

1

## <sup>2</sup> Supporting Information for

### <sup>3</sup> On the Friendship Paradox and Inversity: A Network Property with Applications to <sup>4</sup> Privacy-sensitive Network Interventions

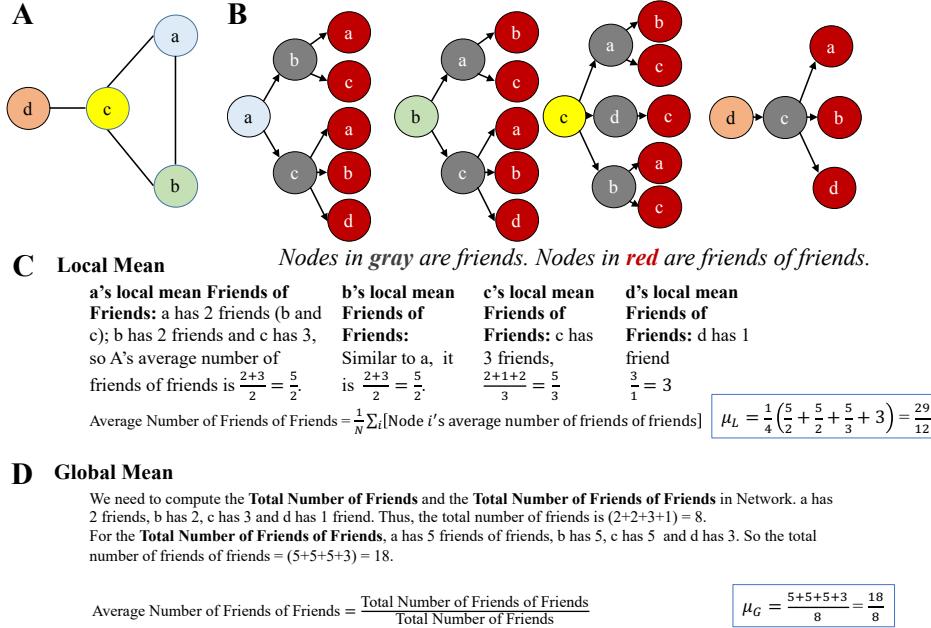
<sup>5</sup> Vineet Kumar, David Krackhardt and Scott Feld

<sup>6</sup> Corresponding Author: Vineet Kumar.  
<sup>7</sup> E-mail: vineet.kumar@yale.edu

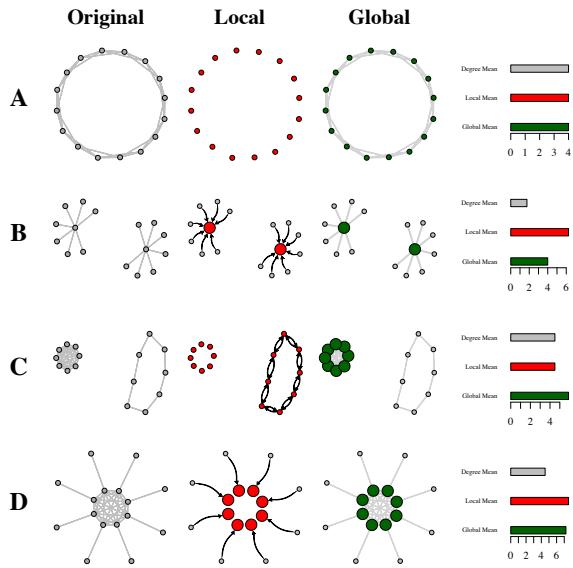
#### <sup>8</sup> This PDF file includes:

<sup>9</sup> Figs. S1 to S7  
<sup>10</sup> Tables S1 to S5  
<sup>11</sup> SI References

12 S.A. Local and Global Mean in Example Networks



**Fig. 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  $\mu_L = \frac{1}{4} \left( \frac{5}{2} + \frac{5}{2} + \frac{5}{3} + 3 \right) = 2.42$ . **(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,  $(2 + 2 + 3 + 1) = 8$ .



**Fig. S2. 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 Detached 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.

13 **S.B. Mathematical Appendix**

14 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  
 15 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  $\hat{E}$   
 16 including both  $(i, j)$  and  $(j, i)$  as distinct elements of  $\hat{E}$  corresponding to an undirected edge  $i \leftrightarrow j$ . Since the purpose is to  
 17 conduct network interventions, we only consider nodes that are connected, and exclude isolates from the network. We detail  
 18 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 $i$	$\frac{1}{D_i} \sum_{j \in \mathcal{N}(i)} D_j$
$\mu_D, \sigma_D^2$	Mean and variance of Degrees	$\frac{1}{N} \sum_i D_i, \frac{1}{N} \sum_i (D_i - \mu_D)^2$
$\mu_L$	Local Mean	$\frac{1}{N} \sum_i F_i$
$\mu_G$	Global Mean	$\frac{\sum_i D_i F_i}{\sum_i D_i}$
$\rho$	Inversity	$\text{Corr} \left( D_i, \frac{1}{D_j} \right) \forall (i, j) \in \hat{E}$

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

22 **Theorem S1.** *For a finite network  $\mathcal{G} = (V, E)$  and  $\mathcal{N}_i$  is the set of  $i$ 's connections. We find the following:*

- 23 (i) *The friendship paradox statement, "on average, your friends have more friends than you do," specified as  $\frac{1}{|\mathcal{N}(i)|} \left( \sum_{j \in \mathcal{N}_i} D_j \right) >$*   
 24  $D_i \forall i \in V$ , *cannot hold for all nodes in  $\mathcal{G}$  or any connected component of  $\mathcal{G}$ .*
- 25 (ii) *There exists  $\mathcal{G}$  for which the friendship paradox statement holds true for all nodes, except one.*

26 *Proof.* Consider a multi-component network with  $C$  components,  $V = \bigcup_{k=1}^C \mathcal{C}_k$ , where each component  $\mathcal{C}_k$  represents the set of  
 27 nodes in a connected network.

28 To prove part (i) of the theorem, first consider each of the components in turn, with  $k = 1$ . First, in the trivial case  
 29 of a degree-regular component, part (i) trivially holds. Next, consider the case with degree variation within component  $k$ .  
 30 Within  $\mathcal{C}_k$ , for a finite network, there must be a finite set of nodes  $\mathcal{V}_k^{\max}$  with maximum degree within this component. At  
 31 least one of the nodes in  $\mathcal{V}_k^{\max}$  must then be connected to a node of lower degree; otherwise, the component would not be  
 32 fully connected. Now, for that node, call it  $i \in \mathcal{V}_k^{\max}$  connected to a node of lower degree, the friendship paradox statement  
 33  $\frac{1}{|\mathcal{N}(i)|} \left( \sum_{j \in \mathcal{N}_i} D_j \right) > D_i \forall i \in \mathcal{C}_k$  cannot hold. Thus, for each component  $k$ , there is at least one node for which the friendship  
 34 paradox statement does not hold. In the overall network  $\mathcal{G}$ , there must be at least  $C$  nodes for which the friendship paradox  
 35 statement cannot hold.

36 For part (ii), we only need to consider the star or hub and spoke network. The friendship paradox statement can easily be  
 37 verified to hold for all nodes except the central node.  $\square$

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

40 *Proof. (as given in Feld, 1991).*  $\mu_G = \frac{\sum_i \sum_j e_{ij} D_j}{\sum_i D_i} = \frac{\sum_i D_i^2}{\sum_i D_i} = \frac{\mu_D^2 + \sigma_D^2}{\mu_D}$ . We note that the above proof is not affected by  
 41 isolates, since they add zero to both the numerator and denominator, leaving  $\mu_G$  unchanged, whether or not we remove these  
 42 isolates.  $\square$

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

$$44 \quad \mu_L = \mu_D + \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[ \frac{(D_i - D_j)^2}{D_i D_j} \right] \quad [1]$$

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

46 *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  
 47 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  
 48 mean is defined as:

$$49 \quad \mu_L = \frac{1}{|V|} \sum_i F_i = \sum_{i \in V} \left[ \frac{1}{D_i} \left( \sum_{j \in \mathcal{N}(i)} D_j \right) \right]$$

50 Rewriting the expression for  $\mu_L$  in terms of the connections (edges) between individuals, we obtain:

$$\begin{aligned} 51 \quad \mu_L &= \frac{1}{|V|} \sum_{i \in V} \left[ \frac{1}{D_i} \left( \sum_{j \in V} e_{ij} D_j \right) \right] = \frac{1}{|V|} \sum_{i \in V} \sum_{j \in V} \left[ e_{ij} \frac{1}{D_i} (D_j) \right] \\ 52 &= \frac{1}{2|V|} \sum_{(i,j) \in V \times V} \left[ e_{ij} \left( \frac{D_j}{D_i} \right) + e_{ji} \left( \frac{D_i}{D_j} \right) \right] = \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[ \frac{D_j}{D_i} + \frac{D_i}{D_j} \right] \\ 53 &= \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[ \frac{D_j^2 + D_i^2}{D_i D_j} \right] = \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[ \frac{(D_i - D_j)^2 + 2D_i D_j}{D_i D_j} \right] \\ 54 &= \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[ \frac{(D_i - D_j)^2}{D_i D_j} \right] + \frac{1}{2|V|} (4|E|) \\ 55 &= \mu_D + \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[ \frac{(D_i - D_j)^2}{D_i D_j} \right] \end{aligned}$$

□

56 Note that what we characterize as the local mean defined as above was independently shown to be greater than the mean  
 57 degree, including by the present authors at (1). It has also been documented by others, including by Christian Borgs & Jennifer  
 58 Chayes in a comment to an article by (2), and by (3). However, the properties of the local mean have not been formally  
 59 examined and characterized.

60 For the results below, we consider networks without isolates, even though the proofs hold in their presence.

61 **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  
 62 connected by the following relationship involving the inversity  $\rho$  and the -1, 1, 2, and 3rd moments of the degree distribution:

$$63 \quad \mu_L = \mu_G + \rho \sqrt{\left( \frac{\kappa_1 \kappa_3 - \kappa_2^2}{\kappa_1} \right) [\kappa_{-1} - (\kappa_1)^{-1}]}$$

64 *Proof.* Define the moments of the degree distribution as:  $\kappa_m = \frac{1}{N} \sum_i D_i^m$ . Since we defined  $\rho$  the measure of inversity as the  
 65 correlation of two distributions that we specify as the origin degree (**O**) and inverse destination degree (**ID**) distributions. The  
 66 **O** distribution consists of the degree of nodes corresponding to edges, and **ID** distribution consists of the inverse degree of  
 67 nodes corresponding to edges. Thus, each connection (edge) contributes two entries to each distribution. For example, if there  
 68 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  
 69 both distributions multiple times based on degree.

70 Next, we detail the mean and variance of the distributions. First, we consider the means. The mean of the origin distribution  
 71 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  
 72 variances. The variance of the origin distribution (**O**) is computed as:

$$\begin{aligned} 73 \quad \sigma_O^2 &= \frac{1}{2|E|} \sum_{(i,j) \in E} (D_i - \mu_O)^2 = \frac{1}{2|E|} \sum_{i \in V} D_i (D_i - \mu_O)^2 \\ 74 &= \frac{1}{N \mu_D} \sum_{i \in V} [D_i^3 - 2\mu_O D_i^2 + (\mu_O)^2 D_i] = \frac{\kappa_3}{\kappa_1} - \left( \frac{\kappa_2}{\kappa_1} \right)^2 \end{aligned}$$

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

$$\begin{aligned} 77 \quad \sigma_{ID}^2 &= \frac{1}{2|E|} \sum_{(i,j) \in E} \left[ \left( \frac{1}{D_j} - \frac{1}{\mu_D} \right)^2 \right] = \frac{1}{2|E|} \sum_{(i,j) \in E} \left( \frac{1}{D_j^2} + \frac{1}{\mu_D^2} - \frac{2}{\mu_D D_j} \right) \\ 78 &= \frac{1}{2|E|} \left[ \sum_{(i,j) \in E} \frac{1}{D_j^2} + \frac{1}{\mu_D^2} \left( \sum_{(i,j) \in E} 1 \right) - \frac{2}{\mu_D} \sum_{(i,j) \in E} \frac{1}{D_j} \right] = \frac{1}{2|E|} \left[ \sum_{j \in V} \frac{1}{D_j} + \frac{1}{\mu_D^2} 2|E| - \frac{2}{\mu_D} N \right] \\ 79 &= \frac{1}{\mu_D N} \left[ \sum_{j \in V} \frac{1}{D_j} \right] - \frac{1}{\mu_D^2} = (\kappa_1)^{-1} [\kappa_{-1} - (\kappa_1)^{-1}] \end{aligned}$$

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

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

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

$$88 \quad \mu_L = \mu_G + \rho \sqrt{\left( \frac{\kappa_1 \kappa_3 - \kappa_2^2}{\kappa_1} \right) [\kappa_{-1} - (\kappa_1)^{-1}]} \quad [2]$$

89  $\square$

90 **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:

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

91  $\square$

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

93 **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 there are nodes  $a, b, c, d \in V$  such that  $(a, b), (c, d) \in E$ , but  $(a, d), (b, c) \notin E$ , then by rewiring the network to  $\mathcal{G}' = (V, E')$ , containing edges  $(a, d), (b, c) \in E'$ , but  $(a, b), (c, d) \notin E'$ , we obtain:  $\mu_L(\mathcal{G}') > \mu_L(\mathcal{G})$  whereas  $\mu_G(\mathcal{G}') = \mu_G(\mathcal{G})$ . Also, it follows that  $\rho(\mathcal{G}') > \rho(\mathcal{G})$ .*

97 *Proof.* First, observe that the degree distribution is unaffected by the change, since each node's degree is unchanged by the  
98 rewiring. Therefore, the global mean (which only depends on mean and variance of the degree distribution) is also unaffected,  
99 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  
100 nodes are the same, and all edges are the same except the two rewired edges, we can write the difference between the local  
101 means the local means as:

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

104 The last inequality follows from the ordering of the node degrees. Note that we actually only require the conditions  $D_b < D_d$   
105 and  $D_a < D_c$  to hold. Since the degree distribution does not change with rewiring, by Theorem S4, we must have an increase  
106 in Inversity,  $\rho(\mathcal{G}') > \rho(\mathcal{G})$ .  $\square$

108 **S.C. Data on Real Networks**

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

111 **A. Koblenz Network Collection.** The networks are selected across several categories (Affiliation, Face-to-face Social, Online  
 112 Social, Computer, Infrastructure and Biological networks), and span a wide range in network characteristics like size and  
 113 density (Table S2). These networks also vary widely in terms of their size, from a low of 25 to networks with millions of nodes  
 114 (e.g. Youtube). All network data was obtained from the Koblenz Network Collection (KONECT) (4). We examine these real  
 115 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	$N$	$ E $	$\mu_D$	$\mu_L$	$\mu_G$	$\rho$
<i>Collaboration</i>							
A1	Actor-Movie	383640	1470338	7.67	36.29	35.12	0.02
A2	Club Mmebers	25	91	7.20	10.39	9.39	0.39
A3	Citation (Physics)	28045	3148413	224.53	569.15	667.24	-0.06
A4	Citation (CS)	317080	1049865	6.62	18.53	21.75	-0.10
<i>Face-to-Face Interaction</i>							
FS1	Physician	117	464	7.93	10.19	9.95	0.09
FS2	Adolescent Health	2539	10454	8.23	9.85	10.49	-0.20
FS3	Contact	274	2124	15.50	74.78	56.69	0.26
FS4	Conference	410	2765	13.49	17.10	18.72	-0.19
<i>Online Social</i>							
OS1	PGP Users	10679	24315	4.55	13.46	18.88	-0.17
OS2	Flickr	105722	2316667	43.83	187.12	349.21	-0.22
OS3	Advogato	5042	40509	15.56	99.31	82.52	0.06
OS4	Twitter	465016	833539	3.58	437.74	226.53	0.65
<i>Topology of Computer Networks</i>							
C1	Internet Topology	34761	107719	6.20	530.34	319.46	0.22
C2	WWW (Google)	855802	4291352	10.03	226.59	170.35	0.06
C3	Gnutella P2P	62561	147877	4.73	13.22	11.60	0.15
<i>Infrastructure</i>							
I1	Power Grid	4941	6593	2.67	3.97	3.87	0.06
I2	US Airports	1572	17214	21.90	120.27	112.23	0.04
I3	CA Roads	1957027	2760387	2.82	3.15	3.17	-0.04
<i>Biological</i>							
B1	Human Protein 1	2783	6222	4.32	19.61	15.78	0.15
B2	Human Protein 2	5973	146385	48.81	117.83	143.31	-0.09
B3	Yeast Protein	1458	1970	2.67	9.65	7.13	0.31
B4	C. Elegans	453	2033	8.94	51.57	40.10	0.21

116 A few observations are worth noting here. First, there are many networks with both positive and negative values of inversty,  
 117 both within and across categories. Second, we do not see Inversty close to  $\pm 1$ . However, the Twitter network is closest in  
 118 magnitude, with an inversty of 0.65. Third, the variation in inversty is low in some categories like Infrastructure, whereas it is  
 119 relatively greater in Online Social networks. Finally, we see that even low values of inversty can impact the difference between  
 120 local and global mean significantly also between these and mean degree, e.g. the WWW (Google) network with an inversty of  
 121 0.06 displays meaningful differences. This is due to the multiplier effect of the moments of the distribution function, detailed in  
 122 equation 2.

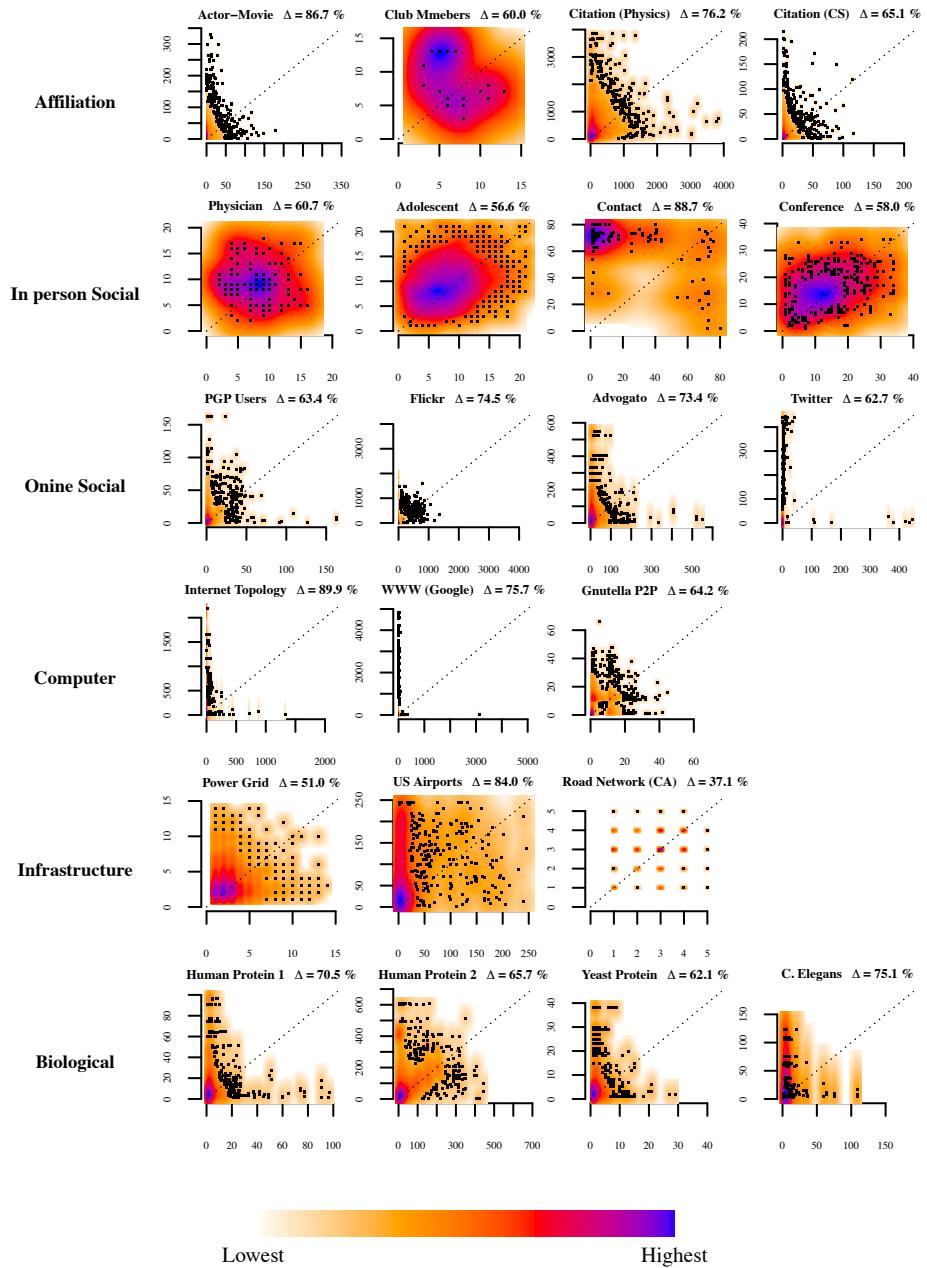
123 **B. India Village Networks.** In addition, we also use data from  $N = 75$  villages in India made publicly available (see (5) for  
 124 details). The summary statistics for those village household networks are detailed in Table S3.

**Table S3. Summary Statistics of Village Networks**

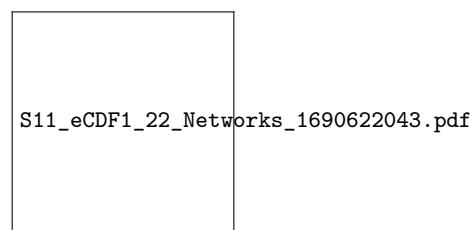
Network Statistic	Mean	SD	Min	Max
Number of households	216.69	61.22	77	356
Number of (undirected) edges	993.31	348.77	334	2015
Density	0.05	0.02	0.02	0.11
Degree Mean	9.10	1.573	6.13	12.78
Degree Variance	52.03	19.88	27.80	124.56

125 **S.D. Individual Friendship Paradox**

126 A basic view of the friendship paradox is developed by plotting the average number of friends (degree) of individual nodes'  
127 "friends" on the vertical axis against the average degree (Fig. S3, Fig. S4). For example, in the Contact (In person Social)  
128 network, we see a deep blue region above and to the left of the  $45^\circ$  line. Although present across all networks, the pattern is  
129 most prominent in the WWW (Google) or Twitter (Online Social) network. Observe also that in the Road Network, only  
130  $\Delta = 37\%$  of nodes have a higher average number of friends of friends than their own degree.



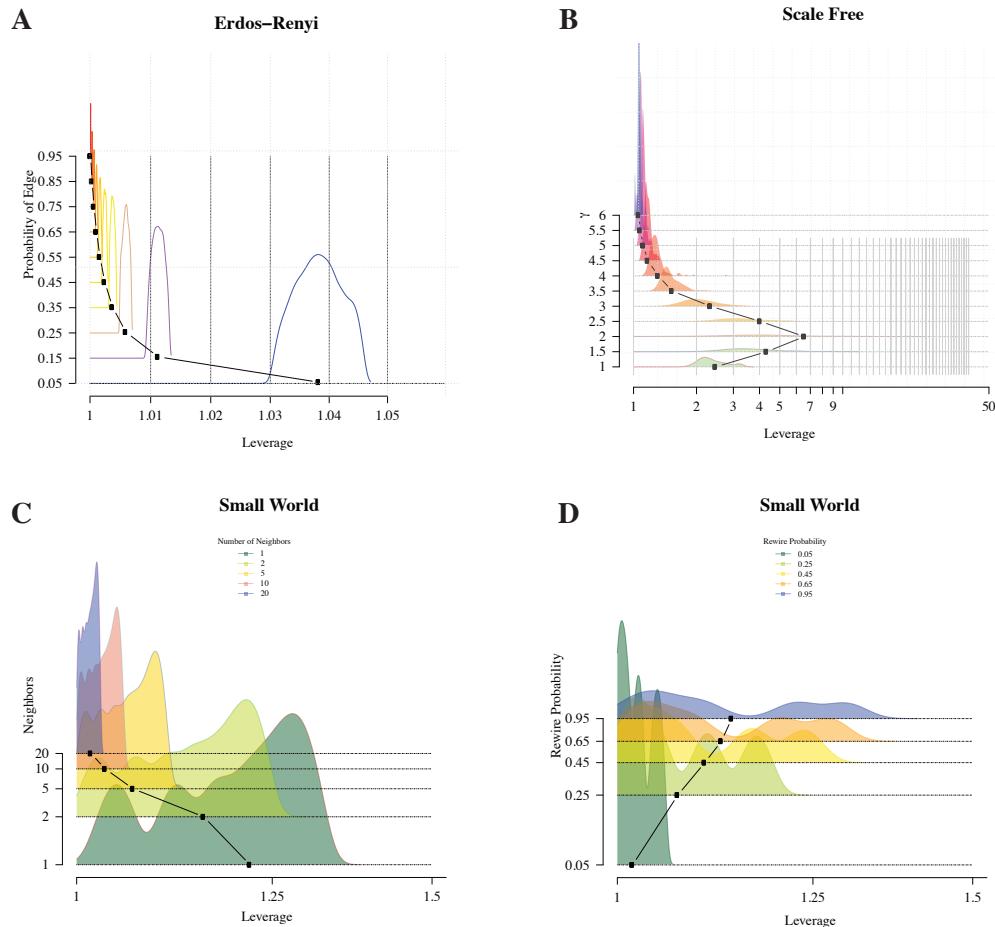
**Fig. S3.** Friendship Paradox at Individual Level. Density plot of average number of friends of nodes compared to node degree in networks.  $\Delta$  indicates the proportion of nodes that have a higher average number of friends than their degree. Lowest density regions within each network are marked 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.



**Fig. S4.** 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 Node Degree (gray), Local Mean of Friends (red) and Global Mean of Friends (green). Across all networks, for lower degrees, the Node Degree curve is to the left of the Local and Global Mean of Friends curves. In several networks, Global Mean of Friends is to the left and higher than Local Mean of Friends (e.g. Contact), whereas in others, it is to the right (e.g. Flickr).

131 We illustrate this “individual friendship paradox” using a scatterplot of the node degree versus the average friend degree in  
 132 Figure S4. Nodes that have a higher degree than their average friends do are colored red, whereas nodes that have lower degree  
 133 are colored blue. Across most real networks, we observe that the blues vastly outnumber the reds. Relatedly, there are several  
 134 nodes with low degrees whose friends on average have a high degree.

135 **S.E. Leverage in Generated Networks**



**Fig. S5.** 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  $\gamma$ , 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.

136 **S.F. Network Features that Affect Inversity**

137 Inversity is strongly dependent on the structure of connections, who is connected to whom. We observe in the main paper that  
138 star type network structures lead to positive values of inversity, whereas clusters or cliques contribute to negative inversity  
139 values.

140 **Star or hub-based networks** First, we observe that there is significant evidence for hub-based network structures appearing in  
141 real-world networks. Such hub-like structures are common across a wide range of networks, including co-author networks,  
142 the underlying interlink network that forms the Internet, as well as Airline networks (6–9). The early networks literature  
143 explaining the emergence of such hub-like patterns posited preferential attachment as a mechanism, where newly joining nodes  
144 connected disproportionately to highly connected nodes (10, 11). The economics literature involving the economic incentives  
145 underlying network formation posits that agents form links based on the expected benefits to such formation. The resulting  
146 network is an equilibrium outcome based on the decisions of each of the agents, who are maximizing their own utilities (12). In  
147 this stream, an influential paper (13) finds that even when agents are homogeneous, where they have identical constraints,  
148 preferences and incentives, *star networks arise across a wide range of equilibria*. Stars are predicted to occur even though  
149 agents are symmetric with identical incentives and opportunities. Complementing this research, star networks are found to  
150 arise over time in experimental settings where agents vary in terms of costs, incentives and even information (14).

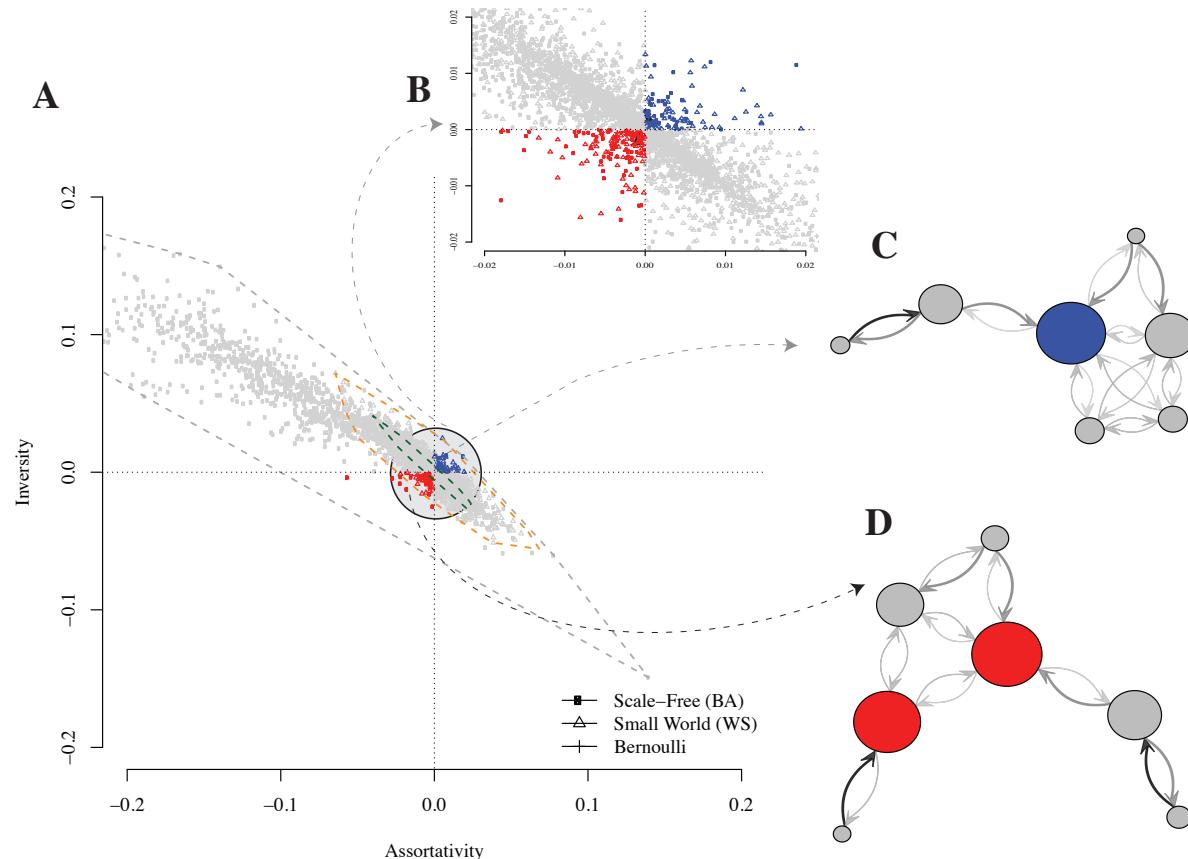
151 **Clusters and Communities in networks** Clusters or communities as well as cliques (fully connected or complete subnetwork) are  
152 commonly observed in networks. A typical conceptualization of community is the following: “Qualitatively, a community is  
153 defined as a subset of nodes within the graph such that connections between the nodes are denser than connections with the  
154 rest of the network.” (15). There are several reasons why communities form, including homophily and social foci. In homophily,  
155 when a number of individuals are similar, then they are much more likely to be connected to each other, and also part of the  
156 same larger grouping or community (16). However, it is to be noted that not all such connections will happen, rather that such  
157 connections and communities are more likely to happen when individuals are homophilous. Homophily has also been tied to  
158 polarization and segregation (17).

159 A prominent theory that explains how communities form is the idea of foci (18). The essential idea is that most ties originate  
160 around foci of activity, where a limited set of people share a focus of activity that organizes activity, and thereby tend to  
161 generate repeated interaction among the same people in the set over time that leads to ties among many of them. Each person  
162 tends to be associated with many different foci. Alters from the same focus tend to be tied to one another, but not those  
163 from separate foci. Consistent with this notion, research has found that the way organizational environments are structured  
164 moderates the tie-formation process (19).

165 An implication of this theory for the present paper is that larger denser foci of activity contribute large numbers of ties for  
166 all their tied participants, and small and/or sparse focused sets generate few for their participants. Thus, the size and density  
167 of focused sets may contribute to positive or negative inversity.

168 **S.G. How Is Inversity Different from Degree Assortativity?**

169 A natural question is whether inversity captures the same information (with opposite sign) as degree assortativity, which  
 170 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  
 171 (20–22). To examine this question, we generate 1000 networks using different generative methods as above. We find that  
 172 assortativity and inversity are not guaranteed to have opposite signs (Fig. S6A). Therefore, the sign of assortativity cannot be  
 173 used to determine whether the local or global mean is greater for a network, unlike with inversity. All 3 network generating  
 174 processes create networks with the same sign for both metrics (detail in Fig. S6B). Example networks for the case of same sign  
 175 assortativity and inversity are illustrated (Fig. S6C, S6D), showing that it is not obvious to predict inversity of a network if we  
 176 know its assortativity.



**Fig. S6.** 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 S5. 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.

177 **S.H. Virus Propagation Models**

178 We detail below several examples of virus propagation models being used for characterizing the transmission and spread of  
 179 diseases. These models build upon the early work of Kermack and McKendrick (23). All individuals in a population (in our  
 180 case, the nodes in a network) are in one of the states, either susceptible (S) or infected (I). Based on the viral propagation, they  
 181 can move to other states like Exposed (X), Recovered (R), or Deceased (D). For example, the SIR model involves individuals  
 182 being in one of three states, (S), (I) or (R) and transitioning between the states probabilistically. Typically, the vast majority  
 183 of nodes are present in the susceptible state (S), in which they might contrast the disease. The exposed state (X) is used to  
 184 indicate a node that has been exposed to the disease, but could be asymptomatic during an incubation period and is not  
 185 capable of infecting others. In contrast, the infected state (I) indicates a node that is capable of infecting others. The (R)  
 186 recovered state implies permanent immunity. There are further extensions possible, e.g. adding infants who have maternal  
 187 antibodies (state M) that provide passive immunity. See (24) or (25) for an overview and survey of these models. These  
 188 models have been extensively used in epidemiological studies to characterize disease dynamics as detailed in Table S4, including  
 189 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 (26, 27). 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:

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

190 .  
 191 Eigenvalue  $\lambda_1$  termed the spectral radius characterizes the connectivity of the network graph. Thus, networks that have  
 192 higher connectivity or  $\lambda_1$  are more likely to allow a contagion different paths to grow into an epidemic, whereas in networks  
 193 with low connectivity, the epidemic is more likely to die out.

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

198 Consider a SIR model for illustration, the results also hold for the other models. The model is parametrized by two rates:  $\beta$   
 199 is the probability of an infected node infecting a susceptible node in a given time period, and  $\delta$  is the probability at which an  
 200 infected node recovers (or is cured) during the period. If time is continuous,  $\beta$  and  $\delta$  can be viewed as the rates of infection

201 and recovery. In either case,  $\mathcal{R}_0$  is defined as  $\boxed{\mathcal{R}_0 = \frac{\beta}{\delta}}.$

The epidemic threshold  $\tau$  is defined as follows (26):

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

202 There are a few observations relevant here. First, the critical value of epidemic threshold is a function of the adjacency  
 203 matrix  $E$  of the network topology (structure)  $\mathcal{G}$ . Second, a network topology with a higher epidemic threshold is less likely to  
 204 have an epidemic. Third, interventions like immunizing nodes or reducing the number of connections (edges) can increase the  
 205 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 ((28))
SIR	Measles (29), Swine Flu H1N1 (30), Ebola (31)
SXIR	Chicken Pox (32), SARS (33), COVID-19 (34)
SIRD	COVID-19 ((35))

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

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

212 The process of immunizing (or vaccinating) a set of nodes involves choosing a proportion of nodes (5%, or 10% or 20%)  
 213 and ensuring that these nodes do not transmit any disease. The nodes for immunization are chosen based on three strategies:  
 214 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 $N_k^{\text{infected}}$ . Thus, $k$ can become infected through any of its infected neighbors. So we have: $P_{S \rightarrow I}^k = 1 - (1 - P_{\text{transmit}})^{N_k^{\text{infected}}}$ . Similarly, $P_{S \rightarrow S}^k = (1 - P_{\text{transmit}})^{N_k^{\text{infected}}}$ .
$n_{\text{infected}}^0$	1%	Proportion of nodes in network that are infected at the beginning
$n_{\text{sim}}$	100	Number of simulations

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

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

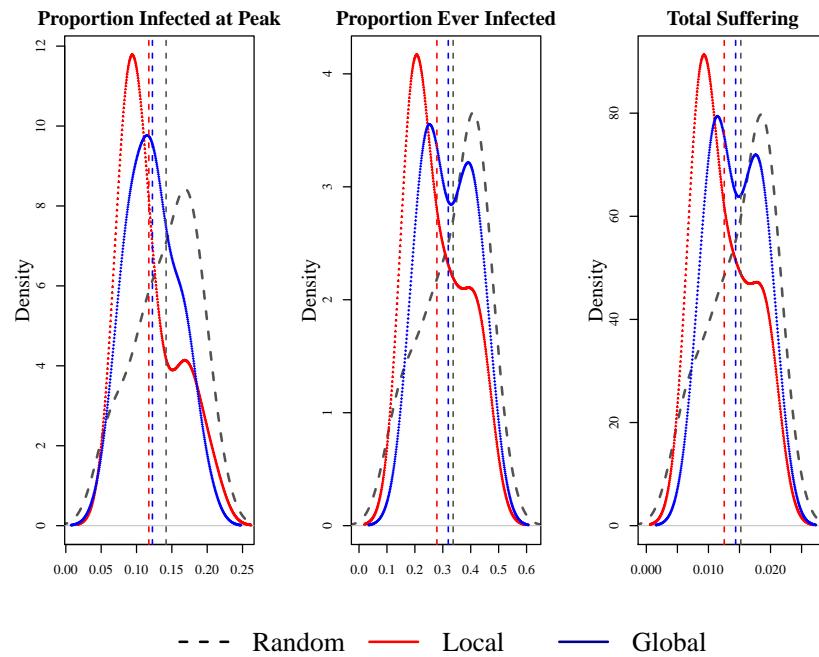
217     **S.I. Epidemic Outcomes**

218     In Figure S7, we examine the epidemic propagation characteristics on the Facebook network (36) using the same parameters  
 219     as detailed in Table S5. The epidemic could either be viewed as an *informational* epidemic propagating through Facebook.  
 220     Alternatively, one might consider the Facebook network structure to serve as an approximation of contact network for the  
 221     purposes of this evaluation.

222     We evaluate epidemics using the following metrics:

- 223     • **Proportion Infected at Peak** =  $\frac{1}{N} \max_t (\sum_i I_{it})$ : Since epidemics increase in intensity and eventually die down, an  
 224       important characteristic is to measure the proportion of the population who are infected at the peak of the epidemic.  
 225       This directly impacts important decisions like hospital capacity planning etc.
- 226     • **Proportion Ever Infected** =  $\frac{1}{N} \sum_i \max_t (I_{it})$ : The proportion of the population that was ever infected by the disease  
 227       is important since it represents the total spread of the disease in the population. It could also represent the number of  
 228       people who might have immunity to future recurrences of the disease.
- 229     • **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  
 230       the length of the infections. This represents the proportion of individual-period combinations with an infection.

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



**Fig. S7.** 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.

## References

- 235 1. V Kumar, D Krackhardt, S Feld, Why your friends are more popular than you are - a revisit in *Archives of the 33rd Sunbelt Conference of International Network for Social Network Analysis*. (INSNA), (2013).
- 236 2. S Strogatz, Friends you can count on. *The New York Times* **September 17** (2012).
- 237 3. MO Jackson, The friendship paradox and systematic biases in perceptions and social norms. *J. Polit. Econ.* **127**, 777–818 (2019).
- 238 4. J Kunegis, Konec: the Koblenz network collection in *Proceedings of the 22nd International Conference on World Wide Web*. (ACM), pp. 1343–1350 (2013).
- 239 5. A Banerjee, AG Chandrasekhar, E Duflo, MO Jackson, The diffusion of microfinance. *Science* **341**, 1236498 (2013).
- 240 6. AL Barabási, et al., Evolution of the social network of scientific collaborations. *Phys. A: Stat. mechanics its applications* **311**, 590–614 (2002).
- 241 7. ME Newman, Coauthorship networks and patterns of scientific collaboration. *Proc. national academy sciences* **101**, 5200–5205 (2004).
- 242 8. S Goyal, MJ Van Der Leij, JL Moraga-González, Economics: An emerging small world. *J. political economy* **114**, 403–412 (2006).
- 243 9. ME O'Kelly, HJ Miller, The hub network design problem: a review and synthesis. *J. Transp. Geogr.* **2**, 31–40 (1994).
- 244 10. AL Barabási, R Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- 245 11. AL Barabási, , et al., Scale-free networks: a decade and beyond. *science* **325**, 412 (2009).
- 246 12. MO Jackson, , et al., *Social and economic networks*. (Princeton university press Princeton) Vol. 3, (2008).
- 247 13. V Bala, S Goyal, A noncooperative model of network formation. *Econometrica* **68**, 1181–1229 (2000).
- 248 14. JK Goeree, A Riedl, A Ule, In search of stars: Network formation among heterogeneous agents. *Games Econ. Behav.* **67**, 445–466 (2009).
- 249 15. F Radicchi, C Castellano, F Cecconi, V Loreto, D Parisi, Defining and identifying communities in networks. *Proc. national academy sciences* **101**, 2658–2663 (2004).
- 250 16. M McPherson, L Smith-Lovin, JM Cook, Birds of a feather: Homophily in social networks. *Annu. review sociology* **27**, 415–444 (2001).
- 251 17. P Dandekar, A Goel, DT Lee, Biased assimilation, homophily, and the dynamics of polarization. *Proc. Natl. Acad. Sci.* **110**, 5791–5796 (2013).
- 252 18. SL Feld, The focused organization of social ties. *Am. journal sociology* **86**, 1015–1035 (1981).
- 253 19. DA McFarland, J Moody, D Diehl, JA Smith, RJ Thomas, Network ecology and adolescent social structure. *Am. sociological review* **79**, 1088–1121 (2014).
- 254 20. ME Newman, Assortative mixing in networks. *Phys. review letters* **89**, 208701 (2002).
- 255 21. ME Newman, J Park, Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122 (2003).
- 256 22. I Sendiña-Nadal, MM Danziger, Z Wang, S Havlin, S Boccaletti, Assortativity and leadership emerge from anti-preferential attachment in heterogeneous networks. *Sci. reports* **6**, 21297 (2016).
- 257 23. A M'Kendrick, Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* **44**, 98–130 (1925).
- 258 24. F Brauer, The Kermack–McKendrick epidemic model revisited. *Math. biosciences* **198**, 119–131 (2005).
- 259 25. HW Hethcote, The mathematics of infectious diseases. *SIAM review* **42**, 599–653 (2000).
- 260 26. D Chakrabarti, Y Wang, C Wang, J Leskovec, C Faloutsos, Epidemic thresholds in real networks. *ACM Transactions on Inf. Syst. Secur. (TISSEC)* **10**, 1–26 (2008).
- 261 27. BA Prakash, D Chakrabarti, M Faloutsos, N Valler, C Faloutsos, Got the flu (or mumps)? check the eigenvalue! *arXiv preprint arXiv:1004.0060* (2010).
- 262 28. D Smith, J Dushoff, R Snow, S Hay, The entomological inoculation rate and plasmodium falciparum infection in african children. *Nature* **438**, 492–495 (2005).
- 263 29. MJ Ferrari, et al., The dynamics of measles in sub-saharan africa. *Nature* **451**, 679–684 (2008).
- 264 30. M Small, C Tse, Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment. *Phys. A: Stat. Mech. its Appl.* **351**, 499–511 (2005).
- 265 31. T Berge, JS Lubuma, G Moremedi, N Morris, R Kondlera-Shava, A simple mathematical model for Ebola in africa. *J. biological dynamics* **11**, 42–74 (2017).
- 266 32. S Deguen, G Thomas, NP Chau, Estimation of the contact rate in a seasonal SEIR model: application to chickenpox incidence in france. *Stat. medicine* **19**, 1207–1216 (2000).
- 267 33. S Riley, et al., Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**, 1961–1966 (2003).
- 268 34. K Prem, 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 Heal.* (2020).
- 269 35. D Caccavo, Chinese and Italian covid-19 outbreaks can be correctly described by a modified sird model. *medRxiv* (2020).
- 270 36. B Viswanath, A Mislove, M Cha, KP Gummadi, On the evolution of user interaction in Facebook in *Proc. Workshop on Online Social Networks*. pp. 37–42 (2009).