**Global efficiency of local immunization on complex networks**
Supplementary Information

Laurent Hébert-Dufresne, Antoine Allard, Jean-Gabriel Young and Louis J. Dubé

http://dynamica.phy.ulaval.ca/

*Département de Physique, de Génie Physique et d'Optique*
*Université Laval, Québec, Québec, Canada G1V0A6*

## Contents

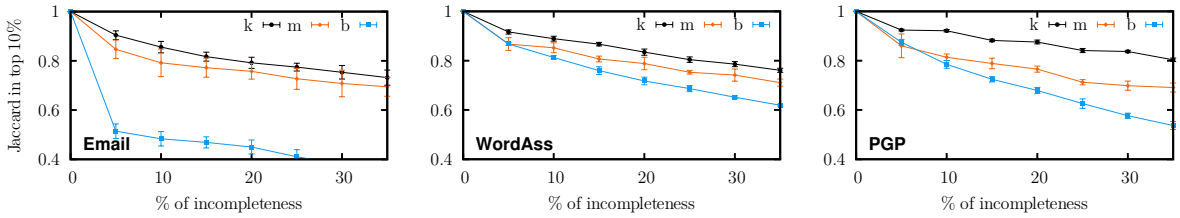
# 1 Supplementary discussions on methods

## Local vs global measures

We differentiate between these two types of measures by the information required to compute them. If this information (per node) is independent of total system size, the measure is considered local; whereas a global measure requires information scaling with system size (often a complete description). For the four properties studied in this paper, we thus consider that:

1. degree is a local measure, as only the number of neighbours of a node is required;
2. membership is a local measure, as the chosen algorithm only requires the neighbourhood of one given node and that of its neighbours to estimate its membership number;
3. coreness is a global measure, as a node's coreness depends on the coreness of its neighbours which in turn depend on the coreness of their neighbours and so on;
4. betweenness centrality is a global measure, since it is calculated by considering the shortest paths between a given node and all of the other nodes in the network.

For obvious reasons, *local measures are less sensitive to incomplete or incorrect information*. Adding, removing or rewiring a link only affects the degree or membership of nodes directly in the neighbourhood of the modification; whereas the same alterations can potentially affect the coreness or betweenness centrality of nodes anywhere in the network through cascading effects. Consequently, measures based on shorter-range information are always more robust to missing information, and sometimes quite significantly, as seen on Suppl. Fig. 1.

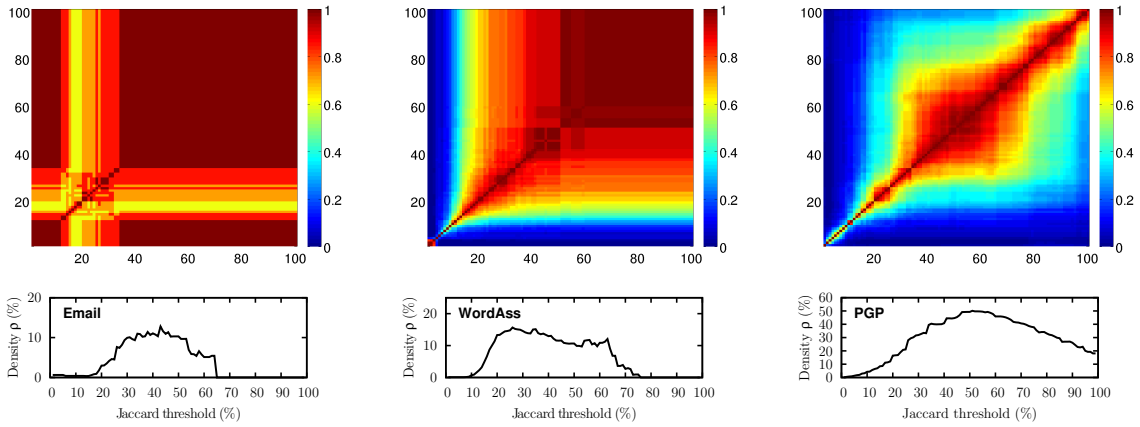



*Supplementary Figure 1:* Robustness of measured from the micro (degree $k$), meso (memberships $m$) and macro (betweenness centrality $b$) scales when information on the network is removed. The robustness is here measured by comparing (with a Jaccard coefficient) the ensemble of nodes identified as being in the top 10% of nodes when a certain fraction of links are randomly removed (horizontal axis) as opposed to the ensemble obtained by considering the complete data.

## Community detection

As mentioned above the *link clustering algorithm* of Ahn *et al.* was chosen in part because it can perform well (and at times even better) using local information instead of the entire network.[1] While we always partitioned the network globally, by setting a resolution threshold, the identification of structural hubs is very robust to this global threshold (see Suppl. Fig. 2). More importantly, this algorithm groups links stemming from a given node in a community based on the similarity of the two neighbourhoods reached through them. Hence, it evaluates the redundancy in second neighbourhoods (how many of my second neighbours are neighbours of more than one of my neighbours?). This redundancy (or overlap) can then serve as an appropriate measure to gauge the major impact of community structure on an epidemic process, namely the loss of potential new infections due to clustering. The link clustering algorithm therefore provides a well-defined method to quantify this loss.

[1] Y.-Y. Ahn, J. P. Bagrow, & S. Lehmann, Link communities reveal multiscale complexity in networks, *Nature*, vol. 466, p.761-764, 2010. See also the corresponding Supplementary Information.




*Supplementary Figure 2:* Robustness of our ability to identify structural hubs (1% of nodes with the most memberships) as the global resolution varies (the final partition is chosen to maximize the community density shown at the bottom). The color map represents the Jaccard coefficient, i.e. the similarity of ensembles, as measured between the structural hubs identified with two different resolution parameter. These ensembles are always very similar when avoiding extreme resolutions (e.g. low density). Note that in the color map, yellow corresponds to a Jaccard coefficient of 0.6 which implies that 75% of the same structural hubs were found by both partitions.

## Supplementary simulation details

**SIS.** All nodes are initially infectious and we relax the system by iterating a discrete time propagation simulation using time step $\Delta t$ chosen such that $\alpha\Delta t$ and $\beta\Delta t$ are less than $10^{-3}$:

i. at each $\Delta t$, every susceptible neighbour $S$ of every infectious individual $I$ is infected with probability $\alpha\Delta t$;
ii. at each $\Delta t$ every infectious individual $I$ recovers with probability $\beta\Delta t$;

the steady-state is averaged over multiple independent simulations to minimize the standard deviation (due to network structure and finite size).

**SIR.** A single node is randomly infected and the following stochastic process is iterated until no infectious nodes remain:

i. $I$ nodes infect each of their $S$ neighbours with probability $T$ and then recover.

The final state, considering only epidemics larger than 1% of the system size, is averaged over multiple independent simulations to minimize the standard deviation (due to network structure and finite size).

# 2 Theoretical modelling

The conclusion drawn in the main text were validated using synthetic networks. In this Section, we present how these synthetic networks are generated. Furthermore, we describe the mathematical framework used to calculate the final outcome of the SIR dynamics on these networks. Finally, we give the parameters used for the main results.

## Synthetic networks

The synthetic networks considered are a clustered and multitype generalisation of the Configuration Model.[2] In these networks, nodes are connected either through single links or through motifs (see Fig. 7 in the main text for an example). Motifs are composed of $M$ nodes which are all connected to one another, and a node belongs to $i$ motifs and has $j$ single links with probability $p(i, j)$. This node therefore has a degree ($k$) equal to $(M-1)i+j$ and a membership ($m$) equal to $i+j$.

Networks are generated using a stub pairing scheme: a node belonging to $i$ motifs and having $j$ single links has $i$ "motif stubs" and $j$ "link stubs". Groups and single links are then formed by randomly choosing $M$ motif stubs and 2 link stubs, respectively, and then by linking the corresponding nodes to one another. This last step is repeated until none of the motif and link stubs remain. The distribution $\{p(i, j)\}_{i,j\in\mathbb{N}}$ therefore defines a random network ensemble, and the results obtained in this Section are averaged over this ensemble.

## Mathematical formalism

As there exists a mapping—under simple assumptions—between the SIR dynamics and bond percolation on networks[3,4]. To calculate the outcome of the SIR dynamics on the networks just described, we use a previously published formalism[2] in which we add the possibility for nodes to exist with a given probability (i.e., site percolation) to simulate immunization strategies. We only give a short outline of this theoretical model as a general and more formal description will be the subject of a subsequent publication.

For each pair $(i, j)$ such that $p(i, j) \neq 0$ we assign a node type denoted by $\{i, j\}$ (the set of such pairs is noted $\mathcal{M}$). As the pair $(i, j)$ is the only information available about the nodes, assigning one node type per pair allows us to simulate very detailed immunization strategies. Indeed, we define $q_{\{i,j\}}$ as the probability for a type-$\{i, j\}$ node to be immunized; the simulated immunization strategy is therefore encoded in the set of probabilities $\{q_{\{i,j\}}\}_{\{i,j\}\in\mathcal{M}}$. Also, as explained in the main text, the infectious agent propagates from an infected node to a susceptible neighbour with probability $T$. From a percolation point of view, $1 - q_{\{i,j\}}$ is the occupation probability of type-$\{i, j\}$ sites (nodes) and $T$ is the occupation probability of bonds (links).

### Solving site/bond percolation in motifs

The mathematical formalism that we have developed relies on probability generating functions (PGFs) and therefore implicitly requires the networks under consideration to have a tree-like structure. As the

[2] A. Allard, L. Hébert-Dufresne, P.-A. Noël, V. Marceau, and L. J. Dubé (2012). Bond percolation on a class of correlated and clustered random graphs. *J. Phys. A*, 45(40):405005.

[3] M. E. J. Newman (2002). Spread of epidemic disease on networks. Phys. Rev. E, 66(1):016128.

[4] E. Kenah and J. Robins (2007). Second look at the spread of epidemics on networks. Phys. Rev. E, 76(3):1-12.

networks we consider contain motifs, which clearly do not comply with that assumption, we need to solve the bond and site percolation within motifs beforehand.

As previously shown[5], the bond percolation outcome—the distribution of the number of nodes that can be reached by following links from an initial node—can be exactly obtained by iterating a set of simple equations. We denote $\boldsymbol{n}$ the $|\mathcal{M}|$-tuple whose elements[6] $n_{\{i,j\}}$ correspond to the number of nodes of each type (i.e., there are $n_{\{i,j\}}$ type-$\{i, j\}$ nodes). For the remaining, each boldfaced variable will correspond to such $|\mathcal{M}|$-tuple.

Let us define $Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n})$ as the probability to find a component of $\boldsymbol{l}$ nodes in a motif of size (and composition) $\boldsymbol{n}$ from an initial type-$\{i, j\}$ node. Although nodes are initially all connected to one another in motifs, we are interested in the number of nodes (and their type) that can be reached from an initial node when links are followed with a probability $T$ (bond percolation). Following previous work[5], $Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n})$ is obtained by iterating

$$Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n}) = Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{l}) \prod_{\{i',j'\}\in\mathcal{M}} \binom{n_{\{i',j'\}} - \delta_{ii'}\delta_{jj'}}{l_{\{i',j'\}} - \delta_{ii'}\delta_{jj'}} \prod_{\{i'',j''\}\in\mathcal{M}} (1-T)^{n_{\{i',j'\}}(n_{\{i'',j''\}} - l_{\{i'',j''\}})} \tag{2.1a}$$

$$Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{l}) = 1 - \sum_{\boldsymbol{m}<\boldsymbol{l}} Q_{\{i,j\}}(\boldsymbol{m}|\boldsymbol{l}) \tag{2.1b}$$

from the initial condition $Q_{\{i,j\}}(\boldsymbol{\delta_{\{i,j\}}}|\boldsymbol{\delta_{\{i,j\}}})$, where $\delta_{ij}$ is the Kronecker delta, and where $\boldsymbol{\delta_{\{i,j\}}}$ is an $|\mathcal{M}|$-tuple whose elements are all equal to 0 except for the $\{i, j\}$-th one that is equal to one. The sum in Eq. (2.1b) is over all $\boldsymbol{m}$ such that $m_{\{i,j\}} \leq l_{\{i,j\}}$ for every node type $\{i, j\}$ but with the additional constraint that $\boldsymbol{m} \neq \boldsymbol{l}$. The initial condition simply states that the probability of finding a component of one type-$\{i, j\}$ node from a type-$\{i, j\}$ node in a motif containing only one type-$\{i, j\}$ node is one (the initial node is always included in the size of the component). Then, Eqs. (2.1) iteratively increase the size of the motif and compute the size distribution along the way until the complete distribution for a motif of the desired size (and composition) is obtained.

Should we be interested in studying bond percolation on motifs solely, we would keep the distribution $\{Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n})\}$ for a given size $\boldsymbol{n}$, and discard the distributions for motifs of intermediate size obtained while iterating Eqs. (2.1). These intermediate distributions can however be used to *exactly* predict the distribution of the number of nodes that can be reached by following links from an initial node in motifs where links *and* nodes exist with given probabilities (bond and site percolation). Indeed, as each node exists independently with a given probability, the probability for a motif of original size $\boldsymbol{n}$ to be of size $\boldsymbol{m}$ after the random removal of its nodes is simply

$$W_{\{i,j\}}(\boldsymbol{m}|\boldsymbol{n}) = \prod_{\{i',j'\}\in\mathcal{M}} \binom{n_{\{i',j'\}} - \delta_{ii'}\delta_{jj'}}{m_{\{i',j'\}} - \delta_{ii'}\delta_{jj'}} \left[1 - q_{\{i',j'\}}\right]^{m_{\{i',j'\}} - \delta_{ii'}\delta_{jj'}} \left[q_{\{i',j'\}}\right]^{n_{\{i',j'\}} - m_{\{i',j'\}}}, \tag{2.2}$$

where we assume that the initial type-$\{i, j\}$ exists. Then, the probability for a type-$\{i, j\}$ node to lead to a component of size $\boldsymbol{l}$ in a motif of original size $\boldsymbol{n}$ but whose links and nodes have been randomly removed is simply

$$P_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n}) = \sum_{\boldsymbol{m}=\boldsymbol{\delta_{\{i,j\}}}}^{\boldsymbol{n}} Q_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{m}) W_{\{i,j\}}(\boldsymbol{m}|\boldsymbol{n}) \,. \tag{2.3}$$

[5] A. Allard, L. Hébert-Dufresne, P.-A. Noël, V. Marceau, and L. J. Dubé (2012). Exact solution of bond percolation on small arbitrary graphs. EPL, 98(1):16001.

[6] $|\mathcal{M}|$ is the number of elements (i.e., the cardinality) of the set $\mathcal{M}$, hence the number of node types.

The site and bond percolation can therefore be exactly solved for motifs by iterating Eqs. (2.1)–(2.3). As expected, if applied to the simplest motifs, i.e. links, the type-$\{i, j\}$ node at the other end of the link will be reached with probability $T(1-q_{\{i,j\}})$ and will not be reached with probability $1-T(1-q_{\{i,j\}})$.

As motifs and single links are built by randomly matching stubs, the probability for a type-$\{i, j\}$ node to appear in a motif [a link] is proportional to $[ip(i, j)]$ $[jp(i, j)]$. The probability for a motif to be composed by $\boldsymbol{n}$ nodes is given by the multinomial distribution

$$R^{(\mathrm{m})}(\boldsymbol{n}) = \frac{M!}{\left[\sum_{\{i,j\}\in\mathcal{M}} ip(i,j)\right]^M} \prod_{\{i,j\}\in\mathcal{M}} \frac{[ip(i,j)]^{n_{\{i,j\}}}}{n_{\{i,j\}}!} \,. \tag{2.4a}$$

The same applies for the composition of links

$$R^{(\mathrm{l})}(\boldsymbol{n}) = \frac{2!}{\left[\sum_{\{i,j\}\in\mathcal{M}} jp(i,j)\right]^2} \prod_{\{i,j\}\in\mathcal{M}} \frac{[jp(i,j)]^{n_{\{i,j\}}}}{n_{\{i,j\}}!} \,. \tag{2.4b}$$

Finally, for the purpose of calculating the outcome of the bond and site percolation on the synthetic networks, let us define the two following generating functions:

$$\theta^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{x}) = \sum_{\boldsymbol{n}} \frac{n_{\{i,j\}}R^{(\mathrm{m})}(\boldsymbol{n})}{\sum_{\boldsymbol{n}'} n'_{\{i,j\}}R^{(\mathrm{m})}(\boldsymbol{n'})} \left[\sum_{\boldsymbol{l}=\boldsymbol{\delta}_{\{i,j\}}}^{\boldsymbol{n}} P^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n}) \prod_{\{i',j'\}\in\mathcal{M}} [x_{\{i',j'\}}]^{l_{\{i',j'\}}-\delta_{ii'}\delta_{jj'}}\right] \tag{2.5a}$$

$$\theta^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{y}) = \sum_{\boldsymbol{n}} \frac{n_{\{i,j\}}R^{(\mathrm{l})}(\boldsymbol{n})}{\sum_{\boldsymbol{n}'} n'_{\{i,j\}}R^{(\mathrm{l})}(\boldsymbol{n'})} \left[\sum_{\boldsymbol{l}=\boldsymbol{\delta}_{\{i,j\}}}^{\boldsymbol{n}} P^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{l}|\boldsymbol{n}) \prod_{\{i',j'\}\in\mathcal{M}} [y_{\{i',j'\}}]^{l_{\{i',j'\}}-\delta_{ii'}\delta_{jj'}}\right] \tag{2.5b}$$

where the superscript "m" (resp. "l") indicate that the quantities have been solved for motifs (resp. links). In other words, the function $\theta^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{x})$ generates the probability distribution for the number of nodes of each type that can be reached from a type-$\{i, j\}$ node in a random motif [i.e., whose composition is averaged over $R^{(\mathrm{m})}(\boldsymbol{n})$]. Specifically, the coefficient in front of $x^s_{\{i',j'\}}$ in Eq. (2.5a) is the probability of reaching $s$ type-$\{i', j'\}$ nodes from a type-$\{i, j\}$ node in a random motif. The same applies to $\theta^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{y})$.

### Calculating the average fate of an outbreak

We are now in a position to solve the bond and site percolation on the synthetic networks defined previously. For the purpose of the present study, we are interested in the quantity $R_{\mathrm{f}}$: the relative size of the extensive (giant) component. To highlight the nontrivial effect of immunization[7], $R_{\mathrm{f}}$ is expressed in terms of the fraction of the *existing* nodes (i.e., $1-\varepsilon$) that are part of the giant component.

It is convenient to introduce the following function

$$g_{\{i,j\}}(\boldsymbol{x}, \boldsymbol{y}) = \left[\theta^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{x})\right]^i \left[\theta^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{y})\right]^j \tag{2.6}$$

generating the distribution of the number of nodes of each type that are in the immediate neighbourhood of a type-$\{i, j\}$ node. The *immediate neighbourhood* refers to the nodes to which the type-$\{i, j\}$

[7] The relative size of the giant component cannot exceed $1-\varepsilon$ on networks for which a fraction $\varepsilon$ of the nodes has been removed. This reduction in size obviously occurs during any immunization strategy and, for comparison purposes, must be taken into account.

node is connected either via its single links or via its motifs. Similarly, we define

$$f^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{x},\boldsymbol{y}) = \left[\theta^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{x})\right]^{i-1}\left[\theta^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{y})\right]^{j} \tag{2.7a}$$

$$f^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{x},\boldsymbol{y}) = \left[\theta^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{x})\right]^{i}\left[\theta^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{y})\right]^{j-1} \tag{2.7b}$$

generating the distribution of the number of nodes of each type that are in the immediate neighbourhood of a type-$\{i, j\}$ node that has been reached by either one of its single links or one of the motifs it is a part of (if applicable). In other words, these functions generate the excess degree distribution.

To calculate $R_{\mathrm{f}}$, let us define $a_{\{i,j\}}$ as the probability that a link to a type-$\{i, j\}$ node *does not* lead to the giant component. Similarly, we define $b_{\{i,j\}}$ as the probability that a type-$\{i, j\}$ node reached through a motif does not lead to the giant component. Due to the effective tree-like structure of the networks—recall that the outcome of percolation on the motifs has already been solved—$a_{\{i,j\}}$ and $b_{\{i,j\}}$ must satisfy the following self-consistency relations

$$a_{\{i,j\}} = f^{(\mathrm{m})}_{\{i,j\}}(\boldsymbol{a},\boldsymbol{b}) \tag{2.8a}$$

$$b_{\{i,j\}} = f^{(\mathrm{l})}_{\{i,j\}}(\boldsymbol{a},\boldsymbol{b})\,. \tag{2.8b}$$

Put simply, these equations state that if a type-$\{i, j\}$ node reached from either a link or a motif does not lead to the giant component, then neither should the nodes that can be reached from it. The probability that a type-$\{i, j\}$ node *is* part of the giant component is then $1 - g_{\{i,j\}}(\boldsymbol{a},\boldsymbol{b})$. The probability that a randomly existing node is part of the giant component—which corresponds to its size as well—is therefore

$$R_{\mathrm{f}} = \sum_{\{i,j\}\in\mathcal{M}} \frac{(1-q_{\{i,j\}})p(i,j)\left[1-g_{\{i,j\}}(\boldsymbol{a},\boldsymbol{b})\right]}{\sum_{\{i',j'\}\in\mathcal{M}}(1-q_{\{i',j'\}})p(i',j')}\,. \tag{2.9}$$

The theoretical predictions (lines) on Fig. 8 in the main text were obtained by solving Eqs. (2.1)–(2.9) for various values of $T$ and $\{q_{\{i,j\}}\}_{\{i,j\}\in\mathcal{M}}$. Comparison with results obtained from numerical simulations (symbols) confirms the validity of our theoretical model.

## Parameters used for theoretical calculations

Table 1 shows the distribution $\{p(i, j)\}_{i,j\in\mathbb{N}}$ used for the synthetic networks considered in the main text. It also gives the degree ($k_{\{i,j\}}$) and the membership ($m_{\{i,j\}}$) of each node type $\{i, j\} \in \mathcal{M}$. Motifs were composed of $M = 4$ nodes, and numerical simulation results (symbols on Fig. 8) were averaged over $5\times10^5$ realisations of networks of $2.5\times10^5$ nodes. For node types with $k_{\{i,j\}} > 2$, we let sets of $M-1$ links to either be part of cliques of $M$ nodes or be single links in order to avoid unintended degree correlations[8]. For a given fraction $\varepsilon$ of the nodes to immunize, we have

$$\sum_{\{i',j'\}\in\mathcal{M}} q_{\{i',j'\}}p(i',j') = \varepsilon\,. \tag{2.10}$$

The probabilities $q_{\{i,j\}}$ are chosen to satisfy this condition and in decreasing order of degree or membership.

[8] I.Z. Kiss & D.M. Green (2008), Comment on "Properties of highly clustered networks", *Phys. Rev. E*, 78:048101.

| $\{i, j\}$ | $p(i, j)$ | $k_{\{i,j\}}$ | $m_{\{i,j\}}$ |
|---|---|---|---|
| $\{0, 1\}$ | 0.43930 | 1 | 1 |
| $\{0, 2\}$ | 0.13179 | 2 | 2 |
| $\{0, 9\}$ | 0.00712 | 9 | 9 |
| $\{1, 3\}$ | 0.25831 | 6 | 4 |
| $\{1, 6\}$ | 0.04982 | 9 | 7 |
| $\{2, 3\}$ | 0.02325 | 9 | 5 |
| $\{3, 0\}$ | 0.09041 | 9 | 3 |

*Supplementary Table 1:* Distribution $\{p(i, j)\}_{i,j\in\mathbb{N}}$ used for the synthetic networks discussed in the main text. The degree and the membership of each node type is computed according to $k_{\{i,j\}} = (M - 1)i + j$ and $m_{\{i,j\}} = i + j$, respectively, with $M = 4$.

## 3 Introduction to the supplementary results

The last sections of this Supplementary Information present a more complete view of the results obtained on empirical networks and are structured as follows. Each section covers one of the 17 datasets used in the study. Firstly, a brief discussion on the nature of each network is given, along with:

- the number of nodes ($N$), of links ($L$) and the degree distribution ($k$ links per node);
- the maximal community density $\rho$ and corresponding Jaccard threshold $J_\rho$.
- the maximal values of degree $k$, coreness $c$, betweenness centrality $b$ and memberships $m$.

Secondly, correlations between degree, betweenness centrality, coreness and memberships are quantified using Spearman's rank correlation coefficient (defined below). We leave to the reader to observe how, given the correlation coefficient between memberships ranking and degree ranking, along with the mean community density, one can somewhat predict if the membership-based immunization will be more or less efficient than the degree-based version. Finally, the results of all immunization methods (random or on the four measures) are presented for SIS and SIR dynamics for a virulence (probability of disease transmission) close and far from the network's epidemic threshold.

## Spearman's rank correlation coefficient

The Spearman's rank correlation coefficient quantifies the statistical dependence of two different orderings of the same set of items (nodes) on a scale of $-1$ (perfectly anti-correlated) to 1 (perfectly correlated).[9]

Consider $x_i$ to be the rank of item $i$ according to measure $X$, and $y_i$ to be the rank of the same item according to a different measure $Y$. If for example, 10 items have the same score according to $X$ and would otherwise be ranked from $x_j$ to $x_{j+9}$, they are all given the rank $\left[\sum_{k=0}^{9} x_{j+k}\right]/10$. The Spearman's rank correlation coefficient $\sigma(X, Y)$ is then given by:

$$\sigma(X, Y) = \left[\sum_i (x_i - \bar{x})\,(y_i - \bar{y})\right] \Big/ \left[\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2\right]^{1/2},$$

where $\bar{u}$ is the average rank according to measure $U$ (the mean of $\{u_i\}$).

[9] C. Spearman (1904), The proof and measurement of association between two things, *Amer. J. Psychol.*, 15:72101.

# 4 arXiv co-authorship

The cond-mat arXiv database uses articles published at *http://arxiv.org/archive/cond-mat* between April 1998 and February 2004. In this network, an article written by $n$ co-authors contributes to a link of weight $(n-1)$ between every pair of authors. The unweighted network was obtained by deleting all links with a weight under the selected threshold of 0.1.[10]

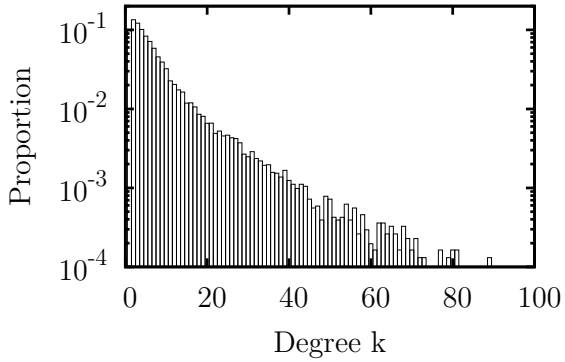
*Supplementary Table 2:* arXiv statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 30561 | 125959 | 191 | 15 | $6.9e+06$ | 127 | 0.35 |

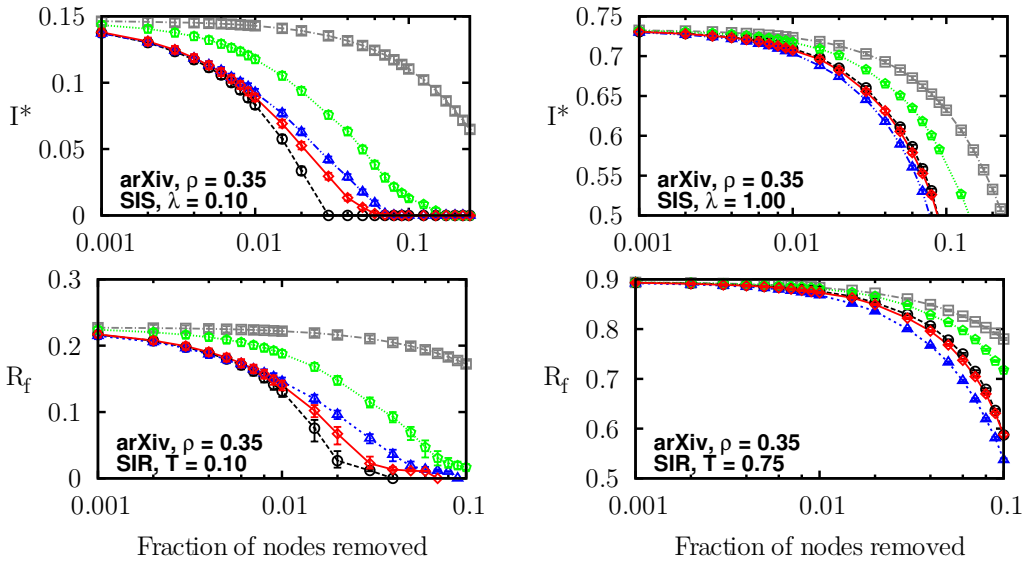
*Supplementary Table 3:* arXiv correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.7766 | 0.7717 | 0.6639 | 0.7461 | 0.9411 | 0.5388 |

*Supplementary Figure 3:* arXiv degree distribution




*Supplementary Figure 4:* Intervention against epidemics on arXiv after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[10]Palla, G., Derenyi, I., Farkas, I. & Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435:814-818

## 5 Brightkite online social network

Brightkite was a location-based online social network where users could "check in" to the physical places they were visiting to connect with nearby friends. This datasets was obtained from a total of 4,491,143 check-ins over the period of Apr. 2008 - Oct. 2010.[11]

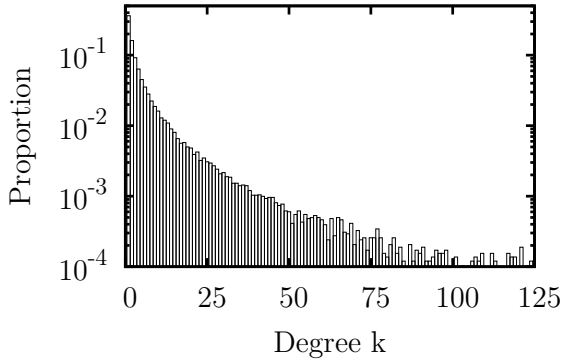
*Supplementary Table 4:* Brightkite statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 58228 | 214078 | 1134 | 52 | $2e+08$ | 1118 | 0.55 |

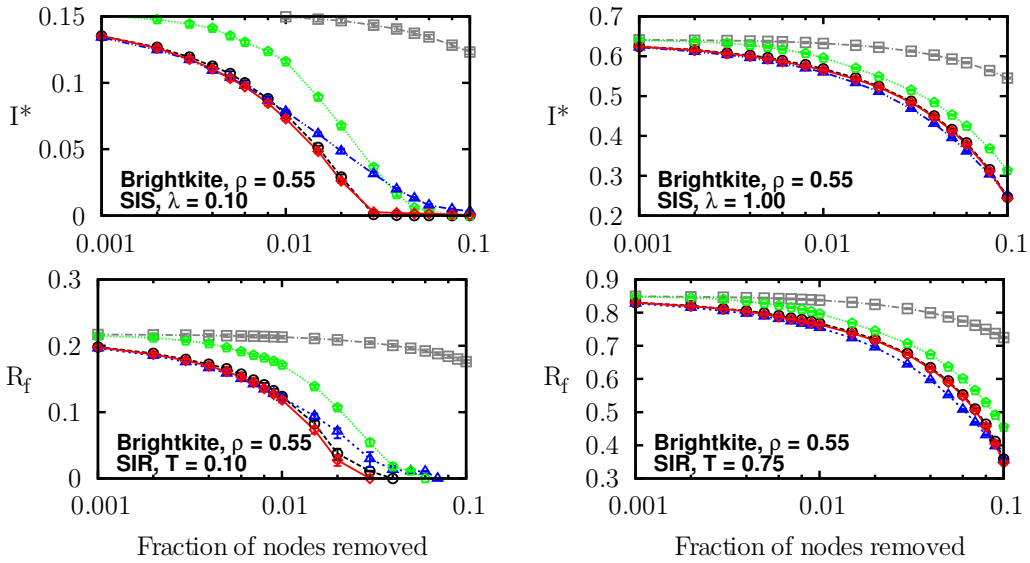
*Supplementary Table 5:* Brightkite correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9845 | 0.8919 | 0.9477 | 0.8822 | 0.9659 | 0.7767 |

*Supplementary Figure 5:* Brightkite degree distribution




*Supplementary Figure 6:* Intervention against epidemics on Brightkite after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[11] Cho, E., Myers, S.A. & Leskovec, J. (2011) Friendship and Mobility: User Movement in Location-Based Social Networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).*

# 6 University email exchange

Network of email communication between accounts from the University Rovira i Virgili.[12]

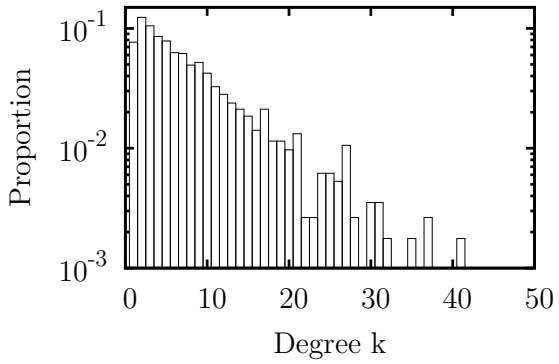
*Supplementary Table 6:* Email statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 1134 | 5143 | 1080 | 8 | $6.1e+05$ | 929 | 0.13 |

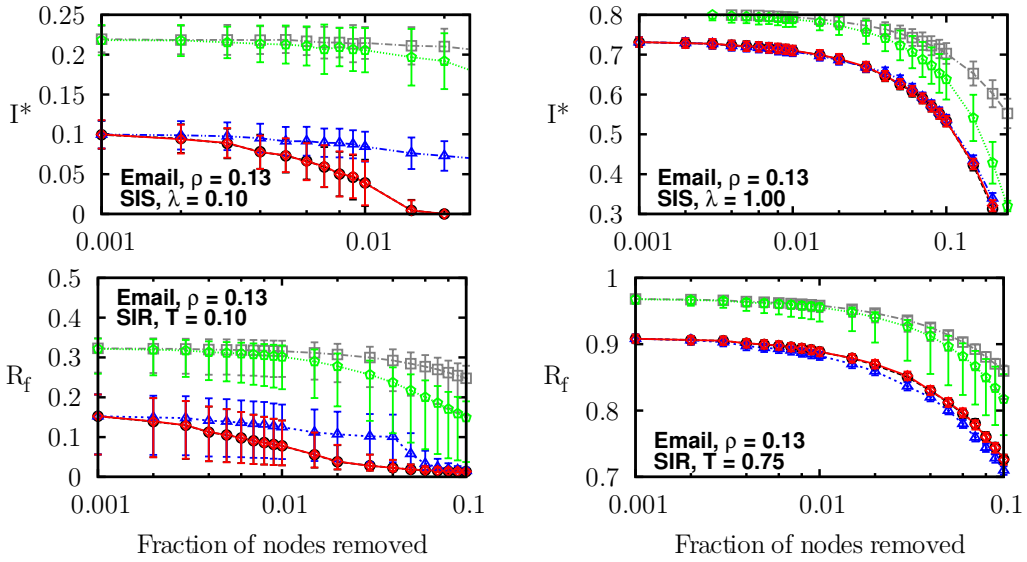
*Supplementary Table 7:* Email correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9900 | 0.9474 | 0.9560 | 0.9447 | 0.9613 | 0.8831 |

*Supplementary Figure 7:* Email degree distribution




*Supplementary Figure 8:* Intervention against epidemics on university email network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[12]Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F. & Arenas, A. (2003) Self-similar community structure in a network of human interactions. *Phys. Rev. E* 68:065103(R)

# 7 Enron email exchange

Network of email interchanges between all different Enron email addresses built from a dataset of around half million emails (made public by the Federal Energy Regulatory Commission).[13]

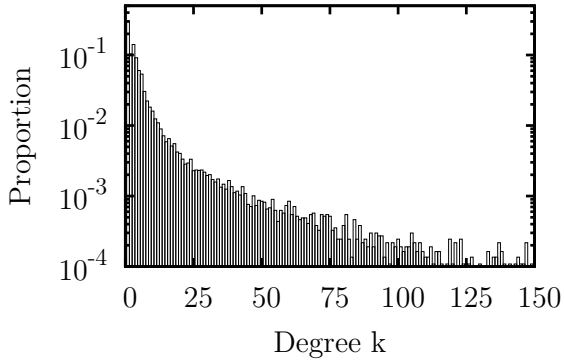
*Supplementary Table 8:* Enron statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 36692 | 183831 | 1383 | 43 | $4.3e+07$ | 1306 | 0.61 |

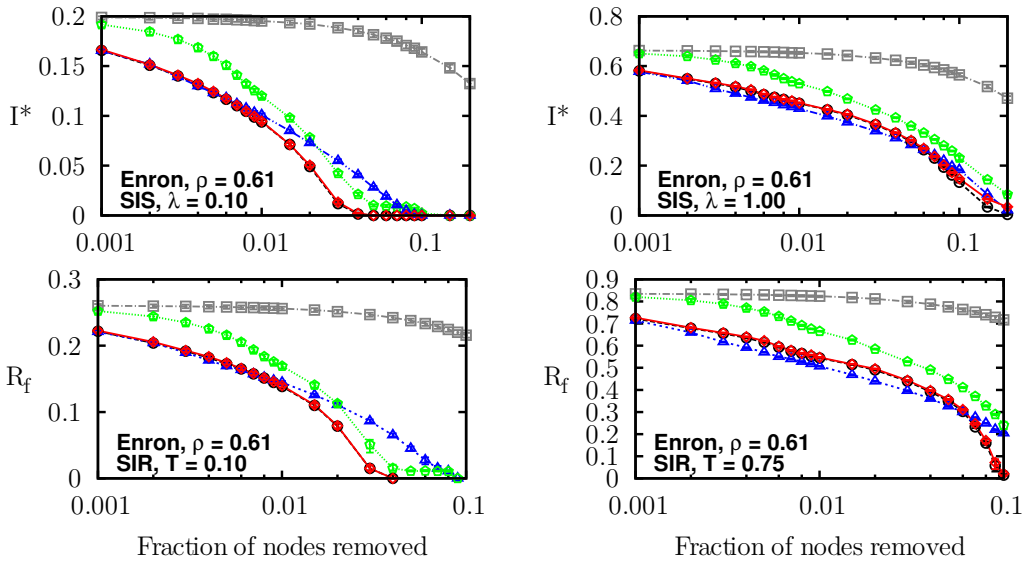
*Supplementary Table 9:* Enron correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9325 | 0.7567 | 0.9173 | 0.7585 | 0.9839 | 0.6862 |

*Supplementary Figure 9:* Enron degree distribution




*Supplementary Figure 10:* Intervention against epidemics on Enron email network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[13] Klimmt, B. & Yang, Y. (2004) Introducing the Enron corpus. *CEAS conference*.

## 8 Gnutella peer-to-peer network

A snapshot of the Gnutella peer-to-peer network, where nodes are hosts and edges connections, from August 30th 2002. The data is originally directed (files taken *from* one host *to* another), but was made undirected for this work.[14]

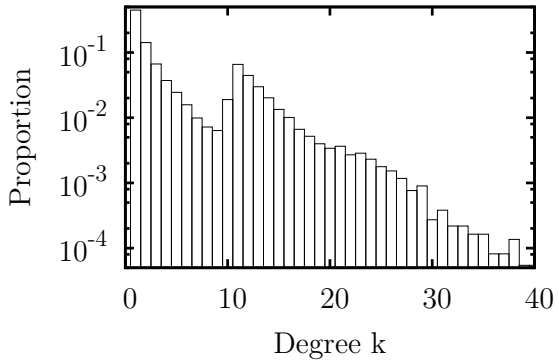
*Supplementary Table 10:* Gnutella statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 36682 | 88328 | 55 | 7 | $5.3e+06$ | 52 | 0.03 |

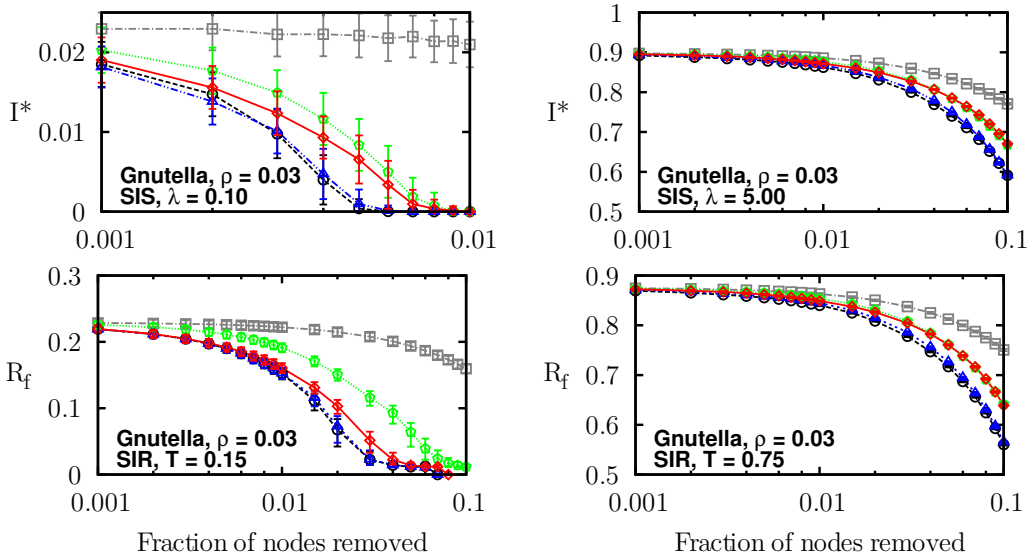
*Supplementary Table 11:* Gnutella correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9849 | 0.9848 | 0.9796 | 0.9925 | 0.9823 | 0.9743 |

*Supplementary Figure 11:* Gnutella degree distribution




*Supplementary Figure 12:* Intervention against epidemics on Gnutella network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[14]Ripeanu, M., Foster, I. & Iamnitchi, A. (2002) Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. *IEEE Internet Computing Journal* 6:50-57

# 9 Google weblinks

Directed network of hyperlinks between Google's webpages (considered undirected for this study).[15]

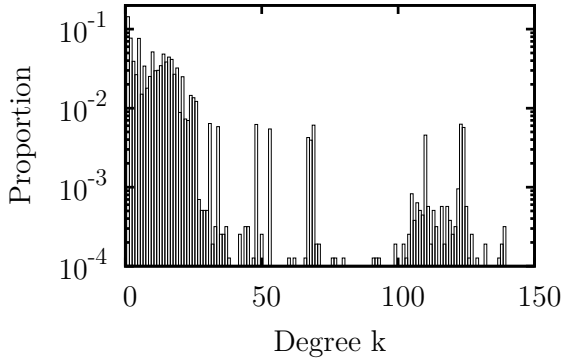
*Supplementary Table 12:* Google statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 15763 | 149456 | 11401 | 102 | $9.0e+07$ | 2883 | 0.49 |

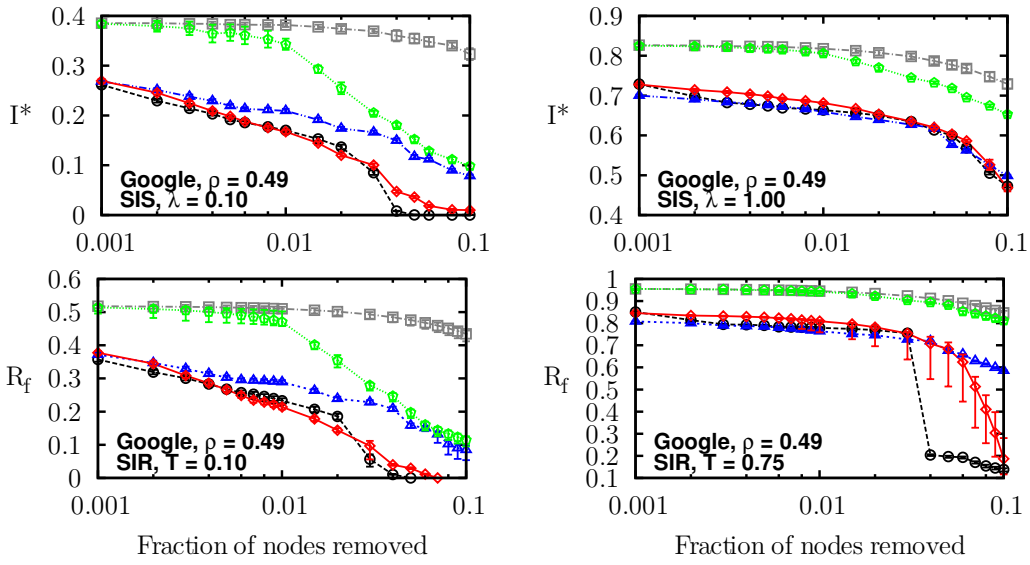
*Supplementary Table 13:* Google correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.8862 | 0.7941 | 0.8401 | 0.7735 | 0.9723 | 0.6995 |

*Supplementary Figure 13:* Google degree distribution




*Supplementary Figure 14:* Intervention against epidemics on Google network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[15]Palla, G., Farkas, I.J, Pollner, P., Derényi, I. & Vicsek, T. (2007) Directed network modules. *New. J. Phys.* 9:186

## 10 Gowalla social network

Gowalla is a location-based social networking website similar to Brightkite. This friendship network is undirected and composed from a total of 6,442,890 check-ins over the period of Feb. 2009 - Oct. 2010.[16]

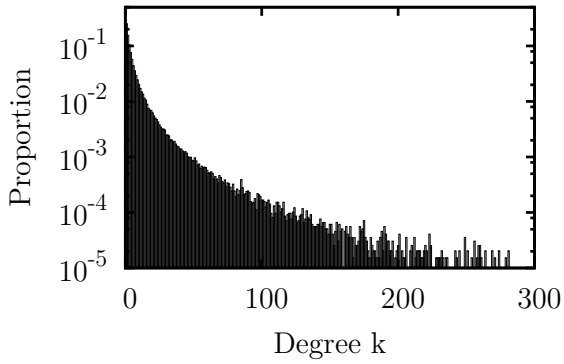
*Supplementary Table 14:* Gowalla statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 196591 | 950327 | 14730 | 51 | $6.3e+09$ | 14600 | 0.54 |

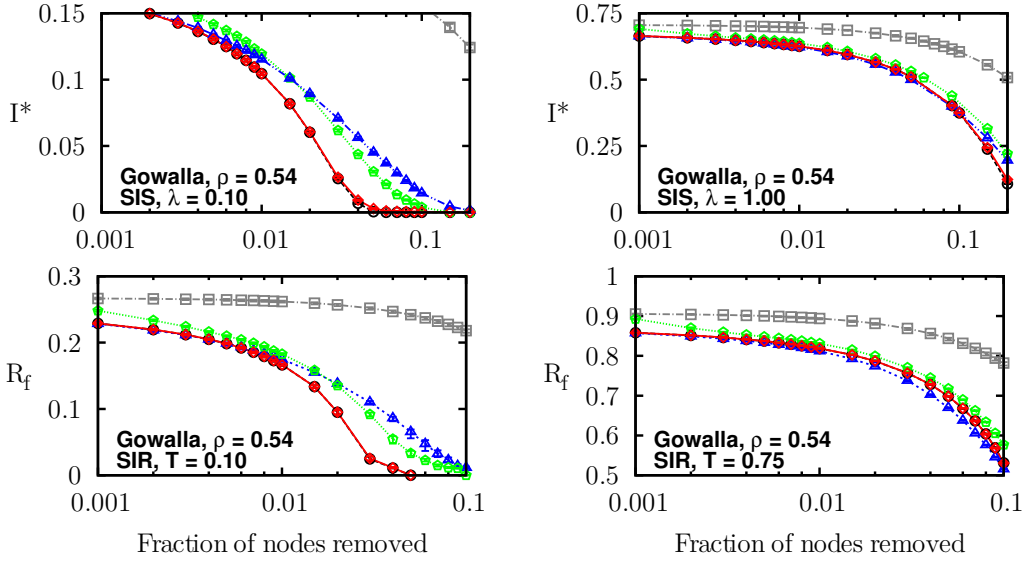
*Supplementary Table 15:* Gowalla correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9792 | 0.8363 | 0.9514 | 0.8311 | 0.9724 | 0.7309 |

*Supplementary Figure 15:* Gowalla degree distribution




*Supplementary Figure 16:* Intervention against epidemics on Gowalla network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[16]Cho, E., Myers, S.A. & Leskovec, J. (2011) Friendship and Mobility: Friendship and Mobility: User Movement in Location-Based Social Networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).*

## 11 Internet autonomous systems

This dataset is a symmetrized snapshot of the structure of the Internet at the level of autonomous systems, reconstructed from BGP tables posted at *archive.routeviews.org*. This snapshot was created by Mark Newman from data for July 22nd 2006.[17]

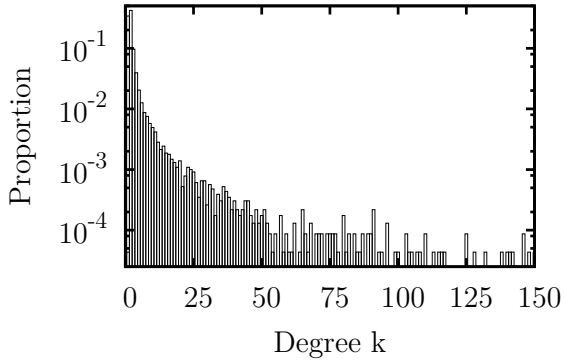
*Supplementary Table 16:* Internet statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 22963 | 48436 | 2390 | 25 | $3.8e+07$ | 1710 | $7e-4$ |

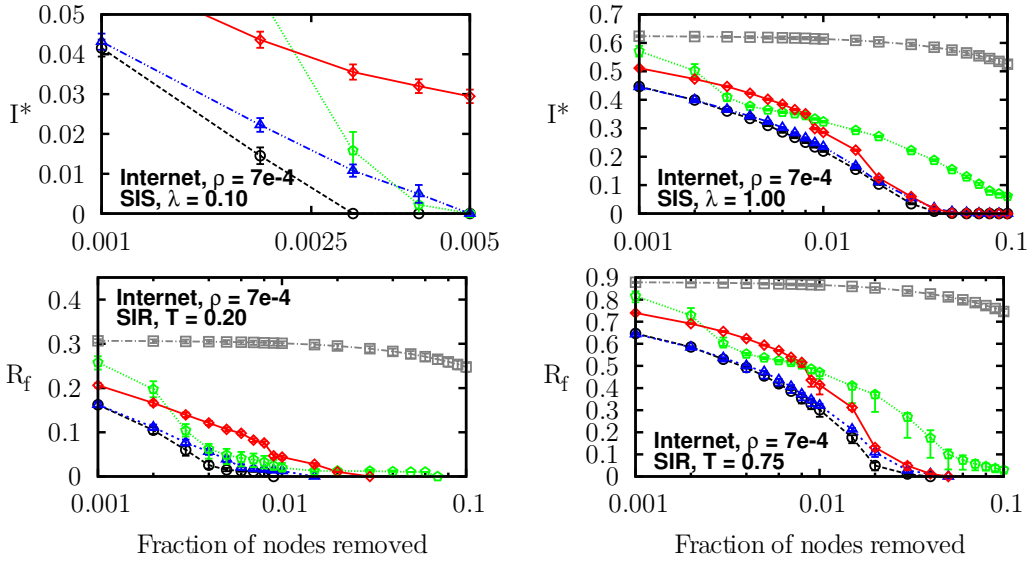
*Supplementary Table 17:* Internet correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9857 | 0.7933 | 0.9469 | 0.7807 | 0.9631 | 0.7079 |

*Supplementary Figure 17:* Internet degree distribution




*Supplementary Figure 18:* Intervention against epidemics on Internet network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[17]Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A. & Dubé, L.J. (2011) Structural Preferential Attachment: Network Organization beyond the Link. *Phys. Rev. Lett.* 107:158702

## 12 Internet Movie Database

This dataset details the co-acting network of for movies released after December 31st 1999 as compiled by IMDb.[18][19]

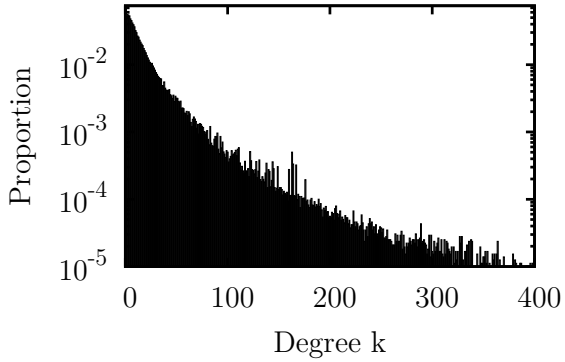
*Supplementary Table 18:* IMDb statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 716463 | 7665259 | 4625 | 192 | N/A | 2152 | 0.52 |

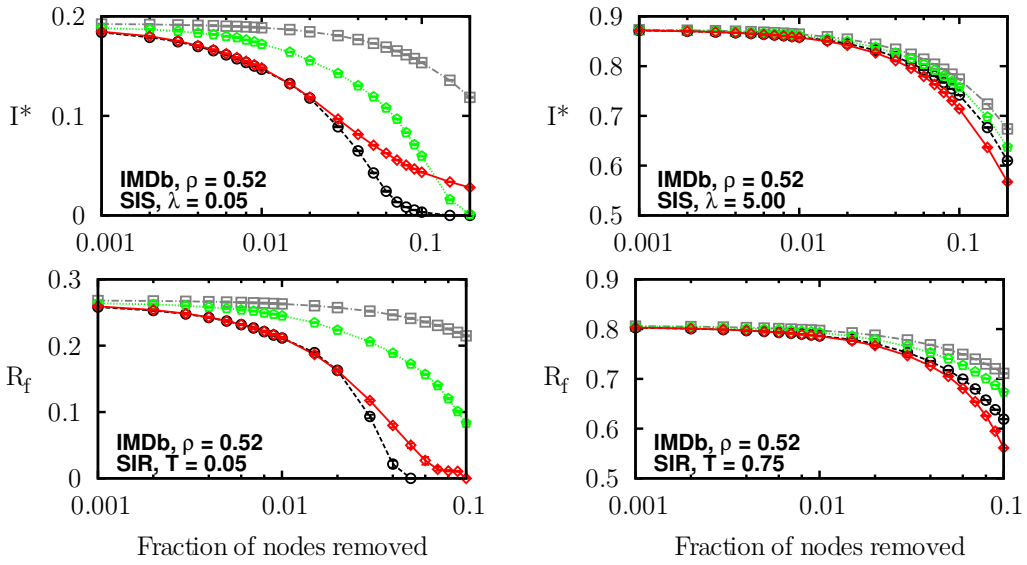
*Supplementary Table 19:* IMDb correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.6830 | N/A | 0.6186 | N/A | 0.9813 | N/A |

*Supplementary Figure 19:* IMDb degree distribution




*Supplementary Figure 20:* Intervention against epidemics on IMDb network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[18] http://www.imdb.com/

[19] Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A. & Dubé, L.J. (2011) Structural Preferential Attachment: Network Organization beyond the Link. *Phys. Rev. Lett.* 107:158702

## 13 MathSciNet co-authorship

Co-authorship network of MathSciNet before 2008.[20][21]

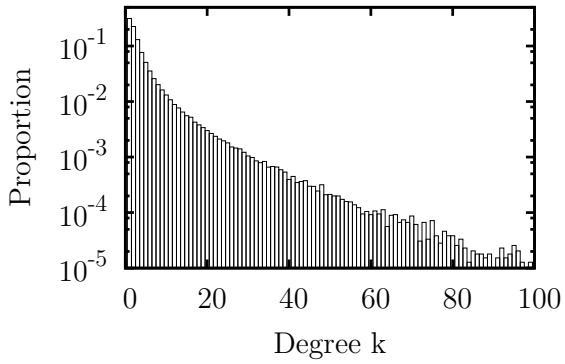
*Supplementary Table 20:* MathSci statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 391529 | 873775 | 496 | 24 | $1.9e+09$ | 485 | 0.40 |

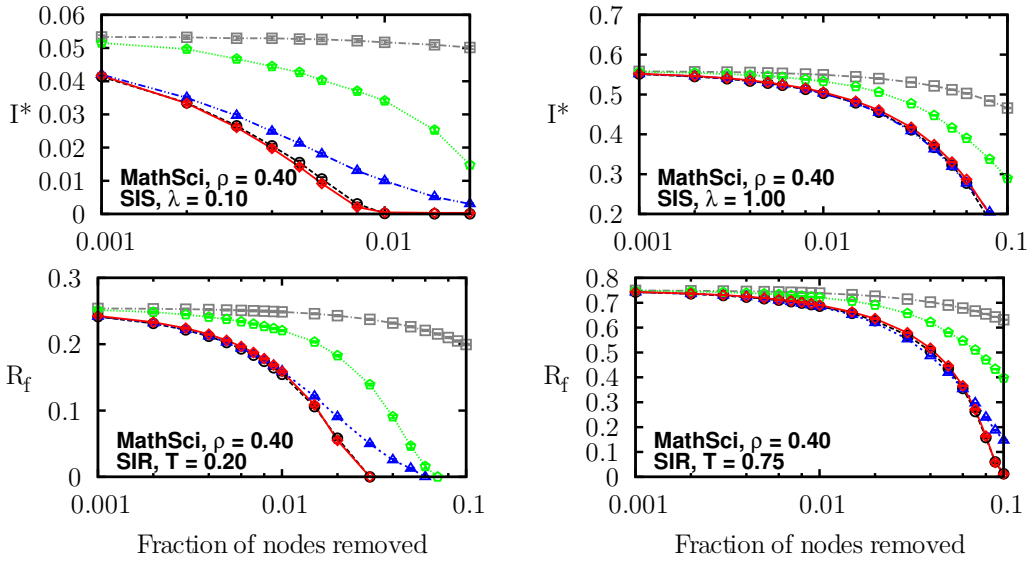
*Supplementary Table 21:* MathSci correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.8645 | 0.8320 | 0.7749 | 0.7835 | 0.9465 | 0.6200 |

*Supplementary Figure 21:* MathSci degree distribution




*Supplementary Figure 22:* Intervention against epidemics on MathSci network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[20] http://www.ams.org/mathscinet/

[21] Palla, G., Farkas, I.J., Pollner, P., Derényi, I. & Vicsek, T. (2008) Fundamental statistical features and self-similar properties of tagged networks. *New J. Phys.* 10:123026

## 14 Myspace online social network

Friendships between the first 100,000 users encountered while crawling Myspace accounts from September to October 2006 (excluding Tom Anderson, the cofounder of MySpace, which is connected to everyone).[22][23]

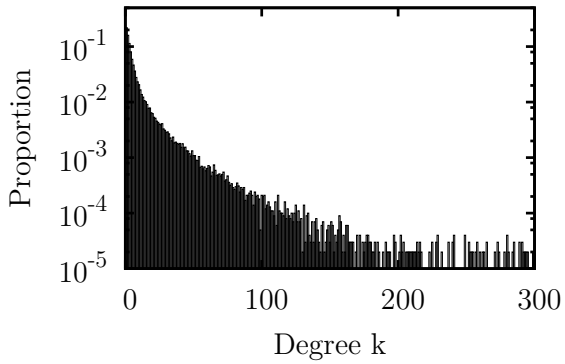
*Supplementary Table 22:* Myspace statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 100000 | 841224 | 59108 | 78 | $2.6e+09$ | 59102 | 0.77 |

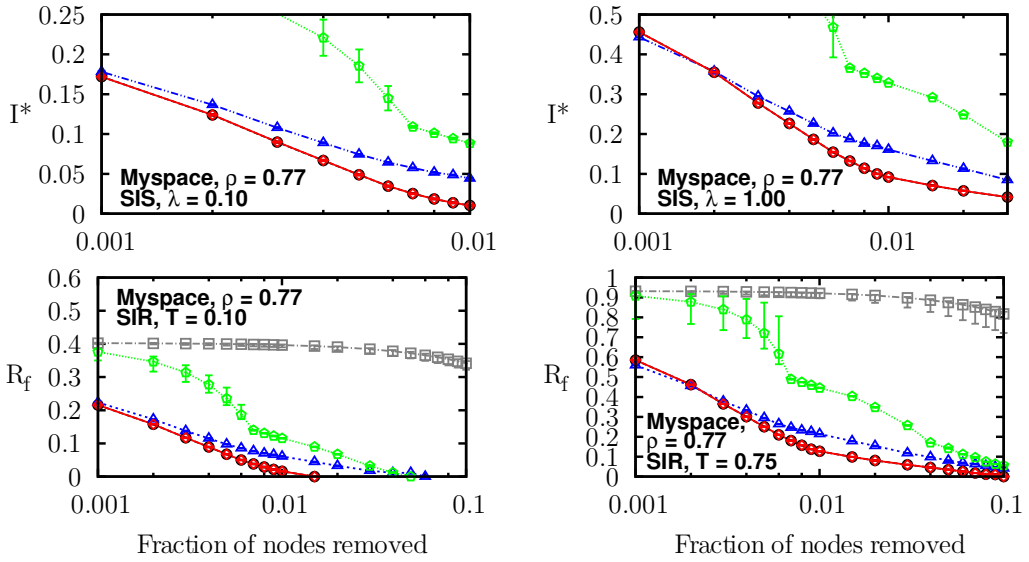
*Supplementary Table 23:* Myspace correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 1.0000 | 0.8667 | 0.9995 | 0.8667 | 0.9995 | 0.8662 |

*Supplementary Figure 23:* Myspace degree distribution




*Supplementary Figure 24:* Intervention against epidemics on Myspace after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[22] http://www.myspace.com/

[23] Ahn, Y.-Y., Han, S., Kwak, H., Moon, S. & Jeong, H. (2007) Analysis of Topological Characteristics of Huge Online Social Networking Services, *Proc. of International World Wide Web Conference*

## 15 Pretty-Good-Privacy data exchange

Dataset describing the giant component in the network of users of the Pretty-Good-Privacy algorithm for information exchange.[24]

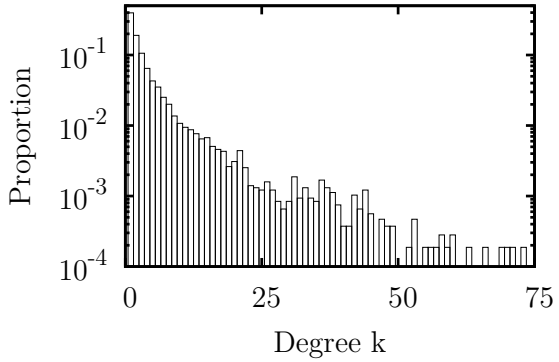
*Supplementary Table 24:* PGP statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 10680 | 24316 | 205 | 31 | $7.5e+06$ | 110 | 0.50 |

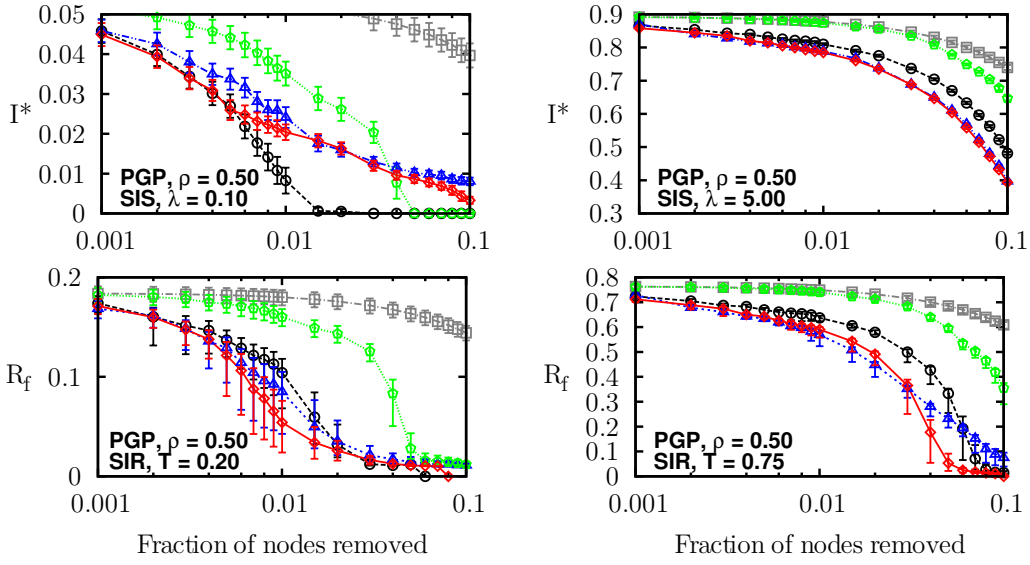
*Supplementary Table 25:* PGP correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.8862 | 0.8599 | 0.7256 | 0.7973 | 0.8973 | 0.5464 |

*Supplementary Figure 25:* PGP degree distribution




*Supplementary Figure 26:* Intervention against epidemics on the PGP network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[24]Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. (2004) Models of social networks based on social distance attachment. *Phys. Rev. E* 70:056122

## 16 Power grid

The topology of the Western States Power Grid of the United States as compiled by Duncan Watts and Steven Strogatz.[25]

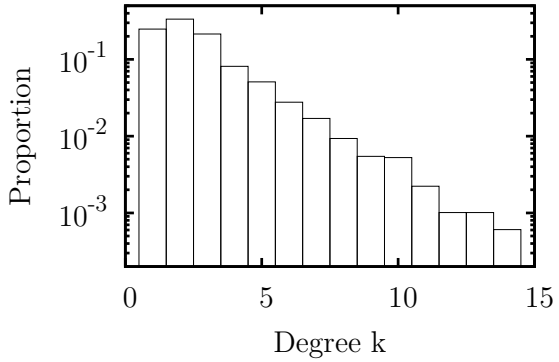
*Supplementary Table 26:* Power grid statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 4941 | 6594 | 19 | 5 | $3.5e+06$ | 18 | 0.49 |

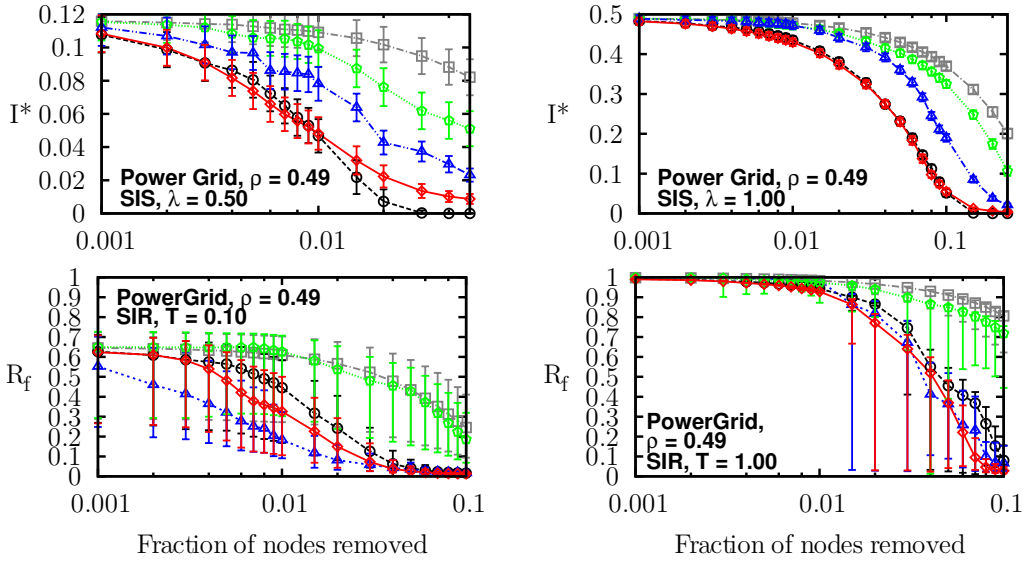
*Supplementary Table 27:* Power grid correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9192 | 0.8605 | 0.6191 | 0.8042 | 0.7342 | 0.5788 |

*Supplementary Figure 27:* Power grid degree distribution




*Supplementary Figure 28:* Intervention against epidemics on the power grid after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[25] Watts, D.J. & Strogatz, S.H. (1998) Collective dynamics of small-world networks. *Nature* 393:440-442

## 17 Protein interactions network

Protein-protein interactions (ProteinCore) in *S. cerevisiae* as listed by the Database of Interacting Proteins.[26][27]

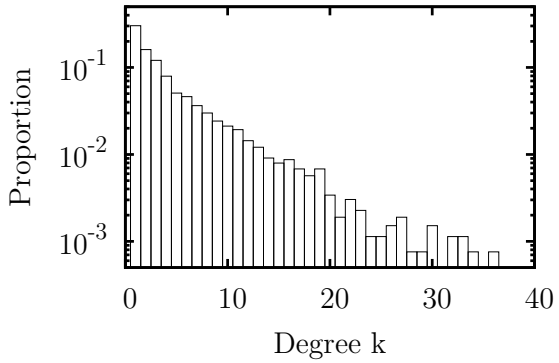
*Supplementary Table 28:* ProteinCore statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 2640 | 6600 | 111 | 8 | $4.0e+05$ | 71 | 0.32 |

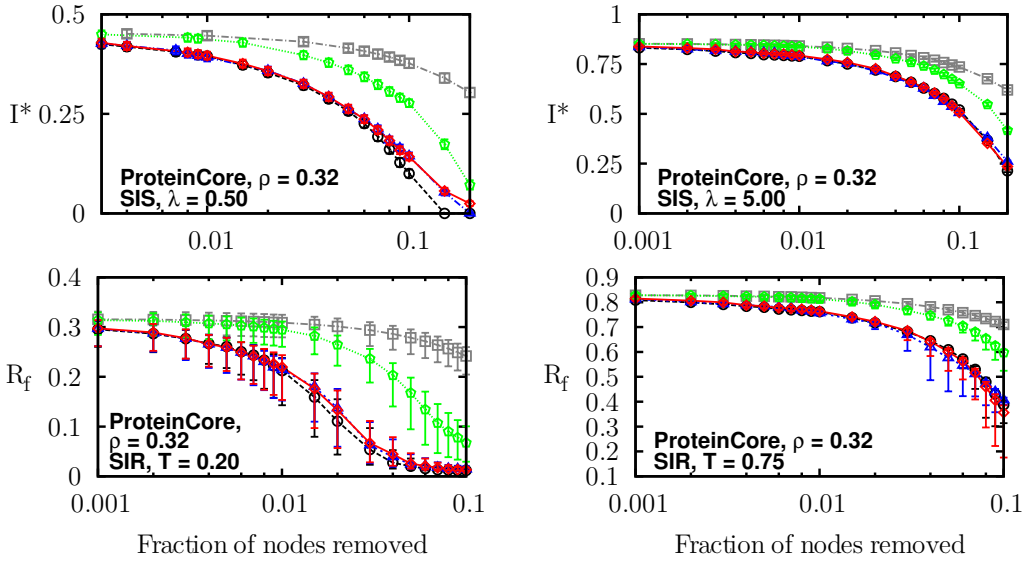
*Supplementary Table 29:* ProteinCore correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.8538 | 0.9118 | 0.7702 | 0.8828 | 0.9543 | 0.7712 |

*Supplementary Figure 29:* ProteinCore degree distribution




*Supplementary Figure 30:* Intervention against epidemics on the protein interactions network after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[26] http://dip.doe-mbi.ucla.edu/

[27] Palla, G., Derényi, I., Farkas, I. & Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435:814-818

## 18 Slashdot online social network

Network of tagged relationships (friends or foes) in the community of the Slashdot news website in November 2008.[28][29]

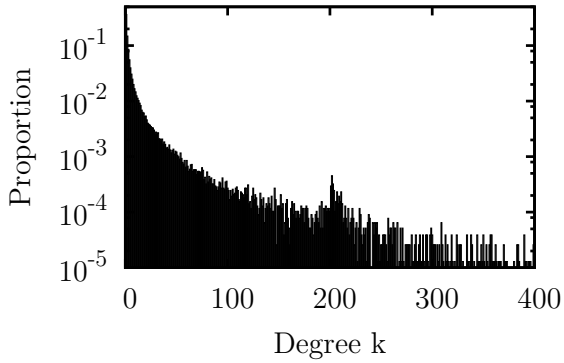
*Supplementary Table 30:* Slashdot statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 77360 | 469180 | 2539 | 54 | $1.2e+08$ | 2506 | 0.46 |

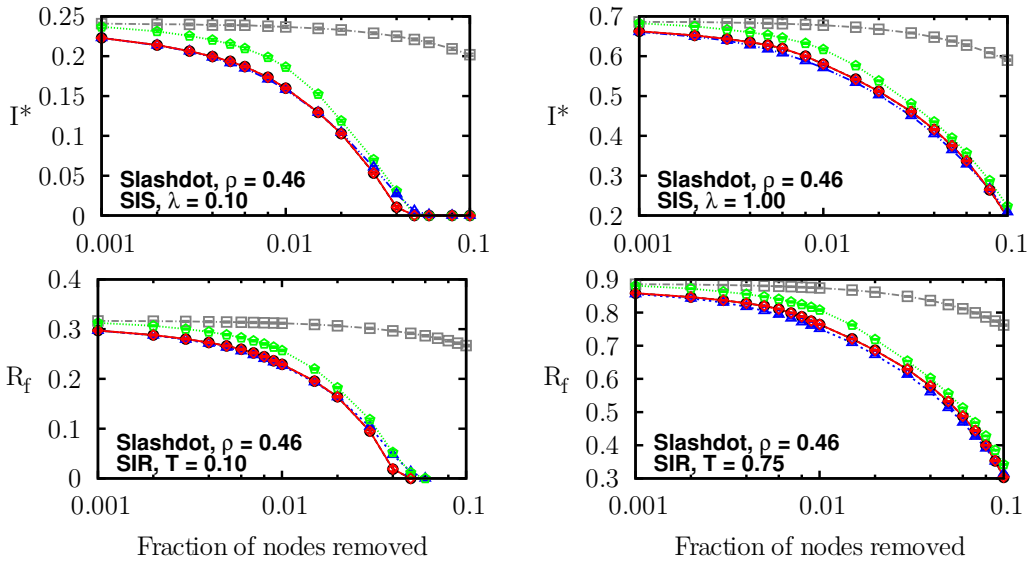
*Supplementary Table 31:* Slashdot correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9958 | 0.9373 | 0.9832 | 0.9358 | 0.9870 | 0.8855 |

*Supplementary Figure 31:* Slashdot degree distribution




*Supplementary Figure 32:* Intervention against epidemics on Slashdot after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[28] http://slashdot.org/

[29] Leskovec, J., Lang, K., Dasgupta, A. & Mahoney, M. (2009) Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics* 6:29-123

## 19 Word association network

Word association graph (built by survey) obtained from the South Florida Free Association norms.[30][31]

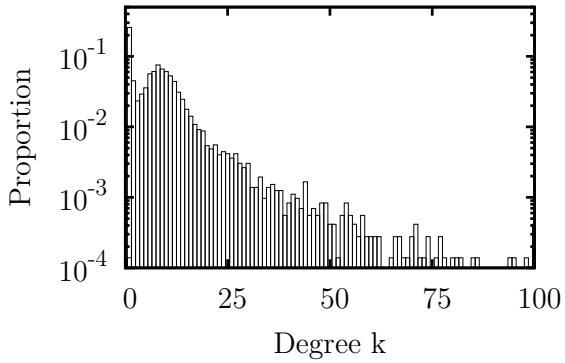
*Supplementary Table 32:* Word ass. statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 7207 | 31784 | 218 | 7 | $1.2e+06$ | 137 | 0.16 |

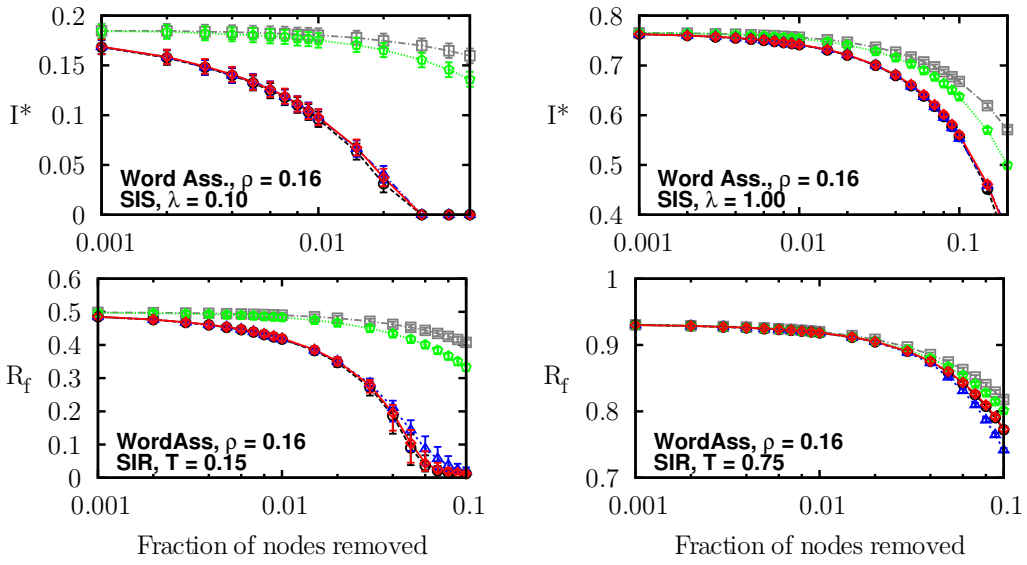
*Supplementary Table 33:* Word ass. correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9698 | 0.9230 | 0.9110 | 0.9229 | 0.9281 | 0.8337 |

*Supplementary Figure 33:* Word ass. degree distribution




*Supplementary Figure 34:* Intervention against epidemics on the word association graph after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[30] http://www.usf.edu/FreeAssociation/

[31] Palla, G., Derényi, I., Farkas, I. & Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435:814-818

## 20 World Wide Web

Network of links between the webpages within *nd.edu* domain and considered undirected for this study.[32]

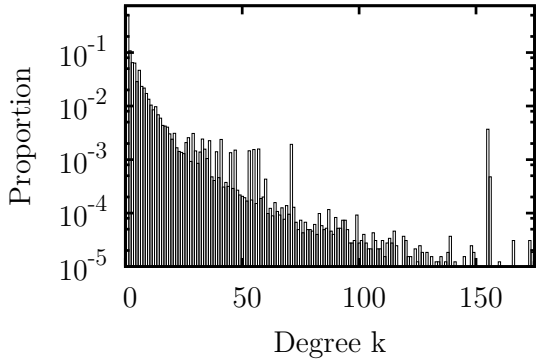
*Supplementary Table 34:* WWW statistics

| $N$ | $L$ | $k_{\max}$ | $c_{\max}$ | $b_{\max}$ | $m_{\max}$ | $\rho$ |
|---|---|---|---|---|---|---|
| 325729 | 1090108 | 10721 | 155 | $2.5e+10$ | 6993 | 0.86 |

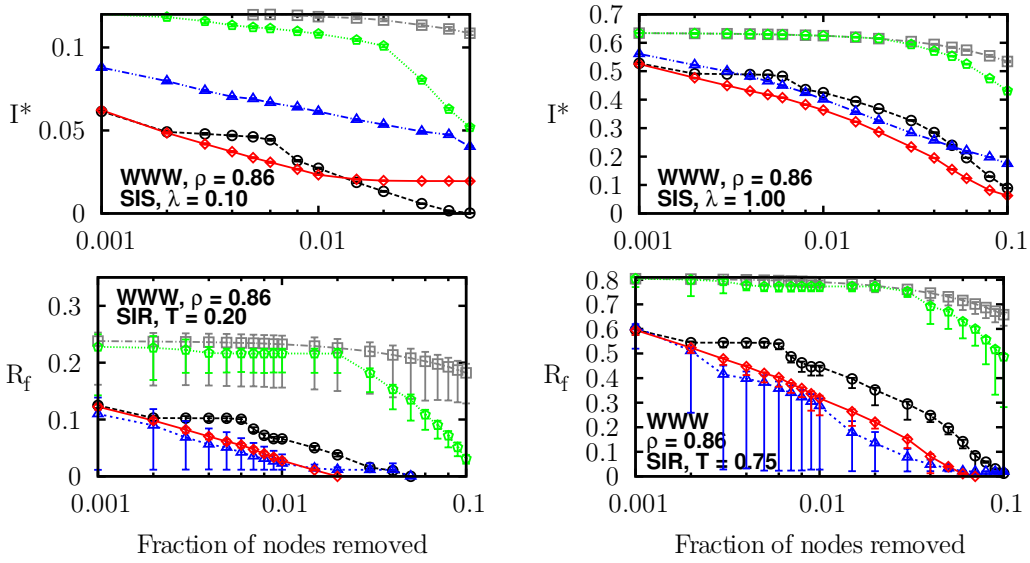
*Supplementary Table 35:* WWW correlations

| $\sigma(k,m)$ | $\sigma(b,m)$ | $\sigma(c,m)$ | $\sigma(k,b)$ | $\sigma(k,c)$ | $\sigma(b,c)$ |
|---|---|---|---|---|---|
| 0.9569 | 0.8683 | 0.9020 | 0.8665 | 0.9614 | 0.7905 |

*Supplementary Figure 35:* WWW degree distribution




*Supplementary Figure 36:* Intervention against epidemics on the WWW after different immunization: randomly (grey squares) and based on coreness (green pentagons), degree (black circles), betweenness centrality (blue triangles) or memberships (red diamonds).




[32]Barabási, A.-L. & Albert, R. (1999) Emergence of scaling in random networks. *Science* 286:509-512