1960).
5. Reis, S. D. S. *et al.* Avoiding catastrophic failure in correlated networks of networks. *Nature Physics* **10,** 762–767 (2014).
6. Radicchi, F. Driving Interconnected Networks to Supercriticality. *Phys Rev X* **4,** 021014 (2014).
7. Bianconi, G. & Dorogovtsev, S. N. Multiple percolation transitions in a configuration model of a network of networks. *Phys Rev E* **89,** 062814 (2014).
8. D'Agostino, G. & Scala, A. *Networks of Networks: The Last Frontier of Complexity* vol. XII of *Understanding Complex Systems* (Springer, 2014).
9. Kleineberg, K.-K. & Boguñá, M. Evolution of the digital society reveals balance between viral and mass media influence. *Phys Rev X* **3,** 031046 (2014).
10. Jacobs, A. Z., Way, S. F., Ugander, J. & Clauset, A. Assembling thefacebook: Using heterogeneity to understand online social network assembly. *ArXiv 1503.06772* (2015).
11. Anderson, R. & May, R. M. *Infectious diseases of humans: Dynamics and control* (Oxford University Press, Oxford, 1992).
12. Ribeiro, B. Modeling and Predicting the Growth and Death of Membership-based Websites. In *Proceedings of the 23rd International Conference on World Wide Web,* WWW '14, 653–664 (International World Wide Web Conferences Steering Committee, 2014).
13. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).

14. Dorogovtsev, S. N., Mendes, J. & Samukhin, A. Size-dependent degree distribution of a scale-free growing network. *Phys Rev E* **63,** 062101 (2001).
15. Bianconi, G. & Barabási, A.-L. Bose-Einstein Condensation in Complex Networks. *Phys Rev Lett* **86,** 5632–5635 (2001).
16. Caldarelli, G., Capocci, A., Rios, P. D. L. & Muñoz, M. A. Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Phys Rev Lett* **89,** 258702 (2002).
17. Vázquez, A. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys Rev E* **67,** 056104 (2003).
18. Pastor-Satorras, R., Smith, E. & Sole, R. V. Evolving protein interaction networks through gene duplication. *J Theor Biol* **222,** 199–210 (2003).
19. Fortunato, S., Flammini, A. & Menczer, F. Scale-Free Network Growth by Ranking. *Phys Rev Lett* **96,** 218701 (2006).
20. D'Souza, R. M., Borgs, C., Chayes, J. T., Berger, N. & Kleinberg, R. D. Emergence of tempered preferential attachment from optimization. *Proc Natl Acad Sci USA* **104,** 6112–7 (2007).
21. Papadopoulos, F., Kitsak, M., Serrano, M., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489,** 537–540 (2012).
22. State, B., Park, P., Weber, I., Mejova, Y. & Macy, M. The Mesh of Civilizations and International Email Flows. *ArXiv 1303.0045* (2013).
23. Takhteyev, Y., Gruzd, A. & Wellman, B. Geography of twitter networks. *Social Networks* **34,** 73–81 (2012).
24. De Domenico, M. *et al.* Mathematical formulation of multilayer networks. *Phys Rev X* **3,** 041022 (2013).
25. Kivelä, M. *et al.* Multilayer networks. *J Complex Netw.* **2,** 203–271 (2014).
26. Serrano, M., Krioukov, D. & Boguñá, M. Self-Similarity of Complex Networks and Hidden Metric Spaces. *Phys Rev Lett* **100,** 078701 (2008).
27. Boguñá, M., Krioukov, D. & Claffy, K. C. Navigability of complex networks. *Nature Physics* **5,** 74–80 (2008).
28. Boguñá, M. & Pastor-Satorras, R. Class of correlated random networks with hidden variables. *Phys Rev E* **68,** 036112 (2003).
29. Cosenza, V. World map of social networks (Date of access: 31/03/2015). URL http://vincos.it/world-map-of-social-networks.
30. Watts, D. J. *Everything Is Obvious: How Common Sense Fails Us* (Crown Business, 2012).

Acknowledgements

Author Contributions

Both authors planned the research and wrote the paper. K.-K.K. performed numerical simulations and analytical calculations and prepared all the figures.

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: Kleineberg, K.-K. and Boguñá, M. Competition between global and local online social networks. *Sci. Rep.* **6,** 25116; doi: 10.1038/srep25116 (2016).