Identification of influential spreaders in complex networks

Maksim Kitsak[1,2], Lazaros K. Gallos[3], Shlomo Havlin[4], Fredrik Liljeros[5], Lev Muchnik[6], H. Eugene Stanley[1] and Hernán A. Makse[3]★

Networks portray a multitude of interactions through which people meet, ideas are spread and infectious diseases propagate within a society[1–5]. Identifying the most efficient 'spreaders' in a network is an important step towards optimizing the use of available resources and ensuring the more efficient spread of information. Here we show that, in contrast to common belief, there are plausible circumstances where the best spreaders do not correspond to the most highly connected or the most central people[6–10]. Instead, we find that the most efficient spreaders are those located within the core of the network as identified by the k-shell decomposition analysis[11–13], and that when multiple spreaders are considered simultaneously the distance between them becomes the crucial parameter that determines the extent of the spreading. Furthermore, we show that infections persist in the high-k shells of the network in the case where recovered individuals do not develop immunity. Our analysis should provide a route for an optimal design of efficient dissemination strategies.

Spreading is a ubiquitous process, which describes many important activities in society[2–5]. The knowledge of the spreading pathways through the network of social interactions is crucial for developing efficient methods to either hinder spreading in the case of diseases, or accelerate spreading in the case of information dissemination. Indeed, people are connected according to the way they interact with one another in society and the large heterogeneity of the resulting network greatly determines the efficiency and speed of spreading. In the case of networks with a broad degree distribution (number of links per node)[6], it is believed that the most connected people (hubs) are the key players, being responsible for the largest scale of the spreading process[6–8]. Furthermore, in the context of social network theory, the importance of a node for spreading is often associated with the betweenness centrality, a measure of how many shortest paths cross through this node, which is believed to determine who has more 'interpersonal influence' on others[9,10].

Here we argue that the topology of the network organization plays an important role such that there are plausible circumstances under which the highly connected nodes or the highest-betweenness nodes have little effect on the range of a given spreading process. For example, if a hub exists at the end of a branch at the periphery of a network, it will have a minimal impact in the spreading process through the core of the network, whereas a less connected person who is strategically placed in the core of the network will have a significant effect that leads to dissemination through a large fraction of the population. To identify the core and the periphery of the network we use the k-shell (also called k-core) decomposition of the network[11–14]. Examining this quantity in a number of real networks enables us to identify the best individual spreaders in the network when the spreading originates in a single node. For the case of a spreading process originating in many nodes simultaneously, we show that we can further improve the efficiency by considering spreading origins located at a determined distance from one another.

We study real-world complex networks that represent archetypical examples of social structures. We investigate (1) the friendship network between 3.4 million members of the LiveJournal.com community[15], (2) the network of email contacts in the Computer Science Department of University College London (Zhou, S., private communication), (3) the contact network of inpatients (CNI) collected from hospitals in Sweden[16] and (4) the network of actors who have costarred in movies labelled by imdb.com as adult[17] (see Supplementary Section SI for details).

To study the spreading process we apply the susceptible–infectious–recovered (SIR) and susceptible–infectious–susceptible (SIS) models[2,3,18] on the above networks (see Methods). These models have been used to describe disease spreading as well as information and rumour spreading in social processes where an actor constantly needs to be reminded[19]. We denote the probability that an infectious node will infect a susceptible neighbour as β. In our study we use relatively small values for β, so that the infected percentage of the population remains small. In the case of large β values, where spreading can reach a large fraction of the population, the role of individual nodes is no longer important and spreading would cover almost all the network, independently of where it originated from.

The location of a node is defined using the k-shell decomposition analysis[11–13]. This process assigns an integer index or coreness, k_S, to each node, representing its location according to successive layers (k shells) in the network. The k_S index is a quite robust measure and the node ranking is not influenced significantly in the case of incomplete information. (For details see Supplementary Fig. S6 in Section SII. Small values of k_S define the periphery of the network and the innermost network core corresponds to large k_S (see Fig. 1a and Supplementary Section SII.) Figure 1b–d illustrates the fact that the size of the population infected in a spreading process (shown in this example in the CNI network)

[1]Center for Polymer Studies and Physics Department, Boston University, Boston, Massachusetts 02215, USA, [2]Cooperative Association for Internet Data Analysis (CAIDA), University of California-San Diego, La Jolla, California 92093, USA, [3]Levich Institute and Physics Department, City College of New York, New York, New York 10031, USA, [4]Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan, Israel, [5]Department of Sociology, Stockholm University, S-10691, Stockholm, Sweden, [6]Information, Operations and Management Sciences Department, Stern School of Business, New York University, New York, New York 10012, USA. *e-mail: hmakse@lev.ccny.cuny.edu.
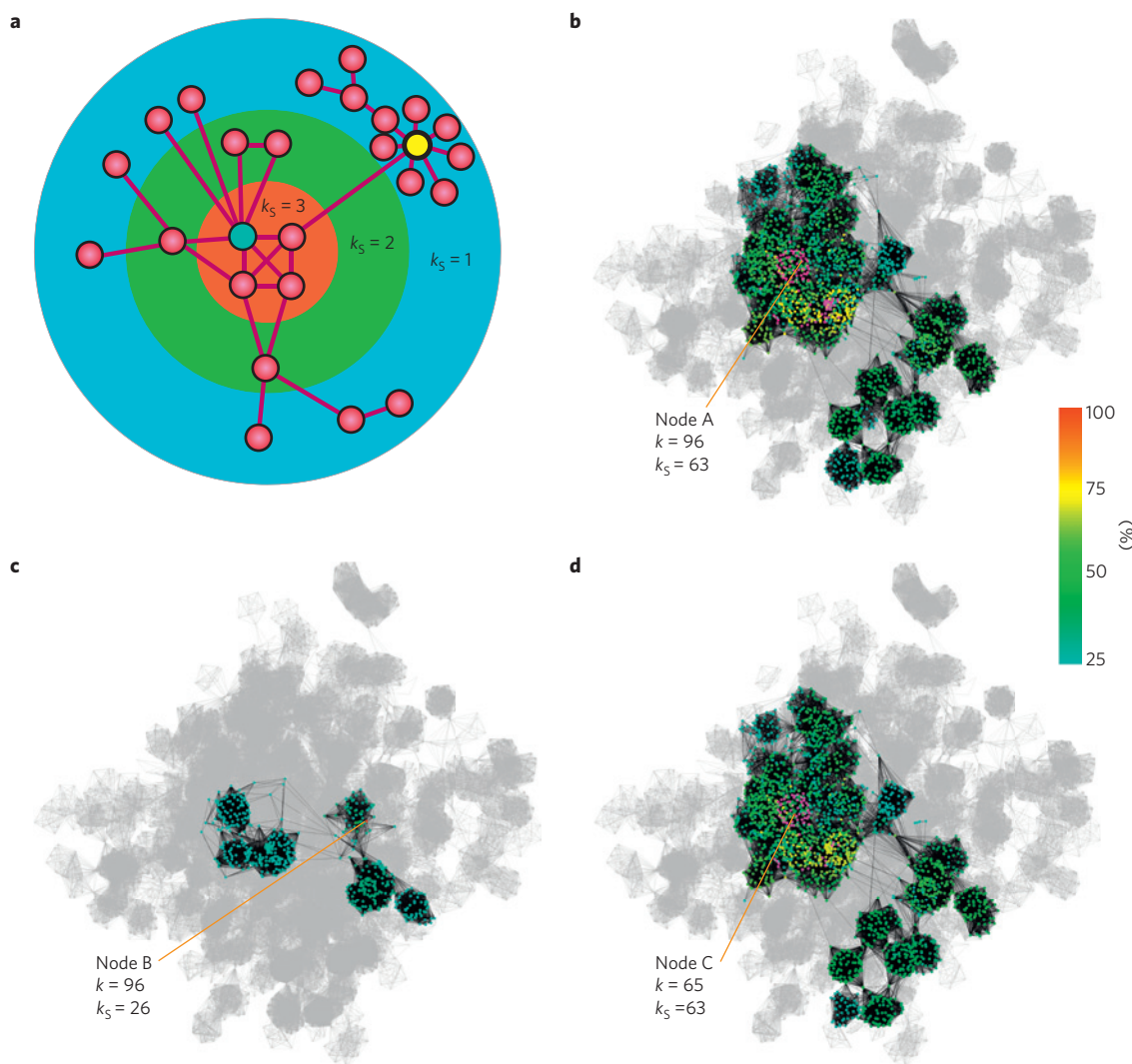
Figure 1 | When the hubs may not be good spreaders. a, A schematic representation of a network under the k-shell decomposition. The two nodes of degree $k = 8$ (blue and yellow nodes) in this network are in different locations: one lies at the periphery ($k_S = 1$) whereas the other hub is in the innermost core of the network, that is, it has the largest k_S ($k_S = 3$). **b–d**, The extent of the efficiency of the spreading process cannot be accurately predicted on the basis of a measure of the immediate neighbourhood of the node, such as the degree k. For the contact network of inpatients (CNI), we compare infections originating from single nodes with the same degree $k = 96$ (nodes A and B) or the same index $k_S = 63$ (nodes A and C), with infection probability $\beta = 0.035$. In the corresponding plots, the colours indicate the probability that a node will be infected when spreading starts in the corresponding origin, as long as this probability is higher than 25%. The results are based on 10,000 different realizations for each case. In the first case, where origin A has $k_S = 63$, spreading reaches a much wider area more frequently, in contrast to origin B ($k_S = 26$), where the infection remains largely localized in the immediate neighbourhood of B. Spreading is very similar between origins A and C, which have the same k_S value, although the degree of C is much smaller than A. The importance of the network organization is also highlighted when we randomly rewire the network (preserving the same degree for all nodes). In this case the standard picture is recovered: the extent of spreading coincides and both hubs contribute equally well to spreading (see Supplementary Section SVI).

is not necessarily related to the degree of the node, k, where the spreading started. Spreading may be very different even when it starts from hubs of similar degrees as comparatively shown in Fig. 1b and c. Instead, the location of the spreading origin given by its k_S index predicts more accurately the size of the infected population. For instance, Fig. 1b and d show that nodes in the same k_S layer produce similar spreading areas even if they have different k (by definition, in a given layer there could be many nodes with $k \geq k_S$).

The above example suggests that the position of the node relative to the organization of the network determines its spreading influence more than a local property of a node, such as the degree k. To quantify the influence of a given node i in an SIR spreading process we study the average size of the population M_i infected in an epidemic originating at node i with a given (k_S, k). The

infected population is averaged over all the origins with the same (k_S, k) values:

$$M(k_S, k) = \sum_{i \in \Upsilon(k_S, k)} \frac{M_i}{N(k_S, k)}$$

where $\Upsilon(k_S, k)$ is the union of all $N(k_S, k)$ nodes with (k_S, k) values.

The analysis of $M(k_S, k)$ in the studied social networks reveals three general results (see Fig. 2): (1) For a fixed degree, there is a wide spread of $M(k_S, k)$ values. In particular, there are many hubs located at the periphery of the network (large k, low k_S) that are poor spreaders. (2) For a fixed k_S, $M(k_S, k)$ is approximately independent of the degree of the nodes. This result is revealed in the vertically layered structure of $M(k_S, k)$, suggesting that infected nodes located in the same k shell produce similar epidemic outbreaks $M(k_S, k)$
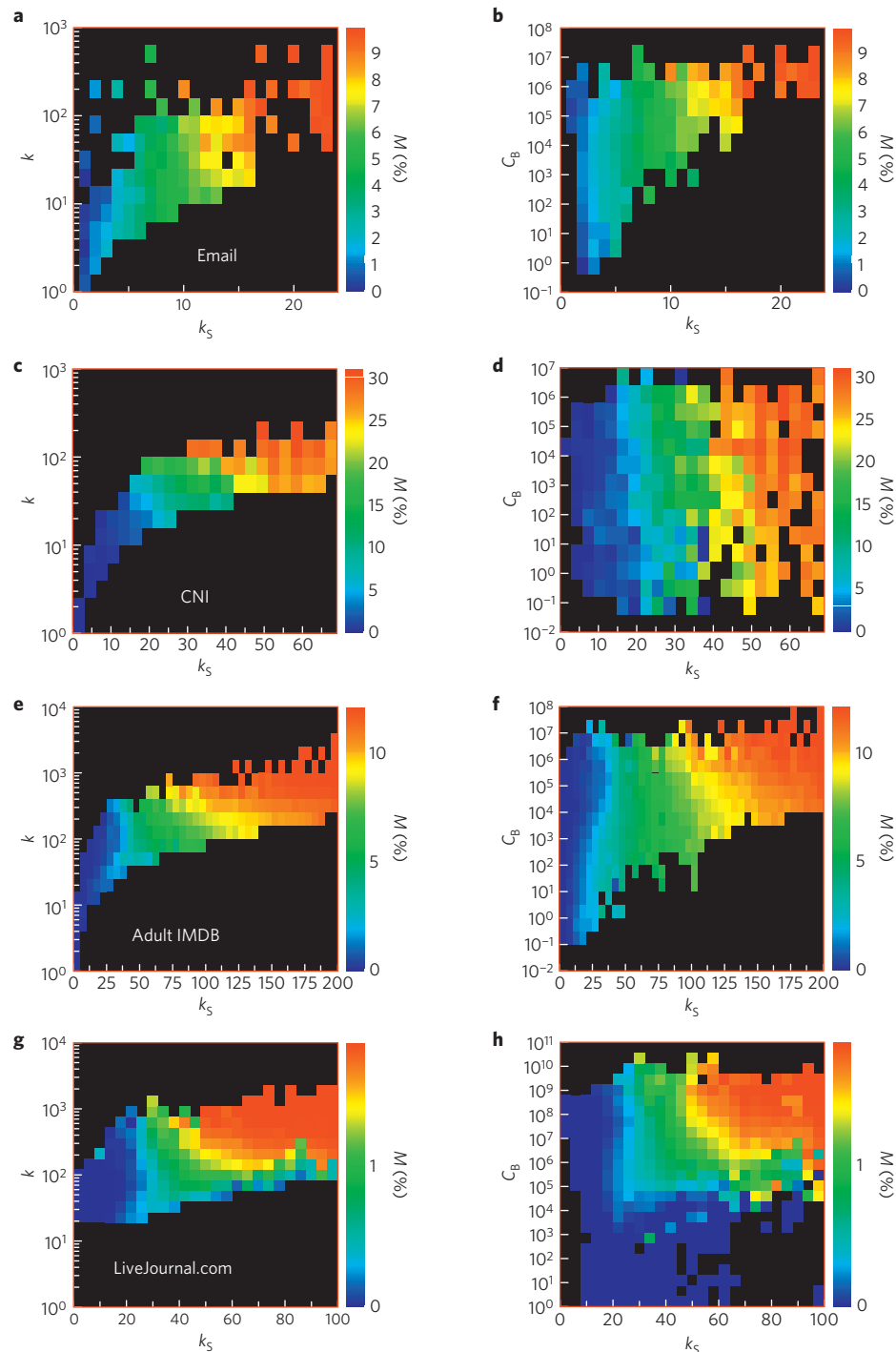
889

Figure 2 | The k-shell index predicts the outcome of spreading more reliably than the degree k or the betweenness centrality C_B. The networks used are (top to bottom) email contacts ($\beta = 8\%$), the CNI network ($\beta = 4\%$), the actor network ($\beta = 1\%$) and the LiveJournal.com friendship network ($\beta = 1.5\%$). **a,c,e,g**, Average infected size of the population $M(k_S, k)$ when spreading originates in nodes with (k_S, k). **b,d,f,h**, The infected size $M(k_S, C_B)$ when spreading originates in nodes of a given combination of k_S and C_B. In both cases, spreading is larger for nodes of higher k_S, whereas nodes of a given k or C_B value can result in either small or large spreading, depending on the value of k_S. (There is an exception at large k_S and small k of the LiveJournal database, which is due to artificial closed groups of virtual characters that connect with one another for the purpose of online gaming and do not correspond to regular users, as the rest of the database.)

independent of the value of k of the infection origin. (3) The most efficient spreaders are located in the inner core of the network (large k_S region), fairly independently of their degree. These results indicate that the k-shell index of a node is a better predictor of spreading influence. When an outbreak starts in the core of the network (large k_S) there exist many pathways through which a virus can infect the rest of the network; this result is valid regardless of the node degree. The existence of these pathways implies that, during a typical epidemic outbreak from a random origin, nodes located in high-k_S layers are more likely to be infected and they will be infected earlier than other nodes (see Supplementary Section SIII). The neighbourhood of these nodes makes them more efficient
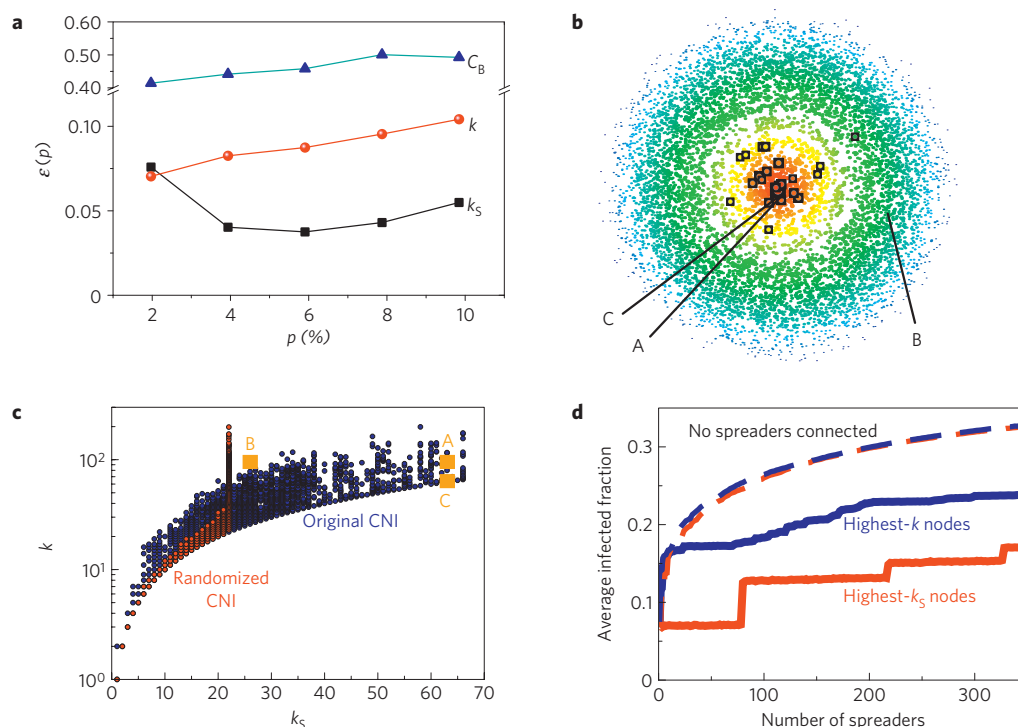
Figure 3 | k-shell structure of the CNI network. a, The imprecision functions $\epsilon_{k_S}(p)$, $\epsilon_k(p)$ and $\epsilon_{C_B}(p)$, for $\beta = 4\%$. Even though both k-shell and k identification strategies yield comparable results for $p = 2\%$, the k-shell strategy is consistently more accurate for $2\% < p < 10\%$, with ϵ_{k_S} approximately half ϵ_k. The C_B identification of the most efficient spreaders is the least accurate, with ϵ_{C_B} exceeding 40%. **b**, We visualize the CNI network as a set of concentric circles of nodes representing inpatients, each circle corresponding to a particular k shell. The k_S indices of a given layer increase as we move from the periphery to the centre of the network[28,29]. Node size is proportional to the logarithm of the degree of the node. We highlight the 25 inpatients with the largest degree values. Note that inpatients with high k values are not concentrated at the 'centre' of the network but instead are scattered throughout different k shells. We highlight the position of the three nodes, A, B and C, of the origins that were used in the example of Fig. 1. **c**, Scatter plot of the node degree k as a function of k_S for all the nodes in the CNI network (black symbols) and the degree-preserving randomized version of the same network (red symbols). Note that there are many inpatients with large k and low k_S values in the original network, whereas in the randomized email network all the hubs are located in the inner core of the network. We also show the positions of the three origins used in Fig. 1. **d**, When spreading starts from multiple origins, the set of nodes with highest degree (blue continuous line) can spread significantly more than the set of highest-k_S nodes (red continuous line), because in the latter case most of these nodes are connected to one another. If we only consider in this set nodes that are not directly linked, then both the sets of highest-k or k_S nodes yield a similar result (dashed lines), where spreading is significantly enhanced. Results are shown for $\beta = 3\%$ in the CNI.

in sustaining an infection in the early stages, thus enabling the epidemic to reach a critical mass such that it can fully develop. Similar results on the efficiency of high-k_S nodes are obtained from the analysis of $M(k_S, C_B)$ in Fig. 2, where C_B is the betweenness centrality of a node in the network[9,10]: the value of C_B is not a good predictor for spreading efficiency.

To quantify the importance of k_S in spreading we calculate the 'imprecision functions' $\epsilon_{k_S}(p)$, $\epsilon_k(p)$ and $\epsilon_{C_B}(p)$. These functions estimate for each of the three indicators k_S, k and C_B how close to the optimal spreading is the average spreading of the pN $(0 < p < 1)$ chosen origins in each case (see Methods and Supplementary Section SIV). The strategy to predict the spreading efficiency of a node based on k_S is consistently more accurate than a method based on k in the studied p range (Fig. 3a). The C_B-based strategy gives poor results compared with the other two strategies.

Our finding is not specific to the social networks shown in Fig. 2. In Supplementary Section SV we analyse the spreading efficiency in other networks not social in origin, such as the Internet at the router level[20], with similar conclusions. The key insight of our finding is that in the studied networks a large number of hubs are located in the peripheral low-k_S layers (Fig. 3b shows the location of the 25 largest hubs in the CNI; see also Supplementary Section SV) and therefore contribute poorly to spreading. The existence of hubs in the periphery is a consequence of the rich topological structure of

real networks. In contrast, in a fully random network obtained by randomly rewiring a real network preserving the degree of each node (such a random network corresponds to the configuration model[21]; see Supplementary Section SVI) all the hubs are placed in the core of the network (see the red scatter plot in Fig. 3c) and they contribute equally well to spreading. In such a randomized structure the same information is contained in the k shell as in the degree classification because there is a one-to-one relation between the two quantities, which is approximately linear, $k_S \propto k$ (Fig. 3c and Supplementary Fig. S13). Examples of real networks that are similar to a random structure are the network of product space of economic goods[22] and the Internet at the AS level (analysed in Supplementary Section SV).

Our study highlights the importance of the relative location of a single spreading origin. Next, we address the question of the extent of an epidemic that starts at multiple origins simultaneously. Figure 3d shows the extent of SIR spreading in the CNI network when the outbreak simultaneously starts from the n nodes with the highest degree k or the highest k_S index. Even though the high-k_S nodes are the best single spreaders, in the case of multiple spreading the nodes with highest degree are more efficient than those with highest k_S. This result is attributed to the overlap of the infected areas of the different spreaders: large-k_S nodes tend to be clustered close to one another, whereas hubs can be more
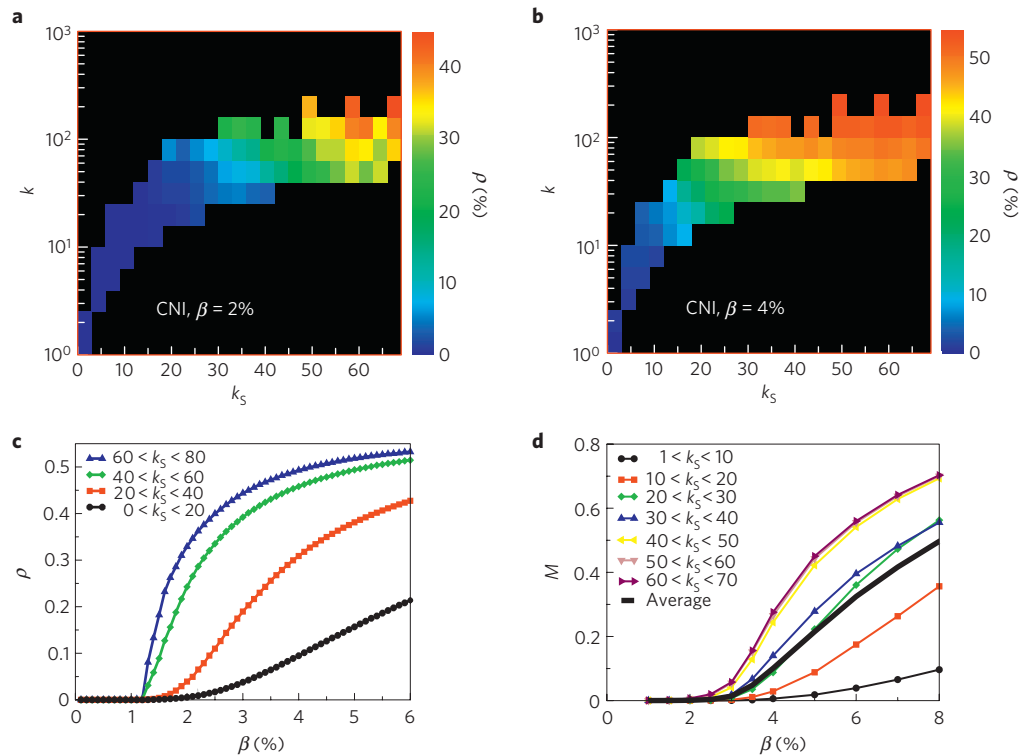
891

Figure 4 | SIS spreading in the CNI network and β dependence for SIS and SIR. a,b, Virus persistence $\rho(k_S, k)$ as a function of k and k_S values of inpatients in the CNI network for $\beta = 2\%$ and $\beta = 4\%$, respectively, where 20% of the individuals are initially infected. The infection survives mainly in nodes with large k_S values. **c,** We form four groups of nodes of the CNI network on the basis of their k-shell values. For all values of β, the average virus persistence ρ is consistently higher in the inner k shells. **d,** Influence of the infection probability β on the spreading efficiency of nodes, grouped according to their k-shell values, for SIR spreading. The solid black line refers to the average infected percentage over all network nodes. Nodes in higher-k shells are consistently the most efficient, independently of the β value.

spread in the network and, in particular, they need not be connected with one another. Clearly, the step-like features in the plot of highest-k_S nodes (red solid curve in Fig. 3d) suggest that the infected percentage remains constant as long as the infected nodes belong in the same k shell. Including just one node from a different k shell results in a significantly increased spreading. This result suggests that a better spreading strategy using n spreaders is to choose either the highest-k or k_S nodes with the requirement that no two of the n spreaders are directly linked to each other. This scheme then provides the largest infected area of the network, as shown in Fig. 3d.

Many contagious infections, including most sexually transmitted infections[23], do not confer full immunity after infection as assumed in the SIR model, and therefore are suitably described by the SIS epidemic model, where an infectious node returns to the susceptible state with probability λ. In an SIS epidemic the number of infectious nodes eventually reaches a dynamic-equilibrium 'endemic' state, where as many infectious individuals become susceptible as susceptible nodes become infectious[18]. In contrast to SIR, in the initial state of our SIS simulations 20% of the network nodes are already infected. The spreading efficiency of a given node i in SIS spreading is the persistence, $\rho_i(t)$, defined as the probability that node i is infected at time t (ref. 7). In an endemic SIS state, $\rho_i(t \to \infty)$ becomes independent of t (see Supplementary Section SVII). Previous studies have shown that the largest persistence $\rho_i(t \to \infty)$ is found in the network hubs, which are re-infected frequently owing to the large number of neighbours[7,24,25]. However, we find that this result holds only in randomized network structures. In the real network topologies studied here, we find that viruses persist mainly in high-k_S layers instead, almost irrespectively of the degree of the nodes in the core.

In the case of random networks, it is found that viruses propagate to the entire network above an epidemic threshold given by $\beta > \beta_c^{\text{rand}} \equiv \lambda \langle k \rangle / \langle k^2 \rangle$ (refs 24,26). In real networks, such as the CNI network, the threshold β_c is different from β_c^{rand}. Furthermore, in real networks, we find that viruses can survive locally even when $\beta < \beta_c$, but only within the high-k_S layers of the network, whereas virus persistence in peripheral k_S layers is negligible (Fig. 4a–c). As the k-shell structure depends on the network assortativity, the lower threshold is in agreement with the observation that high positive assortativity[27] may decrease the epidemic threshold.

The importance of high-k_S nodes in SIS spreading is confirmed when we analyse the asymptotic probability that nodes of given (k_S, k) values will be infected. This probability is quantified by the persistence function

$$\rho(k_S, k) \equiv \sum_{i \in \Upsilon(k_S, k)} \frac{\rho_i(t \to \infty)}{N(k_S, k)}$$

as a function of (k_S, k) at different β values (Fig. 4a and b). High-k_S layers in networks might be closely related to the concept of a core group in sexually transmitted infection research[23]. The core groups are defined as subgroups in the general population characterized by high partner turnover rate and extensive intergroup interaction[23].

Similar to the core group, the dense subnetwork formed by nodes in the innermost k shells helps the virus to consistently survive locally in the inner-core area and infect other nodes adjacent to the area. These k shells preserve the existence of a virus, in contrast to, for example, isolated hubs at the periphery. Note that a virus cannot survive in the degree-preserving randomized version of the CNI network, owing to the absence of high-k shells.

The importance of the inner-core nodes in spreading is not influenced by the infection probability values, β. In both models, SIS and SIR, we find that the persistence ρ or the average infected fraction M, respectively, is systematically larger for nodes in inner k shells compared with nodes in outer k shells, over the entire β range that we studied (Fig. 4c,d). Thus, the k-shell measure is a robust indicator for the spreading efficiency of a node.

Finding the most accurate ranking of individual nodes for spreading in a population can influence the success of dissemination strategies. When spreading starts from a single node the k_S value is enough for this ranking, whereas in the case of many simultaneous origins spreading is greatly enhanced when we additionally repel the spreaders with large degree or k_S. In the case of infections that do not confer immunity on recovered individuals, the core of the network in the large-k_S layers forms a reservoir where infection can survive locally.

Methods

The k-shell decomposition. Nodes are assigned to k shells according to their remaining degree, which is obtained by successive pruning of nodes with degree smaller than the k_S value of the current layer. We start by removing all nodes with degree $k = 1$. After removing all the nodes with $k = 1$, some nodes may be left with one link, so we continue pruning the system iteratively until there is no node left with $k = 1$ in the network. The removed nodes, along with the corresponding links, form a k shell with index $k_S = 1$. In a similar fashion, we iteratively remove the next k shell, $k_S = 2$, and continue removing higher-k shells until all nodes are removed. As a result, each node is associated with one k_S index, and the network can be viewed as the union of all k shells. The resulting classification of a node can be very different than when the degree k is used.

The spreading models. To study the spreading process we apply the SIR and SIS models. In the SIR model, all nodes are initially in the susceptible state (S) except for one node in the infectious state (I). At each time step, the I nodes infect their susceptible neighbours with probability β and then enter the recovered state (R), where they become immunized and cannot be infected again. The SIS model aims to describe spreading processes that do not confer immunity on recovered individuals: infected individuals still infect their neighbours with probability β but they return to the susceptible state with probability λ (here we use $\lambda = 0.8$) and can be reinfected at subsequent time steps, and they remain infectious with probability $1 - \lambda$.

The imprecision function. The betweenness centrality, $C_B(i)$, of a node i is defined as follows: Consider two nodes s and t and the set σ_{st} of all possible shortest paths between these two nodes. If the subset of this set that contains the paths that pass through the node i is denoted by $\sigma_{st}(i)$, then the betweenness centrality of this node is given by

$$C_B(i) = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where the sum runs over all nodes s and t in the network.

The imprecision function $\epsilon(p)$ quantifies the difference between the average spreading between the pN nodes ($0 < p < 1$) with highest k_S, k or C_B and the average spreading of the pN most efficient spreaders (N is the number of nodes in the network). Thus, it tests the merit of using k shell, k and C_B to identify the most efficient spreaders. For a given β value and a given fraction of the system p we first identify the set of the Np most efficient spreaders as measured by M_i (we designate this set by Υ_{eff}). Similarly, we identify the Np individuals with the highest k-shell index (Υ_{k_S}). We define the imprecision of k-shell identification as $\epsilon_{k_S}(p) \equiv 1 - M_{k_S}/M_{\text{eff}}$, where M_{k_S} and M_{eff} are the average infected percentages averaged over the Υ_{k_S} and Υ_{eff} groups of nodes respectively. ϵ_k and ϵ_{C_B} are defined similarly to ϵ_{k_S}.

References

1. Caldarelli, G. & Vespignani, A. (eds) *Large Scale Structure and Dynamics of Complex Networks* (World Scientific, 2007).
2. Anderson, R. M., May, R. M. & Anderson, B. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Science Publications, 1992).
3. Diekmann, O. & Heesterbeek, J. A. P. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (Wiley Series in Mathematical & Computational Biology, 2000).
4. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton Univ. Press, 2008).
5. Rogers, E. M. *Diffusion of Innovation* 4th edn (Free Press, 1995).
6. Albert, R., Jeong, H. & Barabási, A-L. Error and attack tolerance of complex networks. *Nature* **406,** 378–482 (2000).
7. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).
8. Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. Breakdown of the Internet under intentional attack. *Phys. Rev. Lett.* **86,** 3682–3685 (2001).
9. Freeman, L. C. Centrality in social networks: Conceptual clarification. *Social Networks* **1,** 215–239 (1979).
10. Friedkin, N. E. Theoretical foundations for centrality measures. *Am. J. Sociology* **96,** 1478–1504 (1991).
11. Bollobás, B. *Graph Theory and Combinatorics: Proceedings of the Cambridge Combinatorial Conference in Honor of P. Erdös* Vol. 35 (Academic, 1984).
12. Seidman, S. B. Network structure and minimum degree. *Social Networks* **5,** 269–287 (1983).
13. Carmi, S., Havlin, S, Kirkpatrick, S., Shavitt, Y. & Shir, E. A model of Internet topology using k-shell decomposition. *Proc. Natl Acad. Sci. USA* **104,** 11150–11154 (2007).
14. Ángeles-Serrano, M. & Boguñá, M. Clustering in complex networks. II. Percolation properties. *Phys. Rev. E* **74,** 056116 (2006).
15. LiveJournal, http://www.livejournal.com.
16. Liljeros, F., Giesecke, J. & Holme, P. The contact network of inpatients in a regional healthcare system. A longitudinal case study. *Math. Population Studies* **14,** 269–284 (2007).
17. *The Internet Movie Database*, http://www.imdb.com.
18. Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42,** 599–653 (2000).
19. Castellano, C., Fortunato, S. & Loretto, V. Statistical Physics of Social Dynamics. *Rev. Mod. Phys.* **81,** 591–646 (2009).
20. Shavitt, Y. & Shir, E. DIMES: Let the internet measure itself. *ACM SIGCOMM Comput. Commun. Rev.* **35,** 71–74 (2005).
21. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6,** 161–180 (1995).
22. Hidalgo, C. A., Klinger, B., Barabasi, A-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317,** 482–487 (2007).
23. Hethcote, H. & Rogers, J. A. *Gonorrhea Transmission Dynamics and Control* (Springer-Verlag, 1984).
24. Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65,** 036104 (2002).
25. Dezsó, Z. & Barabási, A-L. Halting viruses in scale-free networks. *Phys. Rev. E* **65,** 055103 (2002).
26. Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85,** 4626–4630 (2000).
27. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89,** 208701 (2002).
28. Large Network visualization tool, http://xavier.informatics.indiana.edu/lanet-vi/.
29. Alvarez-Hamelin, J. I., Dallásta, L., Barrat, A. & Vespignani, A. Large scale networks fingerprinting and visualization using the k-core decomposition. *Adv. Neural Inform. Process. Systems* **18,** 41–51 (2006).

Author contributions

All authors contributed equally to the work presented in this paper.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper on www.nature.com/naturephysics. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions. Correspondence and requests for materials should be addressed to H.A.M.