
PRIVACY-SENSITIVE INTERVENTIONS FOR LEVERAGING UNKNOWN NETWORK STRUCTURES TO REGULATE CONTAGION

Vineet Kumar

Yale School of Management

Yale University

vineet.kumar@yale.edu

David Krackhardt

Heinz School of Public Policy and Management

Carnegie Mellon University

krack@cmu.edu

Scott Feld

Department of Sociology

College of Liberal Arts

Purdue University

sfeld@purdue.edu

Spring 2023

Significance Statement:

Networks are powerful tools for interventions due to the cascading impact one individual can have on others. To obtain highly connected individuals for any intervention, we typically need to know the relevant network structure, which is typically not available. We propose intervention strategies termed local and global strategies, based on the friendship paradox to choose individuals for intervention. Both strategies require no prior knowledge of network structure, only sampling from a random individual's friends. These strategies both are guaranteed to obtain greater than average degree for almost any network structure. We characterize the value of these intervention strategies both theoretically and using data from real-world networks. Finally, we identify which strategy performs best with a novel network property called Inversity.

Keywords Contagion | Network Intervention | Friendship Paradox | Inversity

ABSTRACT

Network intervention problems often benefit from selecting a highly-connected node to perform interventions using these nodes, e.g. immunization. However, in many network contexts, the structure of network connections is unknown, leading to a challenge. We develop and examine the mathematical properties of two distinct informationally light strategies, a novel global strategy and local strategy, that yield higher degree nodes in virtually any network structure. We further identify a novel network property called Inversity, whose sign determines which of the two strategies, local or global, will be most effective for a network. We demonstrate that local and global strategies obtain a several-fold improvement in node degree relative to a random selection benchmark for generated and real networks (including contact, affiliation and online networks). In some networks, they achieve a 100-fold improvement. We show how these new strategies can be used to control contagion of an epidemic spreading across a set of village networks, finding that the strategies developed here require far fewer ($< 50\%$) nodes to be immunized, relative to the random strategy baseline. Prior research has typically used the complete network structure to choose nodes for optimal seeding. The relevant network is often costly to collect, and is privacy-invasive, requiring knowing each person's network neighbors, and might not be possible to obtain for time-sensitive interventions. Our interventions are less invasive of individual privacy, since each selected node only needs to nominate some network neighbors for intervention, while mathematically guaranteed to provide better connected nodes.

Network-based interventions are wide-ranging and of crucial importance in any setting where an individual's choice or action has an indirect impact on others, and can have cascading effects. Consider the following: (a) (Reducing) A new infectious disease is spreading through a large population. We want to minimize the number of infected individuals by inoculating using a new vaccine; however, we only have a limited number of doses to administer. (b) (Accelerating) We have a new highly effective medical device with limited samples that we would like to provide to select medical professionals, who can then share information through word of mouth. (c) (Observing) We would like to identify viral contagion as quickly as possible by choosing individuals as observation stations (or for contact tracing).

Although seemingly distinct, whether to reduce, accelerate or observe dynamic contagion, these problems represent a class of network interventions [44] in which we benefit from identifying more central or highly connected individuals in the network.¹

Most current approaches focus on taking advantage of detailed network data on social connections and also on activity to identify influential individuals [25, 24, 48, 2]. However, privacy concerns are increasingly important in such settings, making it challenging to obtain network data [1, 12, 34, 49, 43]. Users are also concerned that their data may be used in algorithms [28], and might even result in discrimination against them [46]. Even, if network data is available, the challenge in many cases is that we do not have access to the *relevant* network structure. In application (a), having the

¹We focus on the class of "simple contagion" problems, which require only one rather than "complex contagion" that require multiple exposures [11, 10].

Facebook (or similar) network structure might not be useful, since the relevant network would be the *physical contact* network, which might be harder to get. In contrast, for application (b), finding a high degree node using a physical contact network of everyone who interacts with a medical professional is unlikely to be informative in characterizing opinion leadership in the profession. For (c), carrying out contact tracing for all individuals can be expensive in effort and time. These factors make it important to be able to leverage the structure of the relevant network while being sensitive to privacy concerns.

We develop, model and characterize privacy-sensitive *novel network intervention strategies*, which obtain highly connected nodes from the *relevant* network by querying randomly chosen individuals, and which do not require knowledge of the underlying network structure. These interventions termed global and local strategies are based on the friendship paradox, and result from theoretical network properties developed here. We also derive a structural network property called *inversivity*, which determines the relative effectiveness of the local and global strategies.

These strategies have several advantages for implementation. First, despite being informationally-light, the strategies here provide provable advantages for virtually any network structures, in contrast to strategies that don't provide such guarantees for general networks. The network structure may be expensive to collect, not be possible to obtain in a timely manner, or may vary over time, making such a method valuable. Second, the strategies are much more privacy-sensitive than mapping out social networks. Third, the strategies can be implemented quickly since they only require local network information obtained by querying individuals or interaction data. Finally, the class of interventions here can be used for both advance and consequent interventions, i.e. for both prevention and treatment interventions.

Friendship Paradox

The Friendship Paradox, which our interventions are based on, is colloquially stated as “your friends have more friends than you” [18, 50].² The intuition for why the friendship paradox helps obtain well-connected nodes is this: since highly-connected nodes (hubs) are connected to many other nodes (by definition), obtaining a random friend (or neighbor) of a random node is likely to result in a hub with greater likelihood, compared to the case of randomly selecting nodes.

We establish that the friendship paradox is actually not just one statement, but a set of distinct claims (All theorems and proofs are in Supplement §S.B). First, we find an impossibility, i.e. the individual-level friendship paradox cannot hold for all individuals in a network (Theorem S1). In practice for real networks, it can hold for a large proportion of nodes in the network (Figs. S2 and S3 in Supplement §S.D). Second, we demonstrate that in contrast to the impossibility of the individual friendship paradox, the network level friendship paradox holds. We find that the *average* number of friends of friends across the network can be characterized in two different ways, using the local and global mean, defined below. Both local and global means are greater than the mean degree of the network, and these means are related through a novel network characteristic we term **inversivity**. While some of these strategies have been suggested

²The phenomenon has also been generalized to the idea that individual attributes and degree are correlated [23], e.g. an individual's co-authors are more likely to be cited [16], or that friends more active on social media [21].

and used as interventions in practice, e.g. [14, 3], their theoretical properties and connections have not been examined and characterized for wide classes of networks.

Local and Global Means We formally characterize the two distinct but related network properties deriving from the friendship paradox relating to the “average number of friends of friends.” Denote a network (see Table S1 for full notation) as an undirected graph $\mathcal{G} = (V, E)$ with V the set of nodes and E the set of edges ($e_{ij} \in \{0, 1\}$ denoting absence or presence of a connection between i and j), D_i refers to the degree of node i , and $N(i)$ the set of i ’s neighbors. We specify the local mean as:

$$\mu_L = \frac{1}{N} \sum_{i \in V} \left[\frac{1}{D_i} \sum_{j \in N(i)} D_j \right] \quad (1)$$

The global mean is defined as the ratio of the total number of friends of friends to the total number of friends in the network, consistent with [18]:

$$\mu_G = \frac{\sum_{i \in V} \left[\sum_{j \in N(i)} D_j \right]}{\sum_{i \in V} D_i} \quad (2)$$

~~The above means arise from differently weighting the average degree across friends. Both means above are consistent with the notion of “average number of friends of friends,” although they are distinct network properties (see Fig. S1 for an example and detailed explanation). The global mean was theoretically investigated earlier and found to be greater than the average degree and is independent of the local structure of connections, given node degrees (Theorem S2). The local mean is also shown to be greater than the mean degree (Theorem S3).³ However, the contrast is that the local mean has distinct properties that depend on local network structure (i.e. who is connected to whom).~~

~~The global mean is greater when there is higher variation across nodes in terms of degree (variance), whereas local mean is greater when we have higher variation across edges, i.e. when edges connect nodes of very dissimilar degree (e.g. with a hub and spoke network). More specifically, the global mean is invariant to rewiring the network while keeping the degree distribution the same, whereas the local mean is impacted by the rewiring (Theorem S6).~~

~~We identify below network structures that result in a greater divergence between the Local and Global mean, and between these means and the average degree (Figure 1, including whether one of the means is always greater than the other, whether they always exhibit correlated variation away from the mean degree, and demonstrate how this matters in network interventions).~~

~~We examine four illustrative network structures (Fig. 1) to answer these questions and to understand the differences between the two means. We find that both local and global mean can be much greater than the mean degree, and between these two means, either of them can be greater than the other, and in some network structures, both can be relatively high compared to the mean degree.~~

³We term these means local or global since the former depends on the local structure (who is connected to whom), whereas the latter only depends on the global network properties (degree distribution).

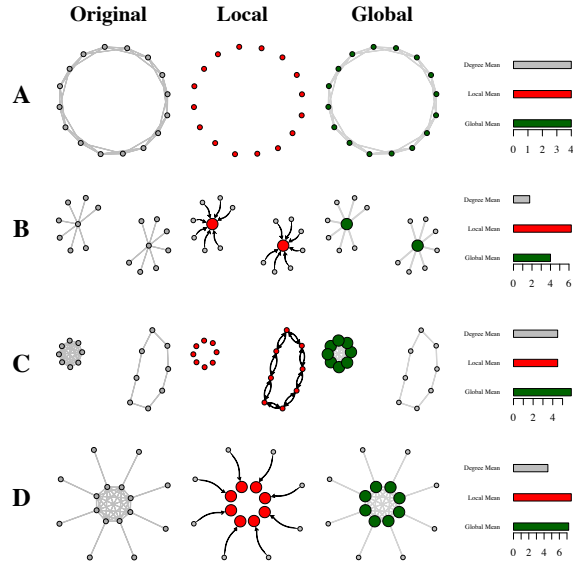


Figure 1: Four Illustrative Networks with Varying Local and Global Means.

Each network in (A)-(D) has the original network plot (left), local weighted network (middle) and global weighted network (right). On the right is a barplot indicating the mean degree, local mean and global mean for each of the networks. *Local Panel (Red)*: In the local weighted network plot (middle), nodes are sized proportional to their weight (w_i^L) in contributing to the local mean. Edges that receive a *higher than median* weight in computing the local mean are in black color. Otherwise, the edges are not plotted in the middle panel. Note that although the original networks are undirected, the selected edges are *directed*. *Global Panel (Green)*: Nodes are sized proportional to their weight (w_i^G) in contributing to the global mean. Edges are all weighted equally in the global weighted network. (A) *Small World Ring*: Each node has four friends, and local and global mean are both equal to average degree (4). None of the edges are shown in the middle panel since all edges have identical weight in computing the local mean. All nodes in both local and global means have the same weight, and size in the middle and right panel. (B) *Two Central Hubs with Spokes*: Each central hub is connected to 7 nodes. The mean degree is lowest in this network. However, local mean is substantially higher than the global mean, and is higher than the mean degree across all networks (a)-(d). In local panel, we see that the weight of central hubs has increased, whereas the corresponding weight for the low degree “spoke” nodes has decreased. In the global panel, the node weights are proportional to degree. (C) *Heavy Core with Attached Cycle*: The global mean is substantially higher than the local mean (and mean degree). Here, we see in the local panel that the weight of each of the nodes has not changed, and all nodes have the same weight. However, in the global panel, we see that the high degree nodes in the complete graph has higher weight compared to the original network, whereas the weights for the nodes in the 2-cycle are lower than in the original network. (D) *Heavy Core with Pendants*: Both the local and global mean are substantially higher than mean degree. In the local panel, the edges connecting core nodes to other nodes (both core and pendant) have a relatively low weight, and are not displayed.

Intervention Strategies

The above formulation of local and global mean suggests distinct intervention strategies. We illustrate *random*, *local* and *global* strategies to choose a “seed” node in the network beginning with an initial randomly chosen node (Table 1). The local strategy would query randomly selected individual nodes with the query, “could you suggest the name of a randomly chosen friend?” as example. The global strategy would ask individual nodes to provide a proportion of their friends, and such a proportion can be set to be small (say 1%) to be sensitive to privacy concerns. Recall that full network strategies, in contrast, require access the entire structure of connections.

We illustrate how the method is able to obtain the *relevant network structure* in a straightforward manner. Specifically, when querying individual nodes to select from the set of friends, we can add a condition to the query. For instance, in application (a) where the focus was on physical contagion network, the query would be "provide a random friend whom you have interacted with in person," where a number of such conditions can be added to the query. Similar conditions

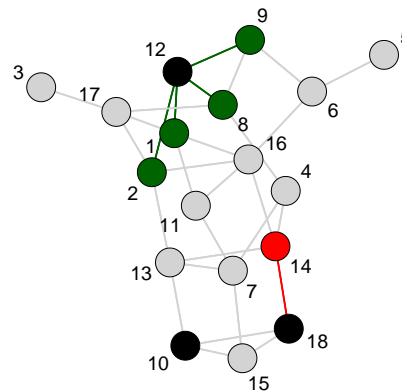
can be used for applications (b) and (c). We can thus view such a query as providing a specific type of network that is more relevant to the application under consideration.

Observe that with the local strategy, the number of seed nodes is fixed, whereas it is probabilistic under the global strategy. For the local strategy, by construction, the expected degree of the obtained seed is equal to the local mean. For the global strategy, we prove that the expected degree of chosen nodes is equal to the global mean (Theorem S5).

Though the local and global strategies appear to be similar in the sense that we are choosing friends of randomly chosen individuals, the crucial distinction lies in whether we are choosing *one* random friend or whether we are choosing among *each* friend. Table 1 details the algorithms to obtain k seeds in a network of size $N \gg k$. This impacts their relative effectiveness as examined below.

Table 1: Implementation of Seeding Strategies

Step	Details
0	Fix $p \in (0, 1]$ (only used for Global strategy in Step 2G).
Repeat Steps 1-2 below until at least k seeds are present in the seed set \mathcal{S} .	
1	Draw a random node r uniformly from set of nodes, V . In Example Network, Nodes 10, 18 and 12 (in black) are drawn for (R), (L) and (G) strategies respectively.
2	Depending on the strategy Random (R), Local (L) or Global (G), do the following:
2R (Random):	Add r to the seed set \mathcal{S} . In Example Network, add node 10 to the seed set.
2L (Local):	Obtain a node s chosen with uniform probability from r 's friends, i.e. $s \in \mathcal{N}_r$. Add the friend s to the seed set \mathcal{S} . In Example Network, one of node 18's friends, node 14 (in red), is chosen at random. Add node 14 to the seed set.
2G (Global):	For each of r 's friends, $s \in \mathcal{N}_r$: With probability p ($0 < p \leq 1$), add s to the seed set \mathcal{S} . In Example Network, each of node 12's friends, nodes 1, 2, 8 and 9 (in green), are added probabilistically (with probability p) to the seed set. <i>Implementation:</i> For each $s \in \mathcal{N}_r$, draw from an independent uniformly distributed random variable $z_s \sim U[0, 1]$. If $z_s < p$, add s to the seed set \mathcal{S} .



Example Network

Note: With Random and Local strategies, we will obtain exactly k nodes in the seed set \mathcal{S} . With the global strategy we might obtain more than k nodes in the seed set. In such a case, we select k nodes at random from the seed set \mathcal{S} without replacement.

Inversity: Connecting Local and Global Means

We next characterize the relative effectiveness of local and global strategies. We identify and define a novel network property, *Inversity*, that determines when the local mean is greater than the global mean. This property captures all local network information related to the local mean and is scale-invariant, i.e. independent of the size or density of the network. We prove that the sign of *inversity* helps us determine which of the local mean or the global mean is higher for any given network. We show how *inversity* is related to but distinct from degree assortativity (in Supplement §S.F).

Inversity is a correlation-based metric that relates the global and local means for any network is obtained as follows. First, define the following edge-based distributions to examine the relationship between the means. The *origin* degree (**O**), $D^O(e)$, *destination* degree (**D**), $D^D(e)$, and *inverse destination* degree (**ID**) distribution, $D^{ID}(e)$, are defined across directed edges $e \in \hat{E}$ as: $D^O(e_{jk}) = D_j$, $D^D(e_{jk}) = D_k$, $D^{ID}(e_{jk}) = \frac{1}{D_k}$. We define the *inversity* across the edge distribution as the Pearson correlation across the origin and inverse degree distributions.

$$\rho = \text{Corr}(D^{\mathbf{O}}, D^{\mathbf{ID}}) \quad (3)$$

We connect (see Theorem S4) the local and global means with *inversity* and the degree distribution ($\kappa_m = \sum_{i \in V} D_i^m$) as:

$$\mu_L = \mu_G + \rho \Psi(\kappa_{-1}, \kappa_1, \kappa_2, \kappa_3) \quad (4)$$

where Ψ is a positive function of the degree distribution.

When connections (edges) are mostly between nodes of similar degree, then *inversity* ρ is likely to be more negative. In such a case, the global mean is greater than the local mean. In contrast, when connections are more likely to be between nodes of dissimilar degree, then *inversity* ρ is positive, and the local strategy is likely to be obtain higher degree nodes. Therefore, if *inversity* is known, we don't need the entire degree distribution to obtain the local mean. Rather, *four* moments of the degree distribution are sufficient for that purpose. *Inversity* captures the local information on imbalances in degree of nodes across edges, whereas the moments of the degree distribution represent global information about the network. *Inversity* ρ has a critical role in determining whether the local or global mean is larger for a network; specifically, $\rho < 0$ indicates the global mean is higher than the local mean, whereas $\rho > 0$ indicates the reverse, implying that knowing *inversity* can help us determine which strategy to use. Even computing *inversity* is information-light, requiring only the $2k$ distribution, which represents the degrees of nodes at the termini of each edge, rather than the entire network [32].

Effectiveness of Strategies: Leverage

To evaluate how much of an improvement over the random strategy is possible, and how this varies across a variety of generated and real networks, we examine the relative effectiveness of strategies, with the random strategy as the

baseline and characterize leverage as the improvement in expected degree. Leverage for strategy s on network \mathcal{G} is defined as $\lambda_s(\mathcal{G}) = \frac{\mu_s(\mathcal{G})}{\mu_D(\mathcal{G})}$ for $s \in \{R, L, G\}$ (since the random strategy obtains the mean degree in expectation, the leverage for R is $\lambda_r(\mathcal{G}) = 1$ and it serves as a baseline). We examine the leverage of both generated and real networks.

Generated Networks: The generated networks using Random, Scale Free and Small World models are detailed in the Supplement §S.E. We examine 3 generative mechanisms: (a) Erdos-Renyi (ER), (b) Scale Free (SF) and (c) Small World (SW) networks (Fig. S4) [17, 5, 47]. We find that for ER networks, at very low density (edge probability), the leverage is very low because most edges connect nodes that have a degree of 1. As density increases, we obtain more variation in degrees, and local leverage increases. However, beyond an edge probability of $p = 0.05$, leverage decreases as the density of the network decreases. Local leverage thus forms a non-monotonic pattern with ER networks. For SF networks, rather than density or edge probability, we initially examine leverage as the network becomes more centralized (as γ increases above 1, very high degree nodes have a lower probability of occurring). We find that as γ increases from 1 to 2, the leverage increases, but then decreases beyond 2. For WS networks, unlike in the ER and SF networks, leverage is monotonically decreasing with number of neighbors (or density), and is monotonically increasing with rewiring probability.

Real Networks: The range of real networks is detailed in Supplement §S.C. First, observing the local strategy (Fig. 2A), we find that for all networks, as expected, the friendship paradox strategies are at least as good as the random strategy. Second, for networks like Twitter (OS4) or Internet Topology (C1), the leverage can be as high as 100. Thus, obtaining a friend of a random node will provide a 100-fold increase in the expected degree of a chosen node. Third, we observe that both local and global leverage (Figs. 2A and 2B) are higher for nodes when average degree is intermediate, i.e. not too low or high. Some networks like the CA Roads network (I3) have very little degree variation and local and global strategies are relatively less effective. Finally, we examine when local and global strategies make a relative difference (Fig. 2B). We find that the highest ratio of local to global mean is for Twitter network (OS4), whereas the lowest ratio (indicating that global strategy has a higher expected mean degree) is shown by Flickr (OS2), both of which belong to the same category of online social networks. Citation networks tend to have higher global mean, whereas for Infrastructure networks, both strategies seem to work just as well.

Application: Controlling Contagion in Networks

We demonstrate an application comparing different strategies to control simple contagion spreading through a network, using the local and global strategies. There are a number of models of contagion, and they can be parametrized several ways. However, remarkably most models of contagion can be characterized by a single parameter termed the *epidemic threshold*. If the ratio of infection to that of recovery is lower than the epidemic threshold, then the epidemic is contained and will die out, whereas if the ratio is above the threshold, then it could turn into an epidemic. The epidemic threshold captures the idea that a contagion when introduced into the network will die out if the reproductive number (\mathcal{R}_0) is

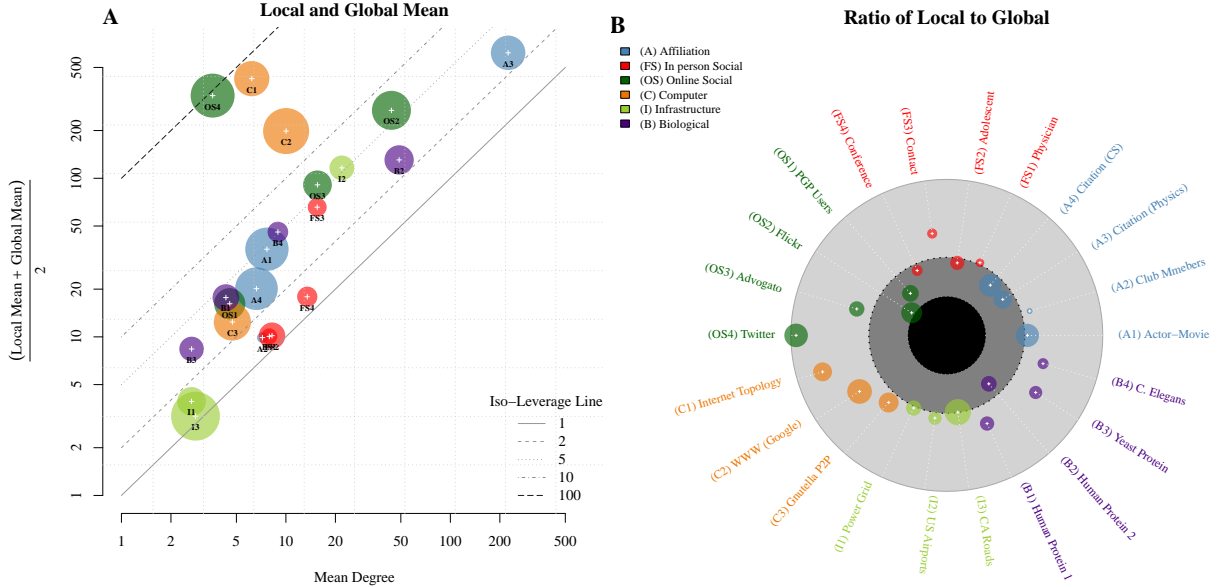


Figure 2: Global and Local Leverage in Real Networks Local and Global Means across Networks (each circle is a network). Area of circles indicates size of networks (number of nodes) in log scale. Color of circle indicates network category. (A) The average of **Local and Global Mean** is higher than mean degree in all real networks, with the highest differences occurring in online social networks and computer networks. Most large networks also tend to show a higher leverage ratio. For in person or face to face networks, the pattern is more variable. The iso-leverage line indicates leverage levels of 1,2,5,10 and 100. We find that all networks have leverage greater than 1, a majority of networks have leverage greater than 5, and 2 networks have leverage close to 100. (B) **Comparison:** Ratio of Local to Global Mean. The ratio of local to global mean $\frac{\mu_L}{\mu_G}$ is represented as follows ($< \frac{1}{2}$ in black circle, $\frac{1}{2} < \frac{\mu_L}{\mu_G} < 1$ in dark gray circle and $1 < \frac{\mu_L}{\mu_G} < 2$ in light gray circle). For example, in the Twitter network, local mean is almost twice the global mean, whereas in the Flickr network, global mean is almost twice the local mean. Computer networks have higher values of the ratio, whereas Infrastructure networks have similar values of local and global means.

below the epidemic threshold, and lead to an epidemic if above the threshold. Thus, a network with a higher epidemic threshold would be able to better withstand or control an infection.

We then examine how the epidemic threshold changes as a function of the proportion of nodes vaccinated (removed), for each strategy (random, local and global). For most *virus propagation models* (VPM) is characterized as the inverse of the greatest (first) eigenvalue of the adjacency matrix A of the network, denoted as $\tau(E) = \frac{1}{\lambda_1(E)}$ (details in §S.G of the Supplement). The threshold in an undirected network is shown to be proportional to the *inverse of the largest eigenvalue* for a wide range of VPMs, including SIR, SEIR, etc. models that have been commonly used for modeling infectious diseases [33]. Nodes are selected for immunization or treatment using each of the intervention strategies (random, local and global).

To ensure realistic in person contact networks for modeling contagion, we examine data in the India villages networks from [4], who collected detailed full census data on the social networks of 75 villages in southern India. The social networks are captured at two different levels of aggregation, at the level of individuals and of households. The advantage of this data set is that villages are relatively geographically isolated and can therefore be treated as separate networks. Details of the network dataset are provided in §S.C of the Supplement.

We find that networks can have either positive or negative inversivity depending on how nodes and edges are defined. Figure 3(a) illustrates the inversivity values across the 75 villages separately for individual and household networks. When nodes as defined as individuals, we find that the networks have negative inversivity, whereas if the nodes are defined as households, the inversivity values of the resulting networks are mostly positive. Inversivity values therefore depend strongly on how the network structure is aggregated. Considering interventions, the inversivity values suggest that a household-based intervention might use the local strategy, whereas the individual-based intervention might use the global strategy. Note that the household-level ties are aggregated from the individual-level ties, implying that networks obtained from similar underlying relationships can result in dramatically different inversivity characteristics, which lead to different interventions.

Our first step is to identify how the epidemic threshold τ changes as we immunize nodes from the network \mathcal{G} . This dataset is especially useful in our analysis since the villages are relatively isolated, implying they can be evaluated separately. While immunizing (or removing) any node from the network is likely to increase the epidemic threshold, immunizing well-connected nodes is likely to prove especially beneficial. We examine the effectiveness of the three strategies (Random, Local and Global) in identifying which nodes to immunize in the network. We next consider Epidemic Thresholds with Immunization in Figure 3(b). We examine 3 strategies (Random, Local and Global) to choose nodes to immunize. The proportion of nodes immunized ranges from 1% - 75%. In both household and individual networks, we find that the friendship paradox strategies obtain higher thresholds than random, for the same proportion of nodes immunized. For instance, in the household networks, to achieve a threshold $\tau = 0.15$, the random strategy needs to have about 50% of nodes immunized, but the Local and Global strategies require less than half of that, at around 25%. For the household networks, we find that the local strategy is better than the global strategies especially at higher levels of removal. However, for individual networks, we find that the Global strategy obtains greater thresholds than Local. This broadly signifies that it is helpful to know which among the global or local strategies to use, and is consistent with the sign of inversivity.

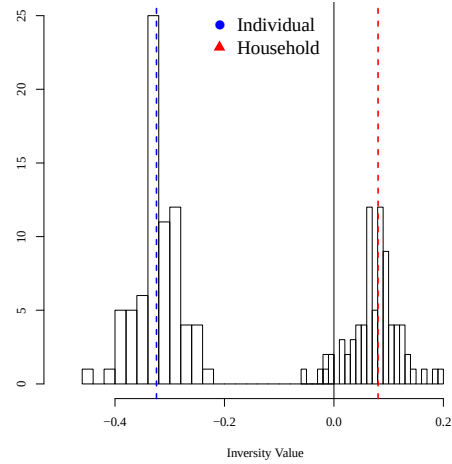
Finally, we simulate an infection process and evaluate the epidemic characteristic of peak infectivity using an SIR virus propagation model (details in §S.G), the most popular class of contagion models [42], with parameters of the simulation detailed in Table S5. We examine peak infection since it is known to be an important characteristic of epidemics [40], directly impacting the load on the healthcare system. Define $I_{it} \in \{0, 1\}$ as an indicator of whether an individual i is infected at time t . We evaluate the epidemics on the proportion of the population infected at the peak of the epidemic $(\frac{1}{N} \max_t (\sum_i I_{it}))$, which is a useful measure in cases where hospital capacity is constrained. There has been much discussion about interventions to avoid precisely such a peak [7]. A strategy with a density plot to the left of another is better in terms of reducing the severity of the epidemic. Thus, for household networks, the local strategy (in red) is better than the global, which in turn is better than the random strategy in reducing peak infection. For the individual networks, however, the global strategy is better.

Conclusion

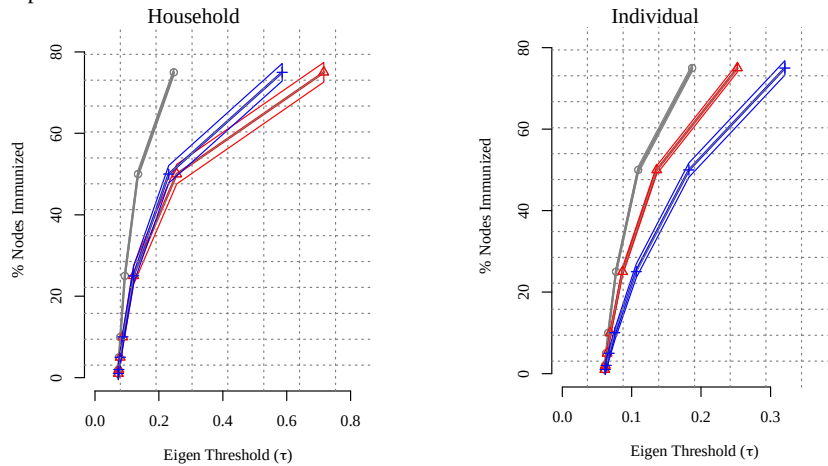
We show that with unknown networks, the friendship paradox can be leveraged to obtain such individuals with minimal informational requirements. We identify intervention strategies (local and global) that have theoretical guarantees on obtaining better-connected individuals. With both generated random networks and real networks, our results show the value of using the local and global strategies to obtain highly connected nodes. In the vast majority of networks, we obtain at least double the average degree, and some networks show increases of several hundred-folds in node degree. We expect the advantages of speed of implementation, generality of application areas for these privacy-sensitive and informationally-light strategies to provide an important tool for network interventions in unknown structures.

Figure 3: Inversity and Epidemic Characteristics on Village Networks. Data for these networks ($N=75$) obtained from Indian villages is publicly available and detailed in [4], and includes both individual-level (Individual network) ties as well as connections between households (Household network). (a) Inversity: Frequency plot of inversity across village networks. (b) Epidemic Thresholds with Immunization: The epidemic threshold ($\tau = \frac{1}{\lambda_1}$) is computed at different levels of immunization, and for different immunization strategies (Random, Local and Global). The strategies are used to select nodes, which upon immunization lose the capability to transmit contagion, and infect other nodes. (c) Epidemic Peak Infection: The proportion of the network nodes that are infected during the peak of the epidemic for immunization level (X) is represented as a density plot. Variation is obtained due to differences in outcomes across village as well as simulation variation. See Table S5 in the Supplement for simulation parameters.

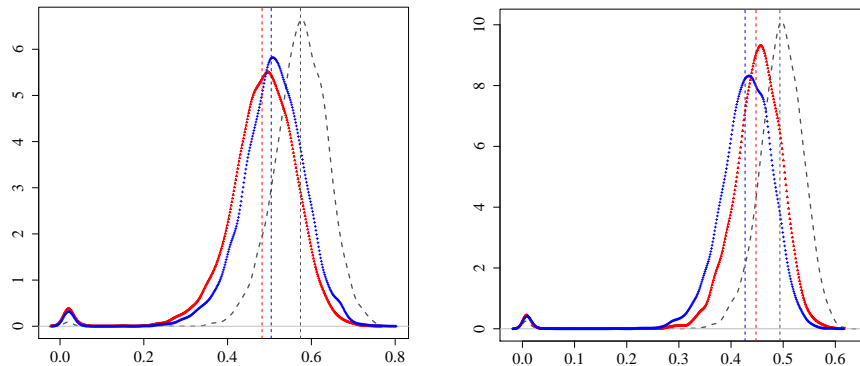
(A) Inversity for Individual and Household Networks



(B): Epidemic Thresholds



(C) Peak Infection Proportion (Density Plot)



References

- [1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*, 30(4):736–758, 2020.
- [2] Mohammed Ali Al-Garadi, Kasturi Dewi Varathan, Sri Devi Ravana, Ejaz Ahmed, Ghulam Mujtaba, Muhammad Usman Shahid Khan, and Samee U Khan. Analysis of online social network connections for identification of influential users: Survey and open research issues. *ACM Computing Surveys (CSUR)*, 51(1):1–37, 2018.
- [3] Marcus Alexander, Laura Forastiere, Swati Gupta, and Nicholas A Christakis. Algorithms for seeding social networks can enhance the adoption of a public health intervention in urban india. *Proceedings of the National Academy of Sciences*, 119(30):e2120742119, 2022.
- [4] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.
- [5] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [6] T Berge, JM-S Lubuma, GM Moremedi, N Morris, and R Kondera-Shava. A simple mathematical model for ebola in africa. *Journal of biological dynamics*, 11(1):42–74, 2017.
- [7] Ross D Booton, Louis MacGregor, Lucy Vass, Katharine J Looker, Catherine Hyams, Philip D Bright, Irasha Harding, Rajeka Lazarus, Fergus Hamilton, Daniel Lawson, et al. Estimating the covid-19 epidemic trajectory and hospital capacity requirements in south west england: a mathematical modelling framework. *BMJ open*, 11(1):e041536, 2021.
- [8] Fred Brauer. The Kermack–McKendrick epidemic model revisited. *Mathematical biosciences*, 198(2):119–131, 2005.
- [9] Diego Caccavo. Chinese and italian covid-19 outbreaks can be correctly described by a modified sird model. *medRxiv*, 2020.
- [10] D. Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.
- [11] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.
- [12] Francesca Cerruto, Stefano Cirillo, Domenico Desiato, Simone Michele Gambardella, and Giuseppe Polese. Social network data analysis to highlight privacy threats in sharing data. *Journal of Big Data*, 9(1):19, 2022.
- [13] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4):1–26, 2008.
- [14] R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901, 2003.

- [15] Severine Deguen, Guy Thomas, and Nguyen Phong Chau. Estimation of the contact rate in a seasonal seir model: application to chickenpox incidence in france. *Statistics in medicine*, 19(9):1207–1216, 2000.
- [16] Young-Ho Eom and Hang-Hyun Jo. Generalized friendship paradox in complex networks: The case of scientific collaboration. *Scientific reports*, 4:srep04603, 2014.
- [17] Paul Erdős and Alfréd Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.
- [18] S.L. Feld. Why your friends have more friends than you do. *American Journal of Sociology*, pages 1464–1477, 1991.
- [19] Matthew J Ferrari, Rebecca F Grais, Nita Bharti, Andrew JK Conlan, Ottar N Bjørnstad, Lara J Wolfson, Philippe J Guerin, Ali Djibo, and Bryan T Grenfell. The dynamics of measles in sub-saharan africa. *Nature*, 451(7179):679–684, 2008.
- [20] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.
- [21] Nathan Oken Hodas, Farshad Kooti, and Kristina Lerman. Friendship paradox redux: Your friends are more interesting than you. *ICWSM*, 13:8–10, 2013.
- [22] Matthew O Jackson. The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, 127(2):777–818, 2019.
- [23] Hang-Hyun Jo and Young-Ho Eom. Generalized friendship paradox in networks with tunable degree-attribute correlation. *Physical Review E*, 90(2):022809, 2014.
- [24] Zsolt Katona, Peter Pal Zubcsek, and Miklos Sarvary. Network effects and personal influences: The diffusion of an online social network. *Journal of marketing research*, 48(3):425–443, 2011.
- [25] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.
- [26] David Krackhardt. Graph theoretical models of structural leverage in marketing. In *Computational Analysis of Social and Organizational Systems Conference, Pittsburgh, PA*, 2002.
- [27] Jérôme Kunegis. Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350. ACM, 2013.
- [28] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)*, 54(2):1–36, 2021.
- [29] AG M’Kendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925.
- [30] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.
- [31] Mark EJ Newman and Juyong Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122, 2003.

- [32] Chiara Orsini, Marija M Dankulov, Pol Colomer-de Simón, Almerima Jamakovic, Priya Mahadevan, Amin Vahdat, Kevin E Bassler, Zoltán Toroczkai, Marián Boguñá, Guido Caldarelli, et al. Quantifying randomness in real networks. *Nature communications*, 6(1):1–10, 2015.
- [33] B Aditya Prakash, Deepayan Chakrabarti, Michalis Faloutsos, Nicholas Valler, and Christos Faloutsos. Got the flu (or mumps)? check the eigenvalue! *arXiv preprint arXiv:1004.0060*, 2010.
- [34] A Praveena and S Smys. Anonymization in social networks: a survey on the issues of data privacy in social network sites. *Journal of International Journal Of Engineering And Computer Science*, 5(3):15912–15918, 2016.
- [35] Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl AB Pearson, James D Munday, et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*, 2020.
- [36] Steven Riley, Christophe Fraser, Christl A Donnelly, Azra C Ghani, Laith J Abu-Raddad, Anthony J Hedley, Gabriel M Leung, Lai-Ming Ho, Tai-Hing Lam, Thuan Q Thach, et al. Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions. *Science*, 300(5627):1961–1966, 2003.
- [37] Irene Sendiña-Nadal, Michael M Danziger, Z Wang, Shlomo Havlin, and Stefano Boccaletti. Assortativity and leadership emerge from anti-preferential attachment in heterogeneous networks. *Scientific reports*, 6:21297, 2016.
- [38] Michael Small and CK Tse. Clustering model for transmission of the sars virus: application to epidemic control and risk assessment. *Physica A: Statistical Mechanics and its Applications*, 351(2-4):499–511, 2005.
- [39] DL Smith, J Dushoff, RW Snow, and SI Hay. The entomological inoculation rate and plasmodium falciparum infection in african children. *Nature*, 438(7067):492–495, 2005.
- [40] Alessandro Soria, Stefania Galimberti, Giuseppe Lapadula, Francesca Visco, Agata Ardini, Maria Grazia Valsecchi, and Paolo Bonfanti. The high volume of patients admitted during the sars-cov-2 pandemic has an independent harmful impact on in-hospital mortality from covid-19. *PLoS one*, 16(1):e0246170, 2021.
- [41] S. Strogatz. Friends you can count on. *The New York Times*, September 17, 2012.
- [42] Juliana Tolles and ThaiBinh Luong. Modeling epidemics with compartmental models. *Jama*, 323(24):2515–2516, 2020.
- [43] Sadeh Torabi and Konstantin Beznosov. Privacy aspects of health related information sharing in online social networks. In *2013 {USENIX} Workshop on Health Information Technologies (HealthTech 13)*, 2013.
- [44] T.W. Valente. Network interventions. *Science*, 337(6090):49–53, 2012.
- [45] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. On the evolution of user interaction in Facebook. In *Proc. Workshop on Online Social Networks*, pages 37–42, 2009.
- [46] Sandra Wachter. Normative challenges of identification in the internet of things: Privacy, profiling, discrimination, and the gdpr. *Computer law & security review*, 34(3):436–449, 2018.

- [47] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998.
- [48] Christo Wilson, Alessandra Sala, Krishna PN Puttaswamy, and Ben Y Zhao. Beyond social graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web (TWEB)*, 6(4):1–31, 2012.
- [49] Wenjing Xie and Kavita Karan. Consumers' privacy concern and privacy protection on social network sites in the era of big data: Empirical evidence from college students. *Journal of Interactive Advertising*, 19(3):187–201, 2019.
- [50] E.W. Zuckerman and J.T. Jost. What makes you think you're so popular? self-evaluation maintenance and the subjective side of the "friendship paradox". *Social Psychology Quarterly*, pages 207–223, 2001.

Supplementary Material for

Privacy-Sensitive Interventions for Leveraging Unknown Network Structures to Regulate Contagion

Vineet Kumar, David Krackhardt and Scott Feld

S.A Local and Global Mean in Example Network

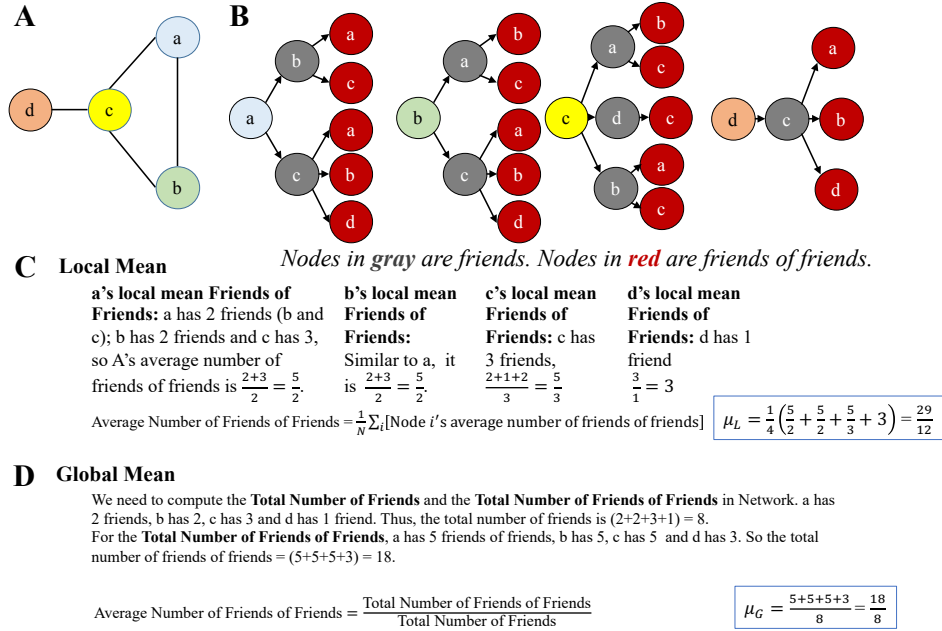


Figure S1: Local and Global Means in Example Network. (A) **Network.** Example network with 4 nodes a, b, c and d . (B) **Illustration of Friends and Friends of Friends.** Each node is mapped out with its friends and friends of friends. The node is in light blue color, Friends are in gray color and Friends of Friends are in red color. Node a has 2 friends b and c . Node a also has 5 friends of friends. (C) **Local Mean.** a has 2 friends, b and c . The total number of friends of friends in a 's network is 5: (a 's friend b has 2 friends, a and c ; a 's friend c has 3 friends, a, b and d). So average number of friends of friends for a is the ratio of the number of nodes in black to the number of nodes in red, i.e. $\frac{2+3}{2} = \frac{5}{2}$. Similarly, for the other nodes, we have $b : \frac{2+3}{2} = \frac{5}{2}$, $c : \frac{2+1+2}{3} = \frac{5}{3}$, $d : \frac{3}{1} = 3$. The local mean of the network is the mean of these local average friends of friends, so $\mu_L = \frac{1}{4} \left(\frac{5}{2} + \frac{5}{2} + \frac{5}{3} + \frac{3}{1} \right) = 2.42$. (D) **Global Mean.** Global mean is the ratio of the total number of friends of friends to the total number of friends. The total number of friends of friends contributed by a is 5. Similarly, b contributes 5, c contributes 5 and d contributes 3 friends of friends. Thus, the total number of friends of friends (i.e. the nodes in red color) are $(5 + 5 + 5 + 3) = 18$. The total number of friends is represented by the nodes in gray, $(2 + 2 + 3 + 1) = 8$.

S.B Mathematical Appendix

Formally, the network graph $\mathcal{G} = (V, E)$ is comprised of a set of N individual nodes and a set of undirected edges E . Each element of E is a pair of nodes, (i, j) indicates an edge (connection) with $e_{ij} \in \{0, 1\}$. We also define the directed edge set \hat{E} including both (i, j) and (j, i) as distinct elements of \hat{E} corresponding to an undirected edge $i \leftrightarrow j$. We detail the table of notation in Table S1.

Table S1: Table of Notation

Symbol	Term	Definition
\mathcal{G}, V, E	Network	Network Graph of Nodes V and Edges E
\hat{E}	Directed Edge Set	Each edge in E is replaced by two directed edges
D_i	Degree	Number of friends of i
$\mathcal{N}(i)$	Neighbors	Set of Friends of i
F_i	Average degree of friends of i	$\frac{1}{D_i} \sum_{j \in \mathcal{N}(i)} D_j$
μ_D, σ_D^2	Mean and variance of Degrees	$\frac{1}{N} \sum_i D_i, \frac{1}{N} \sum_i (D_i - \mu_D)^2$
μ_L	Local Mean	$\frac{1}{N} \sum_i F_i$
μ_G	Global Mean	$\frac{\sum_i D_i F_i}{\sum_i D_i}$
ρ	Inversity	$Corr\left(D_i, \frac{1}{D_j}\right) \forall (i, j) \in \hat{E}$

The basic idea of the friendship paradox can be expressed as "your friends have more friends than you." We examine the degree to which the friendship paradox holds for individual nodes, or the individual friendship paradox. We find in the result below that it cannot hold for all nodes, but can hold for an arbitrarily high proportion (< 1) of nodes.

Theorem S1. *The friendship paradox statement that "your friends have more friends than you" cannot hold for all nodes in a network. Also, the statement can hold for all nodes, except one.*

Proof. Consider a connected network where not all degrees are identical (if all are identical, the statement cannot hold). There must be at least one node that has the highest degree D_{max} and which is connected to at least one node with a lower degree. If not, then the connected network is comprised entirely of highest (identical) degree nodes, thus contradicting the initial statement. If the highest degree node is connected to a lower degree node, then the average friends of friends of the highest degree node must be lower than D_{max} . Thus the statement cannot hold for *all* nodes. To show the second part that it can hold for all nodes except one, consider the star (hub and spoke) network, where all of the nodes except the central node have fewer friends than their friends do. □

Theorem S2. [Feld 1991] *For a network $\mathcal{G} = (V, E)$ with degree mean μ_D and variance σ_D^2 , the global mean of friends of friends is $\mu_G = \left(\mu_D + \frac{\sigma_D^2}{\mu_D}\right)$*

Proof. (as given in Feld, 1991). $\mu_G = \frac{\sum_i \sum_j e_{ij} D_j}{\sum_i D_i} = \frac{\sum_i D_i^2}{\sum_i D_i} = \frac{\mu_D^2 + \sigma_D^2}{\mu_D}$ □

Theorem S3. For any general network $\mathcal{G} = (V, E)$ with mean degree μ_D , the local mean of friends is given by

$$\mu_L = \mu_D + \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{(D_i - D_j)^2}{D_i D_j} \right] \quad (5)$$

where D_i is the degree of node i , and $e_{ij} \in \{0, 1\}$ indicates a connection between i and j .

Proof. Let D_i denote the number of connections of individual i , i.e. $D_i = |\{k \in V : (i, k) \in E\}|$. Denote the set of neighbors of i by $\mathcal{N}(i) = \{k \in V : (i, k) \in E\}$. Define $F_i = \frac{1}{D_i} \sum_{j \in \mathcal{N}(i)} D_j$ as the mean number of friends for friends of i . The local mean is defined as:

$$\mu_L = \frac{1}{|V|} \sum_i F_i = \sum_{i \in V} \left[\frac{1}{D_i} \left(\sum_{j \in \mathcal{N}(i)} D_j \right) \right]$$

Rewriting the expression for μ_L in terms of the connections (edges) between individuals, we obtain:

$$\begin{aligned} \mu_L &= \frac{1}{|V|} \sum_{i \in V} \left[\frac{1}{D_i} \left(\sum_{j \in V} e_{ij} D_j \right) \right] = \frac{1}{|V|} \sum_{i \in V} \sum_{j \in V} \left[e_{ij} \frac{1}{D_i} (D_j) \right] \\ &= \frac{1}{2|V|} \sum_{(i,j) \in V \times V} \left[e_{ij} \left(\frac{D_j}{D_i} \right) + e_{ji} \left(\frac{D_i}{D_j} \right) \right] = \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{D_j}{D_i} + \frac{D_i}{D_j} \right] \\ &= \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{D_j^2 + D_i^2}{D_i D_j} \right] = \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{(D_i - D_j)^2 + 2D_i D_j}{D_i D_j} \right] \\ &= \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{(D_i - D_j)^2}{D_i D_j} \right] + \frac{1}{2|V|} (4|E|) \\ &= \mu_D + \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{(D_i - D_j)^2}{D_i D_j} \right] \quad \square \end{aligned}$$

□

Note that what we characterize as the local mean defined as above was examined by others including [18] etc. and was independently shown to be greater than the mean degree by us ([26]) and others (including by Christian Borgs & Jennifer Chayes in an online comment to an article by [41], and by [22]). However, the properties of the local mean have not been formally examined and characterized.

Theorem S4. Define the m -th moment of the degree distribution by $\kappa_m = \frac{1}{N} \sum_{i \in V} D_i^m$. The local and global means are connected by the following relationship involving the inversity ρ and the -1, 1, 2, and 3rd moments of the degree distribution:

$$\mu_L = \mu_G + \rho \sqrt{\left(\frac{\kappa_1 \kappa_3 - \kappa_2^2}{\kappa_1} \right) \left[\kappa_{-1} - (\kappa_1)^{-1} \right]}$$

Proof. Define the moments of the degree distribution as: $\kappa_m = \frac{1}{N} \sum_i D_i^m$. Since we defined ρ the measure of inversivity as the correlation of two distributions that we specify as the origin degree (**O**) and inverse destination degree (**ID**) distributions. The **O** distribution consists of the degree of nodes corresponding to edges, and **ID** distribution consists of the inverse degree of nodes corresponding to edges. Thus, each connection (edge) contributes *two* entries to *each* distribution. For example, if there is a connection between i and j , i.e. $e_{ij} = 1$, we would have $(D_i, \frac{1}{D_j})$ and $(D_j, \frac{1}{D_i})$. Observe that each individual appears in both distributions multiple times based on degree.

Next, we detail the mean and variance of the distributions. First, we consider the means. The mean of the origin distribution is $\mu_O = \frac{1}{2|E|} \sum_i D_i^2 = \frac{\mu_D^2 + \sigma_D^2}{\mu_D} = \mu_G = \frac{\kappa_2}{\kappa_1}$. Similarly, the **ID** mean is $\mu_{ID} = \frac{1}{2|E|} \sum_i D_i \left(\frac{1}{D_i}\right) = \frac{1}{\mu_D}$. Next, consider the variances. The variance of the origin distribution (**O**) is computed as:

$$\begin{aligned} \sigma_O^2 &= \frac{1}{2|E|} \sum_{(i,j) \in E} (D_i - \mu_O)^2 = \frac{1}{2|E|} \sum_{i \in V} D_i (D_i - \mu_O)^2 \\ &= \frac{1}{N\mu_D} \sum_{i \in V} [D_i^3 - 2\mu_O D_i^2 + (\mu_O)^2 D_i] = \frac{\kappa_3}{\kappa_1} - \left(\frac{\kappa_2}{\kappa_1}\right)^2 \end{aligned}$$

Next, we express the corresponding variance of the inverse destination degree distribution (**ID**), σ_{ID}^2 . Again, recall that $\frac{1}{D_i}$ does not appear just once, but D_i times. Therefore, we have:

$$\begin{aligned} \sigma_{ID}^2 &= \frac{1}{2|E|} \sum_{(i,j) \in E} \left[\left(\frac{1}{D_j} - \frac{1}{\mu_D} \right)^2 \right] = \frac{1}{2|E|} \sum_{(i,j) \in E} \left(\frac{1}{D_j^2} + \frac{1}{\mu_D^2} - \frac{2}{\mu_D D_j} \right) \\ &= \frac{1}{2|E|} \left[\sum_{(i,j) \in E} \frac{1}{D_j^2} + \frac{1}{\mu_D^2} \left(\sum_{(i,j) \in E} 1 \right) - \frac{2}{\mu_D} \sum_{(i,j) \in E} \frac{1}{D_j} \right] = \frac{1}{2|E|} \left[\sum_{j \in V} \frac{1}{D_j} + \frac{1}{\mu_D^2} 2|E| - \frac{2}{\mu_D} N \right] \\ &= \frac{1}{\mu_D N} \left[\sum_{j \in V} \frac{1}{D_j} \right] - \frac{1}{\mu_D^2} = (\kappa_1)^{-1} \left[\kappa_{-1} - (\kappa_1)^{-1} \right] \end{aligned}$$

We next turn to the inversivity and based on the definition we connect it to the local and global means and the degree distribution.

$$\begin{aligned}
\rho &= \left(\frac{1}{2|E|\sigma_O\sigma_{ID}} \right) \sum_{(i,j) \in E} e_{ij} \left[(D_i - \mu_O) \left(\frac{1}{D_j} - \frac{1}{\mu_D} \right) \right] \\
(N\mu_D\sigma_O\sigma_{ID})\rho &= \left[\sum_{(i,j) \in E} e_{ij} \left(\frac{D_i}{D_j} \right) - \mu_O \left(\sum_{(i,j) \in E} \frac{1}{D_j} \right) - \frac{1}{\mu_D} \sum_{(i,j) \in E} D_i + \sum_{(i,j) \in E} e_{ij} \left(\frac{\mu_O}{\mu_D} \right) \right] \\
&= \left[N(\mu_L) - \mu_O \cdot N - \frac{1}{\mu_D} \sum_{(i,j) \in E} D_i + \sum_{(i,j) \in E} e_{ij} \left(\frac{\mu_O}{\mu_D} \right) \right] \\
&= \left[(N\mu_L) - N\mu_O - \frac{1}{\mu_D} \sum_{(i,j) \in E} D_i + 2|E| \left(\frac{\mu_O}{\mu_D} \right) \right] \\
\Rightarrow \mu_L &= \mu_G + \rho \cdot \mu_D \cdot \sigma_O\sigma_{ID}
\end{aligned}$$

Finally, substituting $\mu_D = \kappa_1$ and the expressions for the variances, we obtain:

$$\mu_L = \mu_G + \rho \sqrt{\left(\frac{\kappa_1\kappa_3 - \kappa_2^2}{\kappa_1} \right) \left[\kappa_{-1} - (\kappa_1)^{-1} \right]} \quad (6)$$

□

Theorem S5. *The expected degree of nodes chosen by global strategy is the global mean.*

Proof. To determine the expected degree of a node chosen by the global strategy: Choose $M = 1$ node initially, (say X). With probability q , choose each neighbor of X . For a node k with degree D_k , the probability of being chosen by this process is the first step when any of k 's friends is chosen as the initial node, and the second step is k being chosen with probability q . This probability is $p_k = \frac{1}{N} D_k \times q = \frac{qD_k}{N}$. The expected degree of a chosen ‘‘seed’’ node is then the degree-weighted probability:

$$\frac{\sum_{k \in V} p_k D_k}{\sum_{k \in V} p_k} = \frac{\sum_{k \in V} \frac{1}{N} q D_k^2}{\sum_{k \in V} \frac{1}{N} q D_k} = \frac{\frac{1}{N} \sum_{k \in V} D_k^2}{\frac{1}{N} \sum_{k \in V} D_k} = \frac{\mu_D^2 + \sigma_D^2}{\mu_D} = \mu_G$$

□

Similar logic applies if we choose any arbitrary initial sample of size M as long as the network is large, i.e. $N \gg M$.

Observation. *Denote an undirected tie (a, b) as a connection between nodes a and b . For any network with a given distribution of degrees, the distribution of degrees is unchanged if any two ties (a, b) and (c, d) are rewired to either (i) $(a, c), (b, d)$ or (ii) $(a, d), (b, c)$.*

Each of these nodes loses one tie and gains another and therefore the degrees are unchanged.

Theorem S6. [Rewiring Theorem] Let network $\mathcal{G} = (V, E)$ with $N > 3$ nodes include nodes a, b, c, d with degrees ordered as: $D_a \leq D_b < D_c \leq D_d$. If G containing edges $(a, b), (c, d) \in E$, but $(a, d), (b, c) \notin E$ is rewired to network $\mathcal{G}' = (V, E')$, containing edges $(a, d), (b, c) \in E'$, but $(a, b), (c, d) \notin E'$, then \mathcal{G}' has higher local mean than \mathcal{G} , but both networks have the same global mean.

Proof. First, observe that the degree distribution is unaffected by the change, and therefore the global mean (which only depends on mean and variance of the degree distribution) is also unaffected, i.e. $\mu_G(\mathcal{G}) = \mu_G(\mathcal{G}')$. Recall that the local mean is $\mu^L = \frac{1}{N} \sum_i \sum_j e_{ij} \left[\frac{D_i}{D_j} + \frac{D_j}{D_i} \right]$. Since between \mathcal{G} and \mathcal{G}' the degrees of all nodes are the same, and all edges are the same except the two rewired edges, we can write the difference between the local means the local means as:

$$\begin{aligned} \mu^L(\mathcal{G}') - \mu^L(\mathcal{G}) &= \frac{1}{N} \left[\left(\frac{D_a}{D_d} + \frac{D_d}{D_a} + \frac{D_b}{D_c} + \frac{D_c}{D_b} \right) - \left(\frac{D_a}{D_b} + \frac{D_b}{D_a} + \frac{D_c}{D_d} + \frac{D_d}{D_c} \right) \right] \\ &= \frac{1}{N} \left[(D_d - D_b) \left(\frac{1}{D_a} - \frac{1}{D_c} \right) + (D_c - D_a) \left(\frac{1}{D_b} - \frac{1}{D_d} \right) \right] > 0 \end{aligned}$$

The last inequality follows from the ordering of the node degrees. Note that we actually only require the conditions $D_b < D_d$ and $D_a < D_c$ to hold. □

S.C Data on Real Networks

We use a wide variety of real networks to determine properties and illustrate of the networks as it relates to the interventions detailed in the paper. We use data from two repositories.

S.C.1 Koblenz Network Collection

The networks are selected across several categories (Affiliation, Face-to-face Social, Online Social, Computer, Infrastructure and Biological networks), and span a wide range in network characteristics like size and density (Table S2). These networks also vary widely in terms of their size, from a low of 25 to networks with millions of nodes (e.g. Youtube). All network data was obtained from the Koblenz Network Collection [27]. We examine these real networks on a number of dimensions, the number of nodes, edges and the variation in the degree distribution.

Table S2: Real Network Characteristics

Label	Network Name	Nodes	Edges	Min Degree	Max Degree
<i>Collaboration</i>					
A1	Actor-Movie	383640	1470338	1	655
A2	Club Mmembers	25	91	3	20
A3	Citation (Physics)	28045	3148413	1	4909
A4	Citation (CS)	317080	1049865	1	343
<i>Face-toFace Interaction</i>					
FS1	Physician	117	464	2	26
FS2	Adolescent	2539	10454	1	27
FS3	Contact	274	2124	1	101
FS4	Conference	410	2765	1	50
<i>Online Social</i>					
OS1	PGP Users	10679	24315	1	205
OS2	Flickr	105722	2316667	1	5425
OS3	Advogato	5042	40509	1	803
OS4	Twitter	465016	833539	1	677
<i>Topology of Computer Networks</i>					
C1	Internet Topology	34761	107719	1	2760
C2	WWW (Google)	855802	4291352	1	6332
C3	Gnutella P2P	62561	147877	1	95
<i>Infrastructure</i>					
I1	Power Grid	4941	6593	1	19
I2	US Airports	1572	17214	1	314
I3	CA Roads	1957027	2760387	1	12
<i>Biological</i>					
B1	Human Protein 1	2783	6222	1	129
B2	Human Protein 2	5973	146385	1	855
B3	Yeast Protein	1458	1970	1	56
B4	C. Elegans	453	2033	1	237

S.C.2 India Village Networks

In addition, we also use data from $N = 75$ villages in India made publicly available (see [4] for details). The summary statistics for those village household networks are detailed in Table S3.

Table S3: Summary Statistics of Village Networks

Network Statistic	Mean	SD	Min	Max
Number of households	216.69	61.22	77	356
Number of (undirected) edges	993.31	348.77	334	2015
Density	0.05	0.02	0.02	0.11
Degree Mean	9.10	1.573	6.13	12.78
Degree Variance	52.03	19.88	27.80	124.56

S.D Individual Friendship Paradox

A basic view of the friendship paradox is developed by plotting the average number of friends (degree) of individual nodes' "friends" on the vertical axis against the average degree (Fig. S2, Fig. S3). For example, in the Contact (In person Social) network, we see a deep blue region above and to the left of the 45° line. Although present across all networks, the pattern is most prominent in the WWW (Google) or Twitter (Online Social) network. Observe also that in the Road Network, only $\Delta = 37\%$ of nodes have a higher average number of friends of friends than their own degree.

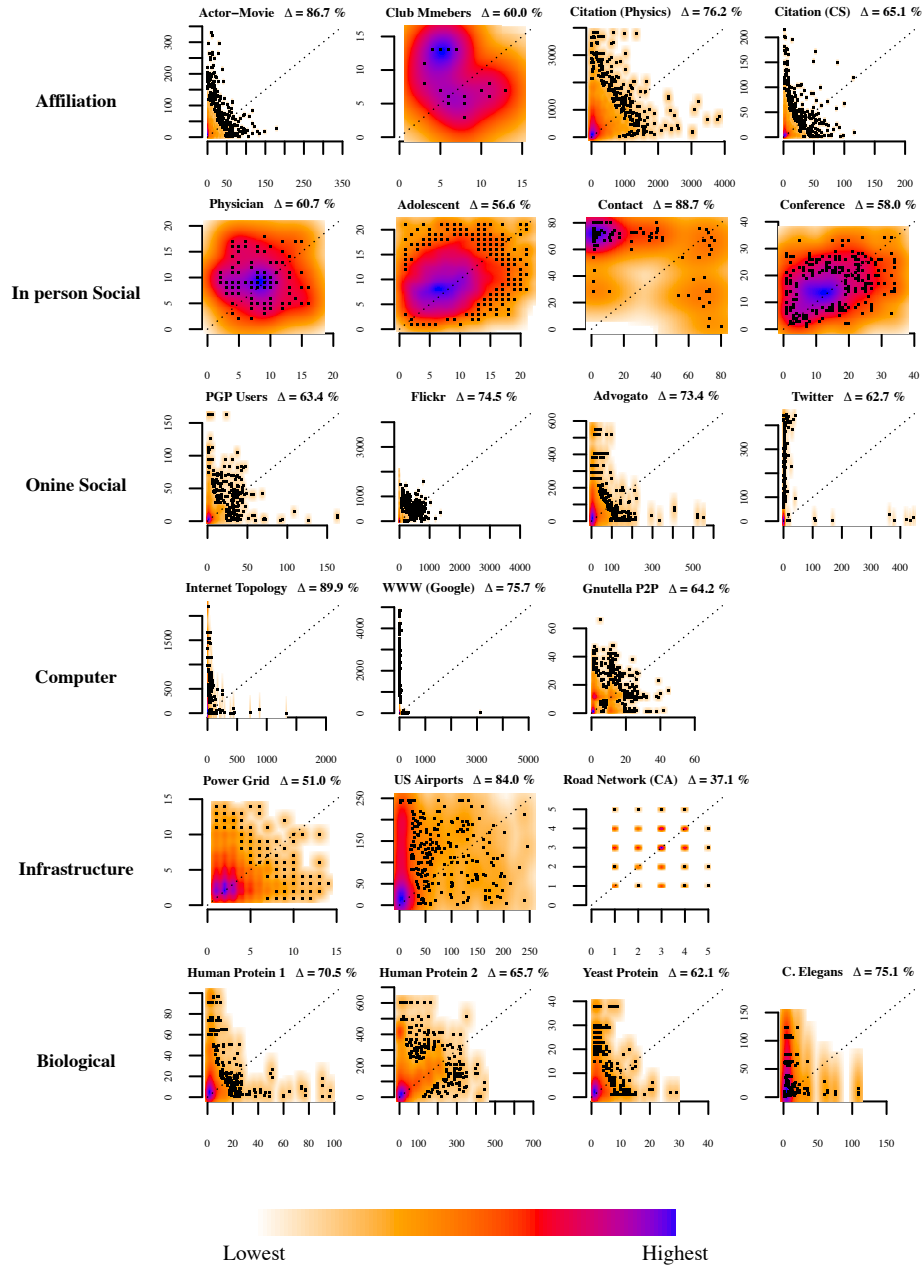


Figure S2: Friendship Paradox at Individual Level. Density plot of average number of friends of nodes compared to node degree in networks. Δ indicates the proportion of nodes that have a higher average number of friends of friends than their degree. Lowest density regions within each network are marked by white / orange, and highest density regions are marked in blue. For all networks, the highest density region lies above and to the left of the 45 degree line. For some networks like Adolescent Health or Road Network (CA), it is relatively more evenly distributed both above and below the 45 degree line, whereas for networks like Internet Topology or Twitter, the distribution is skewed above and to the left.

We illustrate this “individual friendship paradox” using a scatterplot of the node degree versus the average friend degree in Figure S3. Nodes that have a higher degree than their average friends do are colored red, whereas nodes that have lower degree are colored blue. Across most real networks, we observe that the blues vastly outnumber the reds. Relatedly, there are several nodes with low degrees whose friends on average have a high degree.

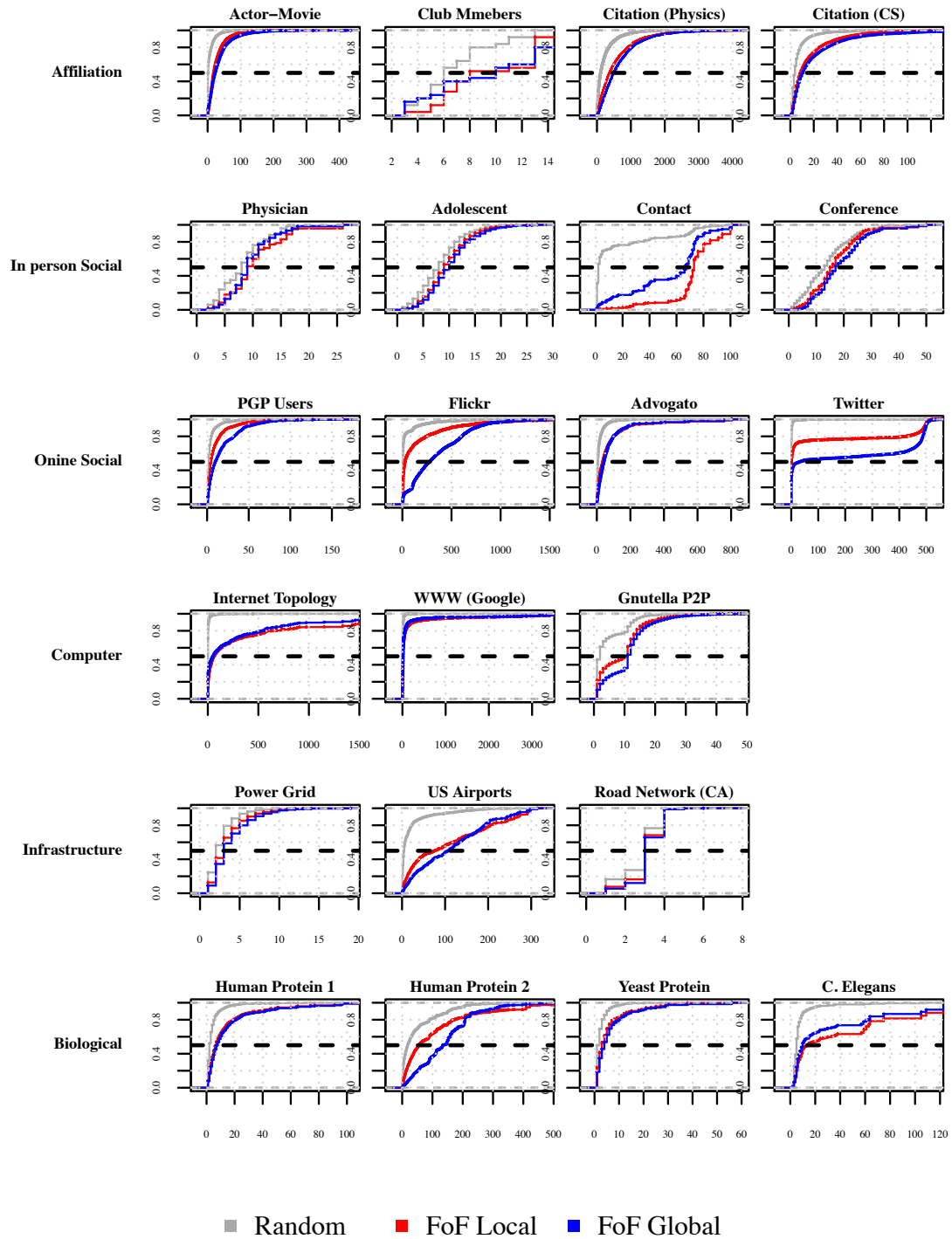


Figure S3: Individual Friendship Paradox. Empirical Cumulative Distribution Functions (CDF) of Real Networks. Panels show the CDF of 3 different network properties at the individual node level. For a specific node degree, the probability that a node with a lower (or identical) degree is chosen by the sampling strategy for random sampling (gray), local FoF sampling (red) and global FoF sampling (green). Across all networks, for lower degrees, the random sampling curve is to the left of the local and global FoF curves. In several networks, global FoF is to the left and higher than local FoF (e.g. Contact), whereas in others, it is to the right (e.g. Flickr).

S.E Leverage in Generated Networks

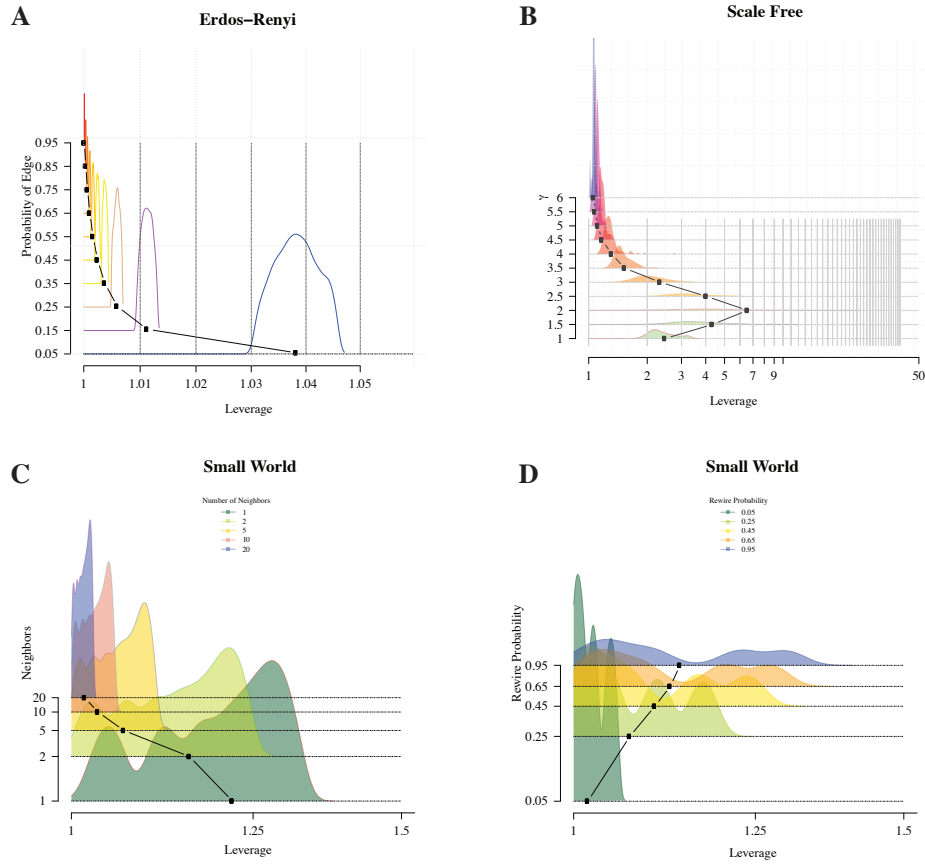


Figure S4: Local Leverage Density in Generated Networks from three different generative models, and spans the parameter space. A sample of 1,000 networks was used for each of the models. (A) Erdos-Renyi (ER) networks generated with edge probabilities, $p \in [0.05, 0.95]$, and size ranging from $N=50$ to $N=1000$ nodes. We find that local leverage is highest for the lowest edge probabilities, and leverage converges to 1 as the networks become more dense. (B) Static Scale Free (BA or Barabasi Albert) networks with scale-free parameter $\gamma \in [1, 6]$. For these networks, observe that the leverage spans a wider range, e.g. for $\gamma = 2$, the samples range from leverage of 1 to over 40. The mean leverage is non-monotonic in terms of γ , increasing when $\gamma < 2$ and decreasing for $\gamma > 2$. The distribution of leverage across the samples also displays decreasing variance when $\gamma > 2$. At very high levels of $\gamma \approx 6$, the local mean converges to the mean degree. With small world (Watts-Strogatz) networks, we have two parameters. First is the number of neighbors each node is connected to initially, n . The edges are then rewired with a specified probability, p_r . First, in panel (C), we find that with a small number of neighbors, the leverage distribution is quite spread out, and there is a substantial leverage effect. However, as we begin to create very dense networks, both the mean and the variance of the leverage distribution leverage diminish substantially. Second, we examine the impact of rewiring probability on the leverage distribution in panel (D). We find that with lower rewiring probabilities, say $p_r = 0.05$, the leverage distribution is closer to 1, whereas with a higher rewiring probabilities, the distributions feature increased variance as well as higher mean leverage.

S.F How Is Inversity Different from Degree Assortativity?

A natural question is whether inversity captures the same information (with opposite sign) as degree assortativity, which is a well known network property $\rho_a = \text{Corr}(D^O, D^D)$ capturing the correlation in degree across all edges in the network [30, 31, 37]. To examine this question, we generate 1000 networks using different generative methods as above. We find that assortativity and inversity are not guaranteed to have opposite signs (Fig. S5A). Therefore, the sign of assortativity cannot be used to determine whether the local or global mean is greater for a network, unlike with inversity. All 3 network generating processes create networks with the same sign for both metrics (detail in Fig. S5B). Example networks for the case of same sign assortativity and inversity are illustrated (Fig. S5C, S5D), showing that it is not obvious to predict inversity of a network if we know its assortativity.

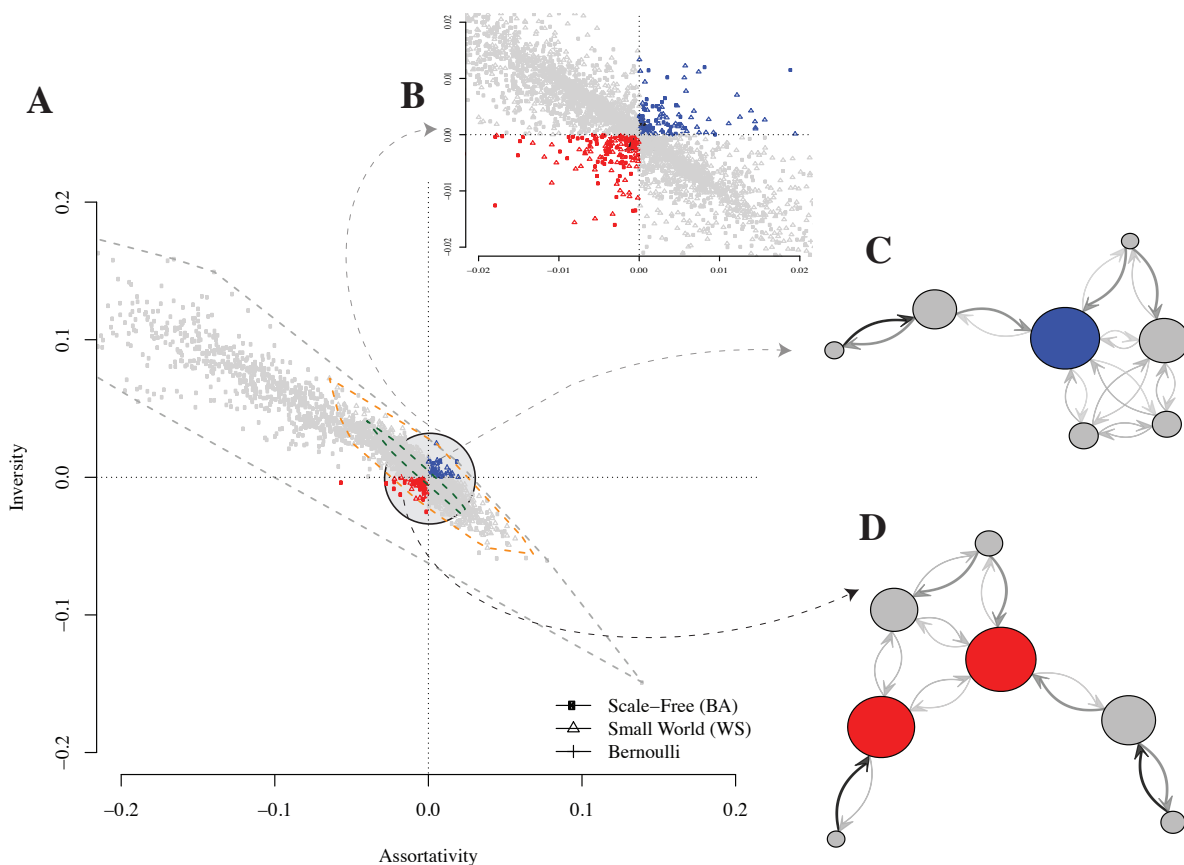


Figure S5: Assortativity and Inversity. $N=1000$ networks are generated from three classes of networks. (A) Erdos-Renyi (ER), Scale Free (BA or Barabasi Albert) and Small World (WS or Watts Strogatz), parameters detailed in legend of Figure S4. Observe the regions in red and blue, where networks have the same sign of assortativity and inversity. (B) Detailed view of region around (0,0) showing all three network types can produce networks with same sign for assortativity and inversity. (C) Example network with $N=7$ nodes where assortativity and inversity are both positive. (D) Similar example where both measures are negative. Overall, it demonstrates that inversity and assortativity are not equivalent measures, e.g. using assortativity in place of inversity could result in using a global strategy when local may be more appropriate.

S.G Virus Propagation Models

We detail below several examples of virus propagation models being used for characterizing the transmission and spread of diseases. These models build upon the early work of Kermack and McKendrick [29]. All individuals in a population (in our case, the nodes in a network) are in one of the states, either susceptible (S) or infected (I). Based on the viral propagation, they can move to other states like Exposed (X), Recovered (R), or Deceased (D). For example, the SIR model involves individuals being in one of three states, (S), (I) or (R) and transitioning between the states probabilistically. Typically, the vast majority of nodes are present in the susceptible state (S), in which they might contract the disease. The exposed state (X) is used to indicate a node that has been exposed to the disease, but could be asymptomatic during an incubation period and is not capable of infecting others. In contrast, the infected state (I) indicates a node that is capable of infecting others. The (R) recovered state implies permanent immunity. There are further extensions possible, e.g. adding infants who have maternal antibodies (state M) that provide passive immunity. See [8] or [20] for an overview and survey of these models. These models have been extensively used in epidemiological studies to characterize disease dynamics as detailed in Table S4, including measles, influenza and COVID-19.

There has been recent notable work that aims to characterize the epidemic thresholds of these compartmental models with disease transmission over a network [13, 33]. The critical idea is that the epidemic threshold of a network can be characterized as the inverse of the greatest (first) eigenvalue of the adjacency matrix A of the network, denoted as:

$$\tau(A) = \frac{1}{\lambda_1(A)}$$

Eigenvalue λ_1 termed the spectral radius characterizes the connectivity of the network graph. Thus, networks that have higher connectivity or λ_1 are more likely to allow a contagion different paths to grow into an epidemic, whereas in networks with low connectivity, the epidemic is more likely to die out.

While there have been a number of epidemic thresholds for specific network generating processes (e.g. small world), the generality of the result above is valuable since it allows: (a) any arbitrary network, without placing restrictions on its topology or structure, (b) a wide range of compartmental models like SIS, SIR and others detailed in Table S4 typically used to model infectious disease.

Consider a SIR model for illustration, the results also hold for the other models. The model is parametrized by two rates: β is the probability of an infected node infecting a susceptible node in a given time period, and δ is the probability at which an infected node recovers (or is cured) during the period. If time is continuous, β and δ can be viewed as the rates of infection and recovery. In either case, \mathcal{R}_0 is defined as $\mathcal{R}_0 = \frac{\beta}{\delta}$.

The epidemic threshold τ is defined as follows [13]:

$$\begin{cases} \mathcal{R}_0 = \frac{\beta}{\delta} < \tau(E) \implies \text{infection dies out over time} \\ \mathcal{R}_0 = \frac{\beta}{\delta} > \tau(E) \implies \text{infection grows over time} \end{cases}$$

There are a few observations relevant here. First, the critical value of epidemic threshold is a function of the adjacency matrix E of the network topology (structure) \mathcal{G} . Second, a network topology with a higher epidemic threshold is less likely to have an epidemic. Third, interventions like immunizing nodes or reducing the number of connections (edges) can increase the threshold $\tau(E)$ so that infections are more likely to die out.

Table S4: Virus Propagation Models Used for Diseases

Virus Propagation Model	Infectious Diseases [References]
SIS	Malaria ([39])
SIR	Measles [19], Swine Flu H1N1 [38], Ebola [6]
SXIR	Chicken Pox [15], SARS [36], COVID-19 [35]
SIRD	COVID-19 ([9])

Note: The states refer to **(S)**usceptible, **(I)**nfectious, **(R)**ecovered / **(R)**emoved, **(X)**Exposed, **(D)**eceased

Implementation of VPM

We begin with a seed set of 1% of the nodes being infected, and evaluate epidemic outcomes using the SIR model. All the nodes in the network that are not infected or recovered are susceptible (S) to the infection. Each infected node can transmit an infection in each period probabilistically to each of its neighbors. The probability of an infection is $P_{\text{transmit}} = \beta$. Thus, a node can become infected (I) from contact with any of its neighbors. In each period, an infected node can be cured or recovered (R) probabilistically, with the likelihood $P_{\text{cure}} = \delta$. Recovered nodes cannot be reinfected and cannot transmit infections.

The process of immunizing (or vaccinating) a set of nodes involves choosing a proportion of nodes (5%, or 10% or 20%) and ensuring that these nodes do not transmit any disease. The nodes for immunization are chosen based on three strategies: random, local and global. The parameters used in the simulation of the epidemic are detailed in Table S5.

Table S5: Parameters of SIR Network Propagation Model

Parameter	Value	Description
$P_{\text{transmit}} = \beta$	0.20	Probability of an infected node transmitting the disease to a susceptible neighbor.
$P_{\text{cure}} = \delta$	0.15	Probability of an infected node recovering. Thus, moving from (I) \implies (R) is $P_{I \rightarrow R} = P_{\text{cure}}$, and $P_{I \rightarrow I} = 1 - P_{\text{cure}}$
$P_{S \rightarrow I}^k$	$1 - (1 - \beta)^{N_k^{\text{infected}}}$	Probability of a susceptible node k becoming infected. Depends on the number of infected neighbors N_k^{infected} . Thus, k can become infected through <i>any</i> of its infected neighbors. So we have: $P_{S \rightarrow I}^k = 1 - (1 - P_{\text{transmit}})^{N_k^{\text{infected}}}$. Similarly, $P_{S \rightarrow S}^k = (1 - P_{\text{transmit}})^{N_k^{\text{infected}}}$.
n_{infected}^0	1%	Proportion of nodes in network that are infected at the beginning
n_{sim}	100	Number of simulations

Note: (S)usceptible, (I)nfected, (R)ecovered / (R)emoved

Thus, a strategy A is better than an alternative strategy B if it results in lower levels of peak infections, total infections and total suffering.

S.H Epidemic Outcomes

In Figure S6, we examine the epidemic propagation characteristics on the Facebook network [45] using the same parameters as detailed in Table S5. The epidemic could either be viewed as an *informational* epidemic propagating through Facebook. Alternatively, one might consider the Facebook network structure to serve as an approximation of contact network for the purposes of this evaluation.

We evaluate epidemics using the following metrics:

- **Proportion Infected at Peak** $= \frac{1}{N} \max_t (\sum_i I_{it})$: Since epidemics increase in intensity and eventually die down, an important characteristic is to measure the proportion of the population who are infected at the peak of the epidemic. This directly impacts important decisions like hospital capacity planning etc.
- **Proportion Ever Infected** $= \frac{1}{N} \sum_i \max_t (I_{it})$: The proportion of the population that was ever infected by the disease is important since it represents the total spread of the disease in the population. It could also represent the number of people who might have immunity to future recurrences of the disease.
- **Total Suffering**: $\frac{1}{NT} \sum_i \sum_t (I_{it})$ Here, the total suffering metric captures not just how many infections occur, but also the length of the infections. This represents the proportion of individual-period combinations with an infection.

For the Facebook network sample, we find that an epidemic's outcomes are better when using the local strategy compared to the global strategy, which in turn are better than the random strategy. This conclusion holds for all the metrics considered above.

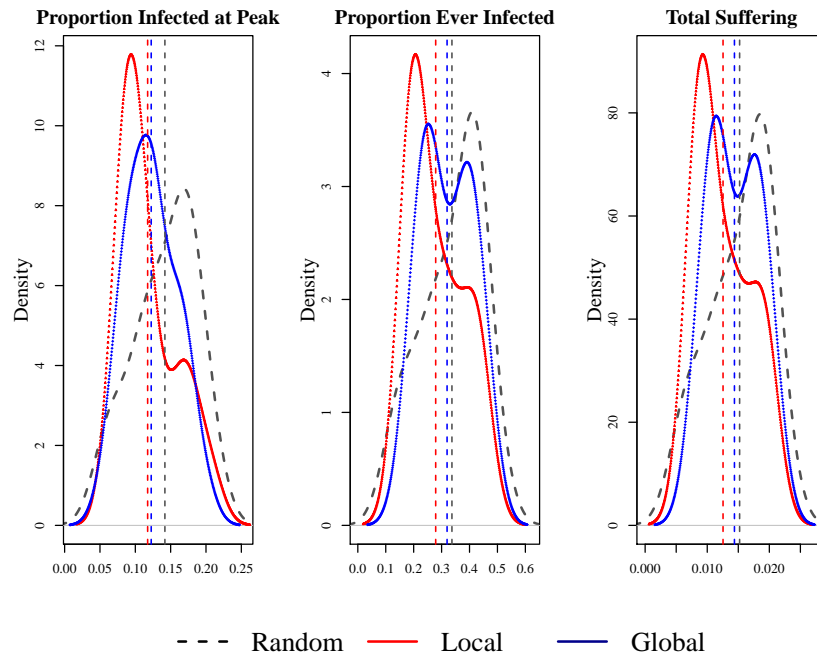


Figure S6: Epidemic Outcomes with Immunization in Facebook Network. See Table S5 for parameters of simulation. All outcomes are density plots. We plot 3 outcomes: (a) the proportion of population infected at the peak, (b) proportion of population that was ever infected, and (c) total suffering. The x -axis represent proportions and the y -axis represent density. We plot the outcomes for 3 strategies: (R)andom, (L)ocal and (G)lobal. The dashed vertical lines represent the means for the 3 strategies. We find that for the Facebook network, the Local strategy is better for all outcomes than the Global, which in turn is better than the Random strategy.