

² Supporting Information for

³ **On the Friendship Paradox and Inversity: A Network Property with Applications to
⁴ Privacy-sensitive Network Interventions**

⁵ Vineet Kumar, David Krackhardt and Scott Feld

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

⁸ This PDF file includes:

⁹ Figs. S1 to S8
¹⁰ Tables S1 to S5
¹¹ SI References

12 S.A. Ego-based and Alter-based Means in Example Networks

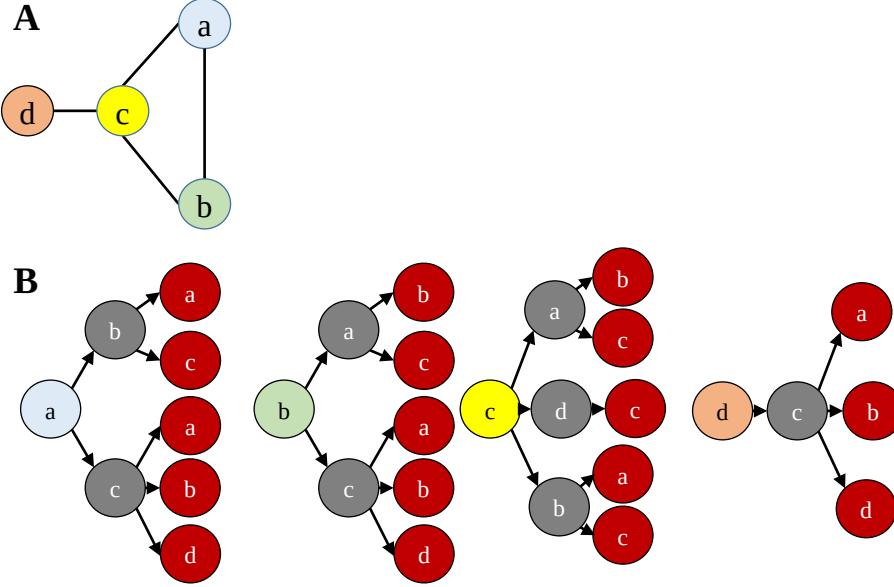


Fig. S1. Ego-based and Alter-based Means in Example Network. **(A) Network.** Example network with 4 nodes a, b, c and d . **(B) Illustration of Neighbors and Neighbors of Neighbors.** Each node is mapped out with its neighbors and neighbors of neighbors. The node is in light blue, neighbors are in gray, and neighbors of neighbors are in red. Node a has 2 neighbors, b and c . Node a also has 5 neighbors of neighbors.

Ego-based Mean: The ego-based mean number of neighbors of neighbors, detailed in (1), addresses the question, How many neighbors of neighbors does each node “experience” in the network? . Specifically, the first node a has two neighbors, and those two neighbors have $(2 + 3) = 5$ neighbors. Thus, on the average, ego a observes that her neighbors have $\frac{5}{2} = 2.5$ neighbors, which is more than she has herself. Similarly, ego b also has two neighbors, who in turn have $(2 + 3) = 5$ neighbors, for an average of $\frac{5}{2} = 2.5$ neighbors per neighbor. c on the other hand has three neighbors, and each neighbor has on the average only $\frac{2+1+2}{3} = \frac{5}{3}$ neighbors; thus, c 's experience is that he has more neighbors than his neighbors do. Finally, ego d only has one neighbor, and that neighbor has 3 neighbors; thus, d 's experience is that he has far fewer neighbors than his neighbors have. Overall, on the average, the four egos experience the following (ego-based) average number of neighbors that their neighbors have:

$$\mu_E = \frac{1}{\text{Number of nodes}} \sum_{i \in \text{Nodes}} (\text{Node } i\text{'s average number of neighbors of neighbors}) = \frac{1}{4} \left(\frac{5}{2} + \frac{5}{2} + \frac{5}{3} + \frac{3}{1} \right) = 2.42$$

Alter-based Mean: The alter-based mean, detailed in (2), addresses a similar idea but focuses on the average number of neighbors each alter has, where alters are the immediate neighbors of egos. For our purposes here, an alter is a special kind of neighbor; it is simply defined as an immediate neighbor of ego. In a network of N nodes, there are precisely N egos; but each node will frequently count more than once in their role as an alter, resulting in many more alters than egos in the network. That is, each node exists once in the network, but they play the role of alter perhaps several times (once for each of the nodes they are connected to). In the present example containing four egos, there are eight alters (colored gray in Panel B in the example above), and each alter has a certain number of adjacent neighbors. Counting how many neighbors each alter has is straightforward: The first alter above (b , who is an alter of a) has two neighbors (a and c); the second alter above (c , who is also an alter of a) has three neighbors (a , b , and d); and so on. Thus, the total number of neighbors of these alters (i.e., all the nodes in red in Panel B) is $(2 + 3 + 2 + 3 + 2 + 1 + 2 + 3) = 18$. The alter-based mean is then:

$$\mu_A = \frac{\text{Total number of neighbors of alters}}{\text{Total number of alters}} = \frac{18}{8} = 2.25$$

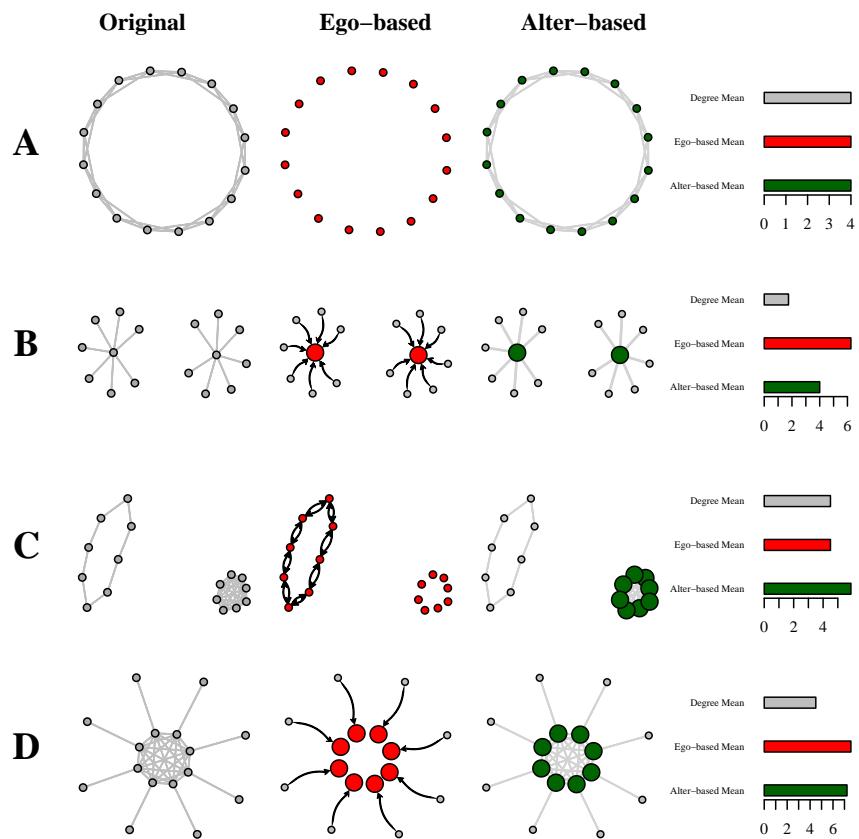


Fig. S2. Four Illustrative Networks with Varying Ego-based and Alter-based Means.

Each network in (A)-(D) has the original network plot (left), ego-based weighted network (middle), and alter-based weighted network (right). On the right is a barplot indicating the mean degree, ego-based mean, and alter-based mean for each of the networks. *Ego-based Panel (Red)*: In the weighted network plot (middle), nodes are sized proportional to their weight (w_i^E) in contributing to the ego-based mean. Edges that receive a *higher than median* weight in computing the ego-based mean are in black. Otherwise, the edges are not plotted in the middle panel. Note that although the original networks are undirected, the selected edges are *directed*. *Alter-based Panel (Green)*: Nodes are sized proportional to their weight (w_i^A) in contributing to the alter-based mean. Edges are all weighted equally in the alter-based weighted network. (A) *Small World Ring*: Each node has four friends, and ego-based and alter-based mean are both equal to the average degree (4). None of the edges are shown in the middle panel since all edges have identical weight in computing the ego-based mean. All nodes in both ego-based and alter-based 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, ego-based mean is substantially higher than the alter-based mean, and is higher than the mean degree across all networks (a)-(d). In the ego-based 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 alter-based panel, the node weights are proportional to degree. (C) *Heavy Core with Detached Cycle*: The alter-based mean is substantially higher than the ego-based mean (and mean degree). Here, we see in the ego-based panel that the weight of each of the nodes has not changed, and all nodes have the same weight. However, in the alter-based panel, we see that the high degree nodes in the complete graph have higher weights 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 ego-based and alter-based mean are substantially higher than the mean degree. In the ego-based 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, and (i, j) indicates an edge (connection) with $e_{ij} \in \{0, 1\}$. We also define the directed edge
 16 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 use neighbor
 17 and friend interchangeably to try and connect with the literature. We note that neighbor is the more general terminology, and
 18 appropriate in this article, since the network phenomena studied are not limited to social networks. We detail the table of
 19 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
$\mathcal{N}(i)$	Neighbors	Set of friends (neighbors) of i , $\mathcal{N}(i) = \{k \in V : (i, k) \in E\}$
D_i	Degree	Number of friends (neighbors) of i , $D_i = \{k \in V : (i, k) \in E\} $
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_E	Ego-based Mean	$\frac{1}{N} \sum_i F_i$
μ_A	Alter-based Mean	$\frac{\sum_i D_i F_i}{\sum_i D_i}$
ρ	Inversity	$\text{Corr}\left(D_i, \frac{1}{D_j}\right) \forall (i, j) \in \hat{E}$

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

23 **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:*

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

27 *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
 28 nodes in a connected network.

29 To prove part (i) of the theorem, first consider each of the components in turn, with $k = 1$. First, in the trivial case
 30 of a degree-regular component, part (i) trivially holds. Next, consider the case with degree variation within component k .
 31 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
 32 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
 33 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
 34 $\frac{1}{|\mathcal{N}(i)|} \left(\sum_{j \in \mathcal{N}_i} D_j \right) > D_i \forall i \in \mathcal{C}_k$ cannot hold. Thus, for each component k , there is at least one node for which the friendship
 35 paradox statement does not hold. In the overall network \mathcal{G} , there must be at least C nodes for which the friendship paradox
 36 statement cannot hold.

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

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

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

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

45
$$\mu_E = \mu_D + \frac{1}{2|V|} \sum_{(i,j) \in V \times V} e_{ij} \left[\frac{(D_i - D_j)^2}{D_i D_j} \right] \quad [1]$$

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

47 *Proof.* Let D_i denote the degree of i . Define $F_i = \frac{1}{D_i} \sum_{j \in N(i)} D_j$ as the mean number of neighbors for neighbors of i . The
48 ego-based mean is defined as:

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

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

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

56 Note that what we characterize as the ego-based mean defined above was independently shown to be greater than the mean
57 degree, including by the present authors at (3). It has also been documented by others, including C. Borgs & J. Chayes in a
58 later comment to an article by (4), and by (5). However, the properties of the ego-based mean have not been formally examined
59 and characterized.

60 For the results below, we consider networks without isolates. □

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 ego-based and alter-based means
62 are connected by the following relationship involving the inversity ρ and the -1, 1, 2, and 3rd moments of the degree distribution:

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

64 We define the moments of the degree distribution as: $\kappa_m = \frac{1}{N} \sum_i D_i^m$. We defined inversity ρ as the correlation
65 of two distributions that we specify as the origin degree (**O**) and inverse destination degree (**ID**) distributions. The **O**
66 distribution consists of the degree of nodes corresponding to edges, and the **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 $\mu_O = \frac{1}{2|E|} \sum_i D_i^2 = \frac{\mu_D^2 + \sigma_D^2}{\mu_D} = \mu_A = \frac{\kappa_2}{\kappa_1}$. 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, we consider the
72 variances. The variance of the origin distribution (**O**) is computed as:

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

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

$$\begin{aligned} 77 \quad \sigma_{ID}^2 &= \frac{1}{2|E|} \sum_{(i,j) \in E} \left[\left(\frac{1}{D_j} - \frac{1}{\mu_D} \right)^2 \right] = \frac{1}{2|E|} \sum_{(i,j) \in E} \left(\frac{1}{D_j^2} + \frac{1}{\mu_D^2} - \frac{2}{\mu_D D_j} \right) \\ 78 \quad &= \frac{1}{2|E|} \left[\sum_{(i,j) \in E} \frac{1}{D_j^2} + \frac{1}{\mu_D^2} \left(\sum_{(i,j) \in E} 1 \right) - \frac{2}{\mu_D} \sum_{(i,j) \in E} \frac{1}{D_j} \right] = \frac{1}{2|E|} \left[\sum_{j \in V} \frac{1}{D_j} + \frac{1}{\mu_D^2} 2|E| - \frac{2}{\mu_D} N \right] \\ 79 \quad &= \frac{1}{\mu_D N} \left[\sum_{j \in V} \frac{1}{D_j} \right] - \frac{1}{\mu_D^2} = (\kappa_1)^{-1} [\kappa_{-1} - (\kappa_1)^{-1}] \\ 80 \quad & \end{aligned}$$

81 We next turn to the inveristy, and based on the definition, we connect it to the ego-based and alter-based means and the degree
 82 distribution.

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

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

$$89 \quad \mu_E = \mu_A + \rho \sqrt{\left(\frac{\kappa_1\kappa_3 - \kappa_2^2}{\kappa_1} \right) [\kappa_{-1} - (\kappa_1)^{-1}]} \quad [2]$$

90 \square

91 **Theorem S5.** *The expected degree of nodes chosen by alter-based strategy is the alter-based mean.*

Proof. To determine the expected degree of a node chosen by the alter-based 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 neighbors 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:

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

92 \square

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

94 **Theorem S6.** *[Rewiring Theorem] Let network $\mathcal{G} = (V, E)$ with $N > 3$ nodes include nodes a, b, c, d with degrees ordered
 95 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
 96 the network to $\mathcal{G}' = (V, E')$, containing edges $(a, d), (b, c) \in E'$, but $(a, b), (c, d) \notin E'$, we obtain: $\mu_E(\mathcal{G}') > \mu_E(\mathcal{G})$ whereas
 97 $\mu_A(\mathcal{G}') = \mu_A(\mathcal{G})$. Also, it follows that $\rho(\mathcal{G}') > \rho(\mathcal{G})$.*

98 *Proof.* First, observe that the degree distribution is unaffected by the change, since each node's degree is unchanged by the
 99 rewiring. Therefore, the alter-based mean (which only depends on mean and variance of the degree distribution) is also
 100 unaffected, i.e., $\mu_A(\mathcal{G}) = \mu_A(\mathcal{G}')$. Recall that the ego-based mean is $\mu_E = \frac{1}{N} \sum_i \sum_j e_{ij} \left[\frac{D_i}{D_j} + \frac{D_j}{D_i} \right]$. Since between \mathcal{G} and \mathcal{G}'
 101 the degrees of all nodes are the same, and all edges are the same except the two rewired edges, we can write the difference
 102 between their ego-based means as:

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

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

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

110 We use a wide variety of real networks to characterize their properties, and illustrate how relate to the interventions detailed in
 111 the paper. We use data from two repositories.

112 **A. Koblenz Network Collection.** The networks are selected across several categories (Affiliation, Face-to-face Social, Online
 113 Social, Computer, Infrastructure and Biological networks), and span a wide range in network characteristics like size and
 114 density (Table S2). These networks also vary widely in terms of their size, from a low of 25 to networks with millions of nodes
 115 (e.g., Youtube). All network data was obtained from the Koblenz Network Collection (KONECT) (6). We examine these real
 116 networks on a number of dimensions, including the number of nodes, edges, and the variation in the degree distribution.

Table S2. Real Network Characteristics

Label	Network	N	$ E $	μ_D	μ_E	μ_A	ρ
<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

117 A few observations are worth noting here. First, there are many networks with both positive and negative values of inversty,
 118 both within and across categories. Second, we do not see Inversty ρ close to ± 1 . However, the Twitter network is closest in
 119 magnitude, with an inversty of $\rho = 0.65$. Third, the variation in inversty is low in some categories like Infrastructure, whereas
 120 it is relatively greater in Online Social networks. Finally, we see that even low values of inversty can impact the difference
 121 between the ego-based and alter-based means substantially, as well as between each of these and the mean degree. The WWW
 122 (Google) network, for example, displays such meaningful differences, even with a low inversty value of 0.06. This is due to the
 123 multiplier effect of the moments of the distribution function, detailed in equation (3).

124 **B. India Village Networks.** In addition, we also use data from $N = 75$ villages in India made publicly available (see (7) for
 125 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

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. S3, Fig. S4). 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.

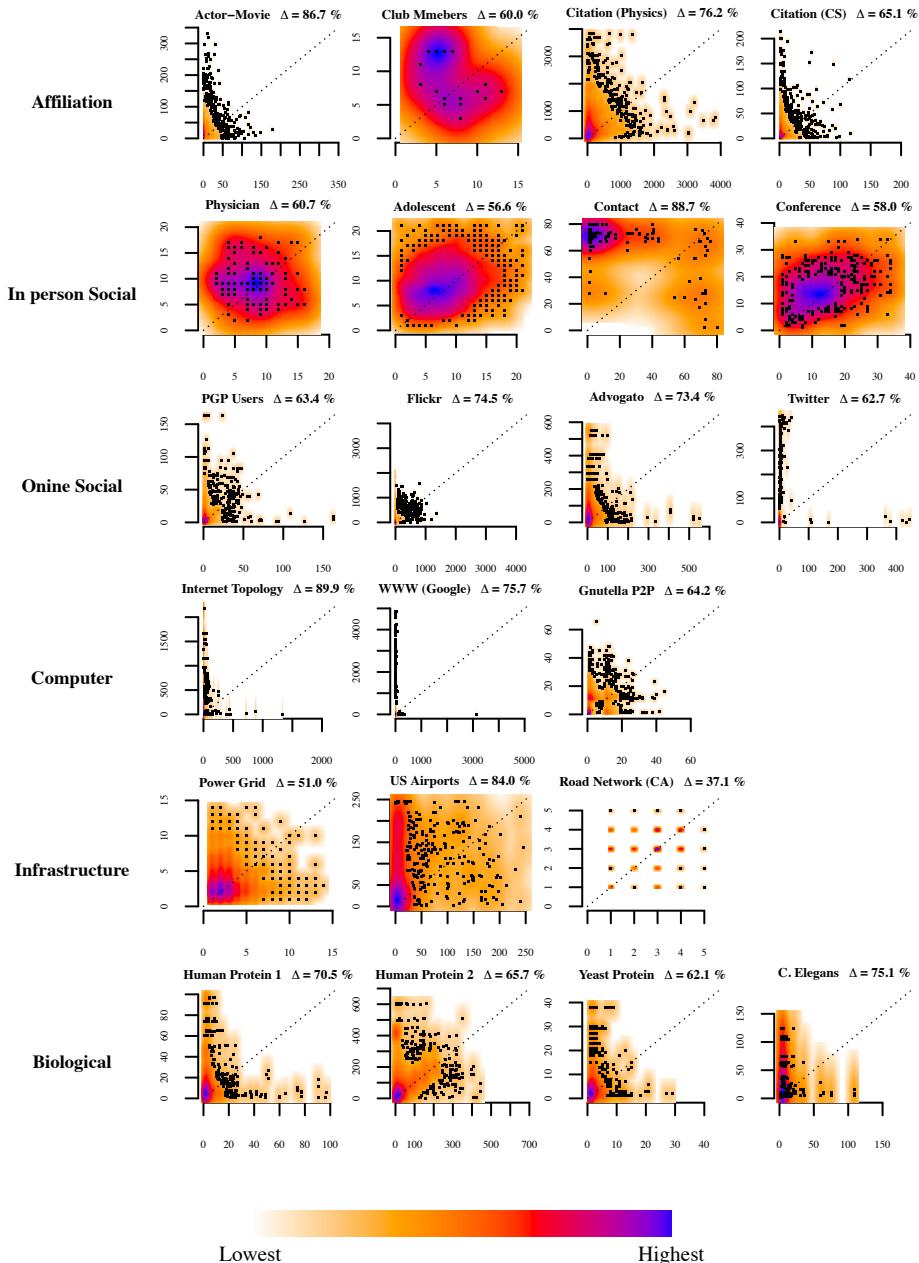


Fig. S3. Friendship Paradox at Individual Level. Density plot of average number of friends of nodes compared to node degree in networks. Δ indicates the proportion of nodes that have a higher average number of friends of friends than their degree. Lowest density regions within each network are marked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

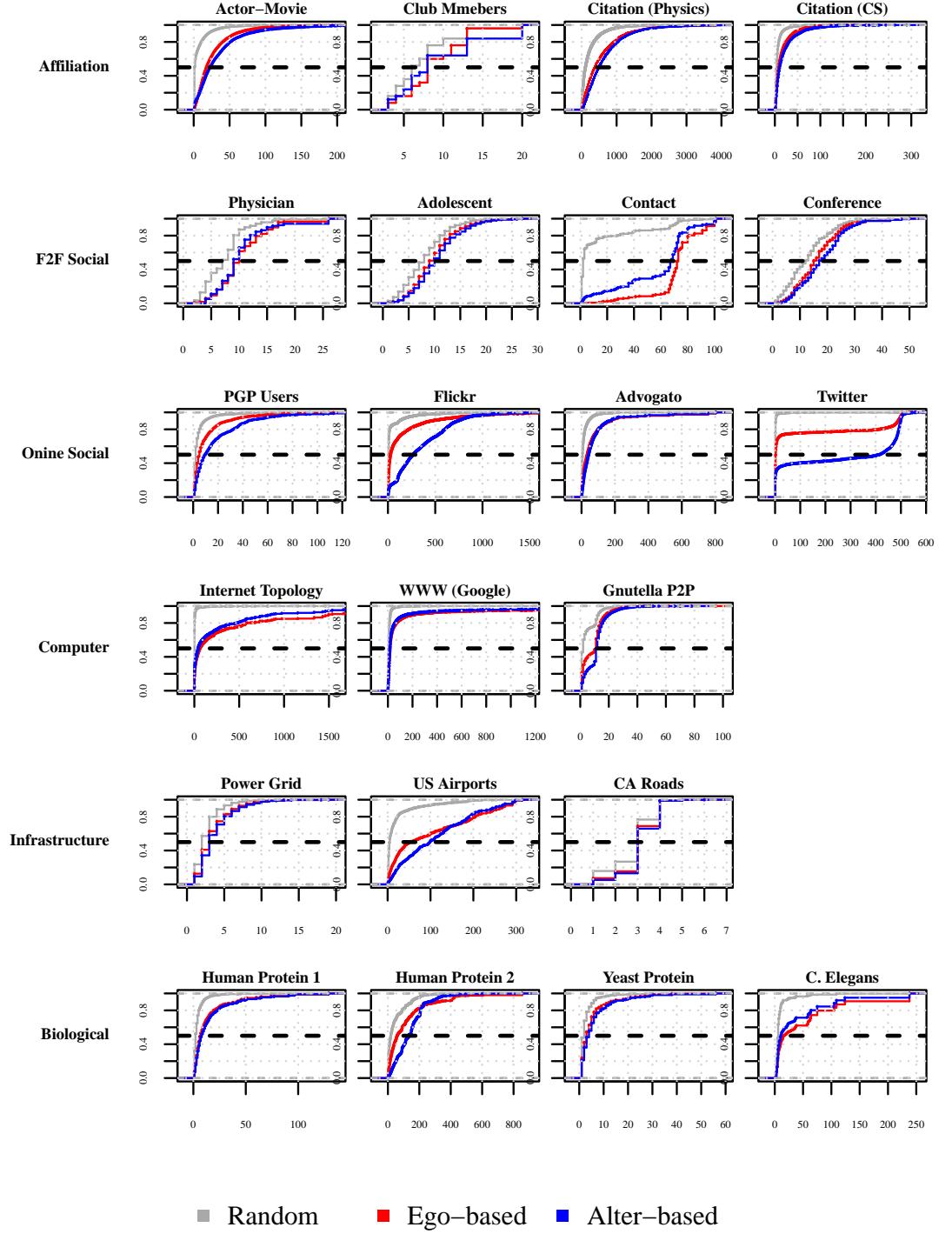


Fig. S4. Individual Friendship Paradox. Empirical Cumulative Distribution Functions (CDF) of Real Networks. Panels show the CDF of 3 different network properties at the individual node level. For a specific node degree, the probability that a node with a lower (or identical) degree is chosen by the sampling strategy for node degree (gray), ego-based mean of Friends of Friends (in red), and alter-based mean of friends of friends (in blue). Across all networks, for lower degrees, the node degree curve is to the left of the ego-based and alter-based mean of friends curves. In several networks, alter-based mean of friends is to the left and higher than ego-based mean of friends (e.g., Contact), whereas in others, it is to the right (e.g., Flickr).

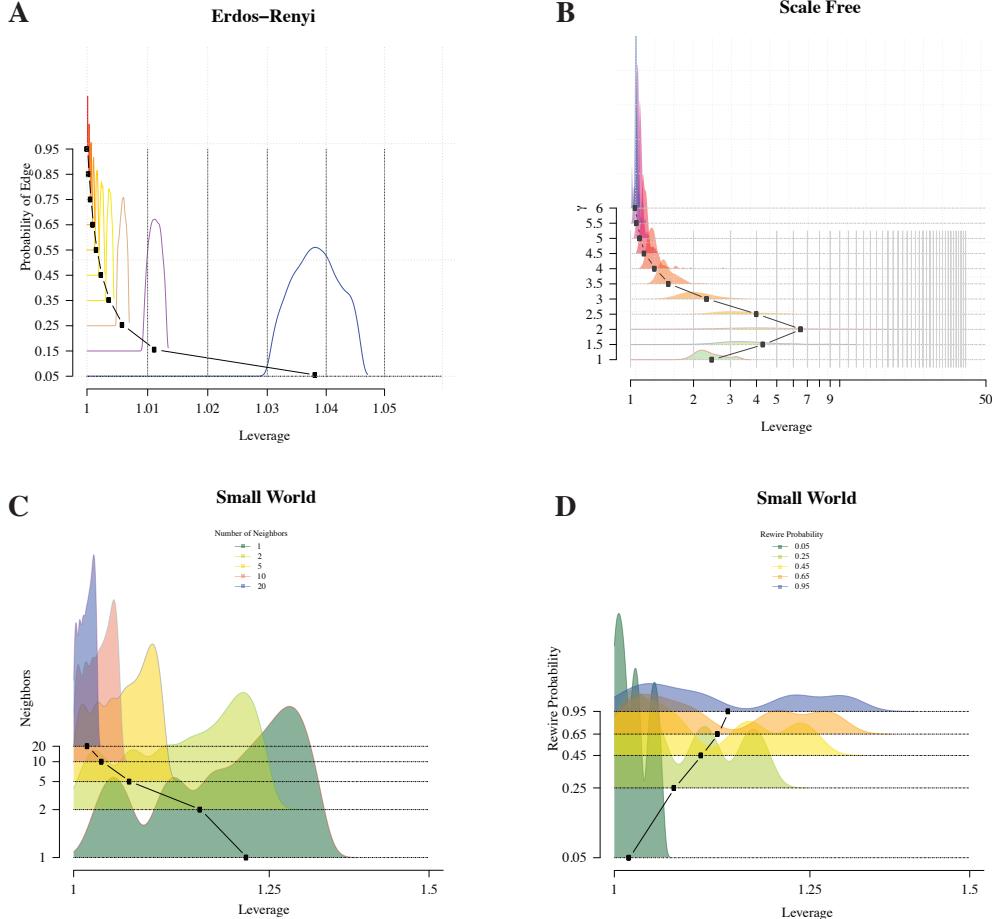


Fig. S5. Ego-based Leverage Density in Generated Networks from three different generative models, spanning the parameter space of each. A sample of 1,000 networks was used for each of the models. (A) Erdos-Renyi (ER) networks generated with edge probabilities, $p \in [0.05, 0.95]$, and size ranging from $N=50$ to $N=1,000$ nodes. We find that ego-based leverage is highest for the lowest edge probabilities, and leverage converges to 1 as the networks become more dense. (B) (SF) Static Scale Free networks with scale-free parameter $\gamma \in [1, 6]$ (8). 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 ego-based mean converges to the mean degree. With (SW) small world networks, we have two parameters (9). 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 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 higher rewiring probabilities, the distributions feature increased variance as well as higher mean leverage.

133 **S.F. Network Features that Impact Inversity**

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

137 **Star or hub-based networks** First, we observe that there is significant evidence for hub-based network structures appearing
138 in real-world networks. Such hub-like structures are common across a wide range of networks, including co-author networks,
139 the underlying interlink network that forms the Internet, as well as Airline networks (10–13). The early networks literature
140 explaining the emergence of such hub-like patterns posited preferential attachment as a mechanism, where newly joining nodes
141 connected disproportionately to highly connected nodes (8, 14). The economics literature involving the economic incentives
142 underlying network formation posits that agents form links based on the expected benefits to such formation. The resulting
143 network is an equilibrium outcome based on the decisions of each of the agents, who are maximizing their own utilities (15). In
144 this stream, an influential paper (16) finds that even when agents are homogeneous — where they have identical constraints,
145 preferences, and incentives — *star networks arise across a wide range of equilibria*. Stars are predicted to occur even though
146 agents are symmetric with identical incentives and opportunities. Complementing this research, star networks are found to
147 arise over time in experimental settings where agents vary in terms of costs, incentives, and even information (17).

148 **Clusters and Communities in networks** Clusters or communities as well as cliques (fully connected or complete subnetworks) are
149 commonly observed in networks. A typical conceptualization of community is the following: “Qualitatively, a community is
150 defined as a subset of nodes within the graph such that connections between the nodes are denser than connections with the
151 rest of the network.” (18). There are several reasons why communities form, including homophily and social foci. In homophily,
152 when a number of individuals are similar, then they are much more likely to be connected to each other, and also part of the
153 same larger grouping or community (19). However, it should be noted that not all such connections will happen; rather, such
154 connections and communities are more likely to happen when individuals are homophilous. We note that homophily has also
155 been tied to polarization and segregation (20).

156 A prominent theory that explains how communities form is the idea of foci (21). The essential idea is that most ties originate
157 around foci of activity, where a limited set of people share a focus that organizes activity, and thereby tend to generate repeated
158 interaction among the same people in the set over time that leads to ties among many of them. Each person tends to be
159 associated with many different foci. Alters from the same focus tend to be tied to one another, but not those from separate
160 foci. Consistent with this notion, research has found that the way organizational environments are structured moderates the
161 tie-formation process (22).

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

165 **S.G. Inversity and Assortativity: Connections and Differences**

166 A natural question is whether inversity, $\rho = \text{Corr} \left(D^O, \frac{1}{D^D} \right)$, captures the same information (with opposite sign) as degree
 167 assortativity, which is a well known network property, $\rho_a = \text{Corr} \left(D^O, D^D \right)$ (23–25). Inversity and assortativity are negatively
 168 correlated, as we might expect.

169 **Can Assortativity Be Used as a Proxy for Inversity?**. There are several specific reasons why we don't think it is a good idea to
 170 use assortativity as a *proxy metric* in place of inversity. We demonstrate specifically how using assortativity as a proxy for
 171 inversity would lead to incorrect choices to the network intervention questions below.

172 **Network Intervention Questions:** We begin with our objective for the network intervention strategies, which is to choose a
 173 strategy that maximizes the expected degree of target nodes. We have two questions that need to be answered.

- 174 (a) Identify whether the ego-based strategy leads to higher expected degree than the alter-based strategy (or vice versa), i.e.
 175 whether $\mu_E > \mu_A$ or $\mu_A > \mu_E$.
- 176 (b) Evaluate the improvement in expected degree offered between any of random, ego-based and alter-based strategies, i.e.
 177 $(\mu_E - \mu_D)$, $(\mu_A - \mu_D)$ and $(\mu_E - \mu_A)$.

178 The requirements for problem (a) are different from those of (b). For (a), we need to just know the correct ordering of
 179 ego-based and alter-based strategies, and not the magnitude. Inversity (ρ) gives us a direct answer to question (a), since
 180 $\mu_E > \mu_A \iff \rho > 0$. For (b), the ordering is not sufficient, and we need to know the magnitude, and we explore this below.

181 **Why Magnitudes of Differences Matter:** The decision maker could evaluate the benefit of using a friendship paradox strategy is
 182 worth the potential cost relative to the random strategy, which requires the least amount of information and effort from the
 183 people originally selected. Thus, the decision maker would trade off the increase in expected degree relative to the marginal
 184 cost of using the strategy (26). This logic implies that knowing just the ordering of the different strategies is not sufficient, and
 185 we would need to know the magnitudes of the differences in order to select a strategy.

186 Inversity allows for a direct linear transformation from alter-based mean (μ_A) to ego-based mean (μ_E) calculation of
 187 neighbors of neighbors. We can see this using the formulation from equation (2) in page 5 of this Supplement:

$$188 \mu_E = \mu_A + \rho \underbrace{\left(\sqrt{\left(\frac{\kappa_1 \kappa_3 - \kappa_2^2}{\kappa_1} \right) [\kappa_{-1} - (\kappa_1)^{-1}]} \right)}_{\omega} = \mu_A + \rho \omega$$

189 where the κ s represent the moments of the degree distribution. This formula implies that $\mu_E = \mu_A + \omega \rho \implies \mu_E - \mu_A = \omega \rho$,
 190 where ω is a function of the moments of the degree distribution.

191 We have found no comparable transformation between alter-based and ego-based means using assortativity, and argue
 192 that such a relationship is unlikely to exist (see point below about monotonic ordering between these metrics). Thus, to use
 193 assortativity in place of inversity, we must first assume $\rho_a \approx -\rho$ in order to make an approximation of the form: $\mu_E - \mu_A \approx -\omega \rho_a$.
 194 We don't know whether and when this approximation would be valid.*

195 If we do make this approximation, high values of the quantity ω can amplify small differences in the values of ρ . Therefore,
 196 even *small errors* in approximating inversity with assortativity would be highly problematic in evaluating magnitude of effect
 197 sizes. We also see that this impact is practically important. In Table S2 of Supplement §S.C, for instance, we observe that even
 198 small values of inversity can result in large differences between the different means. Please see the statistics corresponding to
 199 the A3 (Citation) and C2 (WWW) networks.

*We further note that without Theorem S5, it would not be possible to quantify the differences between ego-based and alter-based means in terms of the degree distribution and inversity.

200 **Decisions using Assortativity versus Inversity:** Inversity is sensitive to and positively affected by stars and star-like structures.
201 Assortativity, on the other hand, is more sensitive to and positively affected by cliques and clique-like structures. Supplement
202 §S.F shows that cliques and stars are commonly present in real world networks.

203 *Illustrative Example Networks by Simulation:* We explore this logic quantitatively by simulating specific type of networks
204 with stars and cliques to make things concrete. We examine how inversity and assortativity change as the size of the star
205 network is varied, keeping the cliques constant. To make this deterministic, we have fixed the cliques to include all cliques from
206 size 2 to size 20. We add only one star network and vary the size of the star network, with the number of nodes in the star
207 varying from 1 to 61.

208 We plot in Figure S6 both the assortativity and inversity levels for each network as a function of how large the star is in the
209 network. All other parameters remain fixed. The blue circles indicate the inversity (ρ) for the generated network; the red
210 plus-signs indicate the assortativity (ρ_a) for the same generated network. The plot also includes inset sociograms (A)-(D) at
211 specific points of interest.

212 First, we note that inversity and assortativity are highly correlated, with $\text{cor}(\rho, \rho_a) = -0.992$. Similarly, we test a linear
213 regression model of ρ and ρ_a and find the R^2 of the model to be $R^2 = 0.983$. Thus, the measures seem on the surface to be
214 very closely related. However, we observe if we use assortativity as a proxy for inversity, we would not make the right decisions
215 for questions (a) or (b). Using assortativity as proxy, we would approximate $\rho \approx -\rho_a$.

216 We now detail the problems with using assortativity:

217 (1) In the shaded region, where the size of the star networks is between 29 and 47 (approximately), we observe that the sign
218 of assortativity and inversity is the same, i.e., *both are positive*. In such a case, the answers to both questions (a) and (b)
219 would be incorrect.

220 (2) To the left of the shaded region, where the size of the star network varies from 2 to 29 (approximately), we note that
221 $|\rho_a| > |\rho|$, implying that even though the sign of ρ_a and ρ are opposite, the magnitudes are quite different. More
222 specifically, assuming $\rho \approx -\rho_a$, we would overestimate the benefit of the ego-based strategy, leading to errors in answering
223 question (b).

224 (3) To the right of the shaded region, i.e., when star networks have size greater than 47, we observe that $|\rho| \gg |\rho_a|$. This
225 difference in magnitudes implies that using the proxy assumption $\rho \approx -\rho_a$ would be highly problematic for obtaining
226 magnitudes required for (b). More specifically, we would substantially underestimate the benefits of the ego-based strategy
227 relative to alter-based and random.

228 Broadly, in these networks, assortativity cannot reliably help us determine which mean (ego-based or alter-based) is larger,
229 or which strategy dominates. We also don't know the conditions under which it could serve (or not) as a reasonable proxy
230 for inversity for these or more general networks. Thus, even when these metrics are highly correlated, using the proxy of
231 assortativity in place of inversity could lead to errors over a substantial range of networks.

232 **Monotonic Ordering between Assortativity and Inversity?** Since Assortativity (ρ_a) as a proxy is directionally the *reverse* of Inversity
233 (ρ), we would naturally expect that it to vary monotonically with inversity. Specifically, we would expect for any pair of
234 networks \mathcal{G}_1 and \mathcal{G}_2 , when $\rho(\mathcal{G}_1) > \rho(\mathcal{G}_2) \iff \rho_a(\mathcal{G}_1) < \rho_a(\mathcal{G}_2)$. However, there are (many) examples of pairs of networks where
235 both assortativity are directionally the same, i.e. $\rho(\mathcal{G}_1) > \rho(\mathcal{G}_2)$ and $\rho_a(\mathcal{G}_1) > \rho_a(\mathcal{G}_2)$. Using common network generation
236 algorithms, it is easy to obtain many such pairs of networks, where one of the networks has both **higher assortativity and**
237 **higher inversity** than the other. Thus, we find that assortativity as a proxy metric for inversity does not even preserve
238 ordering, making it problematic to rely on it as a proxy for network intervention problems. When we examined inversity and
239 assortativity for a set of 45 village networks from India, we found 3 of the 45 village networks had both positive inversity and
240 assortativity.

241 **Variation of Inversity and Assortativity with Set of Cliques.** [†] We also experimented with varying sets of cliques to understand
242 how and when inversity and assortativity diverge. First, we begin with all cliques of size 2 to 15 included in a network. We
243 then alter the network to selectively remove cliques. The sociograms indicate the cliques that are included in the network.

[†]We thank an anonymous reviewer for suggesting this exercise.

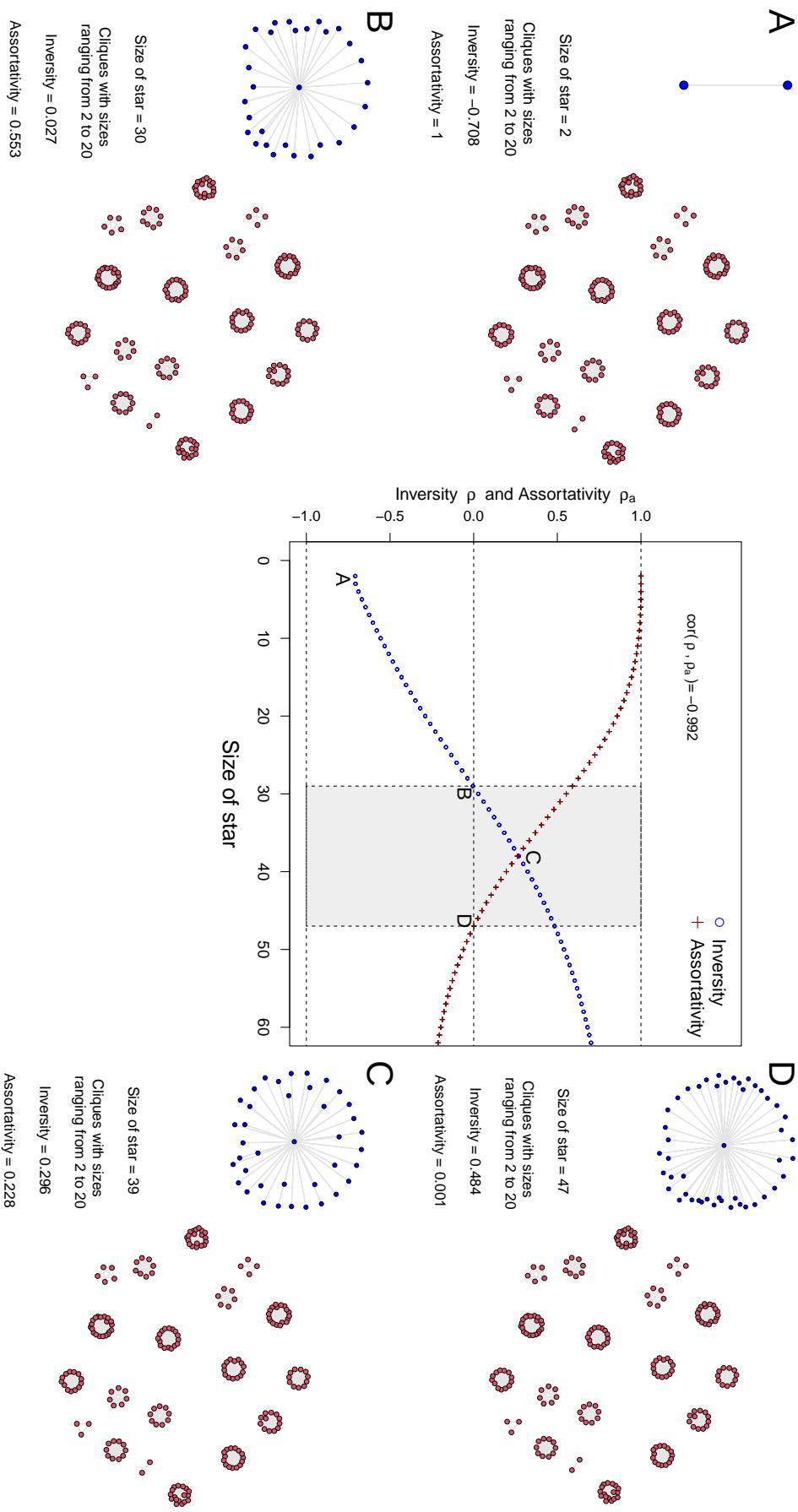
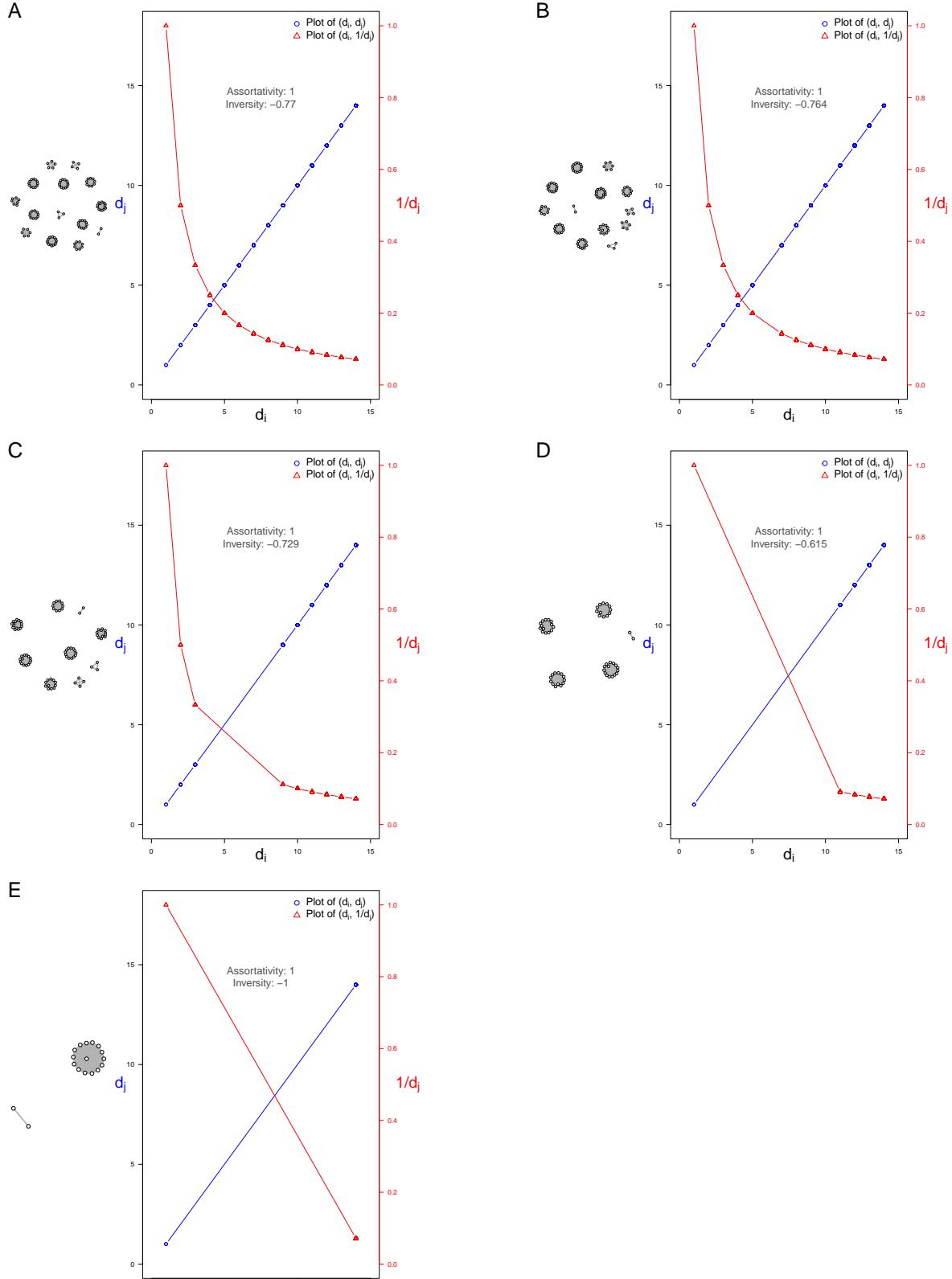


Fig. S6. Variation of Inversity with Star Size

Note: We simulate networks with one star and several cliques. The size of the cliques ranges from 2 to 20 nodes. The size of the star network is varied between 2 and 60 (one hub connected to 60 spokes). The set of cliques is maintained exactly the same across all of the networks. Each network appears in the plot as a point. We observe that inversity is increasing with size of the star network, whereas assortativity is decreasing. The shaded area represents a region among these networks where *both* *inversity* and *assortativity* are *positive*. We illustrate the network topologies corresponding to 4 different networks (A)–(D), where (A) corresponds to the a star of size 2, (B) has a star of size 30, (C) has a star of size 39, and (D) has a star of size 47.

Fig. S7. Networks with varying number of cliques

Each panel corresponds to a network, which here is a set of cliques, ranging in size from 2 to 15. We illustrate the degrees of nodes connected by edges in each panel, with each edge in the network corresponding to a point in the plot. The points (and lines) in blue corresponding to (d_i, d_j) depicting edge (i, j) that connects nodes i and j . This illustration views the degrees the way assortativity would see them. The points (and lines) in red plots $(d_i, 1/d_j)$, corresponding to the way inveristy would view the edge (i, j) .



244 We then plot the degrees of nodes that connect an edge in the x-axis and y-axis. For edge (i, j) , we include (d_i, d_j) in the
 245 plot in blue. For the same edge, we also include $(d_i, 1/d_j)$ in red. Thus, we observe a few different things.

- 246 1. We observe that assortativity for all the sets of cliques we have tested is always 1. However, the inversty is only -1 in the
 247 trivial case where we have 2 cliques.
- 248 2. Inversty and assortativity have a non-linear relationship with the degree as expected. Specifically, when we have low
 249 degree nodes, they contribute disproportionately more to inversty.
- 250 3. The inversty across the set of cliques appears to vary in a non-monotonic manner. Specifically, we have $\rho = -0.77$ in
 251 panel (A), where we have all the cliques between size 2 and 15 included in the network. As we remove cliques, we find
 252 that in panel (B), the inversty $\rho = -0.764$ is closer to zero. This pattern continues in (C) and (D) when we remove even
 253 more cliques, we observe that inversty increases and becomes less negative. However, in the extreme case when we have
 254 only 2 cliques (size 2 and size 15), we observe that inversty $\rho = -1$. Thus, we see that there does not appear to be a
 255 simple pattern for how inversty varies with the number or set of cliques. We think future research might find it useful to
 256 examine the patterns of variation of inversty, and how it contrasts with assortativity.

257 S.H. Virus Propagation Models

258 We detail below several examples of virus propagation models being used for characterizing the transmission and spread of
 259 diseases. These models build upon the early work of Kermack and McKendrick (27). All individuals in a population (in our
 260 case, the nodes in a network) are in one of the states, either susceptible (S) or infected (I). Based on the viral propagation, they
 261 can move to other states like Exposed (X), Recovered (R), or Deceased (D). For example, the SIR model involves individuals
 262 being in one of three states, (S), (I) or (R), and transitioning between the states probabilistically. Typically, the vast majority
 263 of nodes are present in the susceptible state (S), in which they might contrast the disease. The exposed state (X) is used to
 264 indicate a node that has been exposed to the disease, but could be asymptomatic during an incubation period and is not
 265 capable of infecting others. In contrast, the infected state (I) indicates a node that is capable of infecting others. The (R)
 266 recovered state implies permanent immunity. There are further extensions possible, e.g., adding infants who have maternal
 267 antibodies (state M) that provide passive immunity. See (28) or (29) for an overview and survey of these models. These
 268 models have been extensively used in epidemiological studies to characterize disease dynamics as detailed in Table S4, including
 269 measles, influenza, and COVID-19.

270 There has been recent notable work that aims to characterize the epidemic thresholds of these compartmental models with
 271 disease transmission over a network (30, 31). The critical idea is that the epidemic threshold of a network can be characterized
 272 as the inverse of the greatest (first) eigenvalue of the adjacency matrix A of the network, denoted as:

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

273 . Eigenvalue λ_1 , termed the spectral radius, characterizes the connectivity of the network graph. Thus, networks that have
 274 higher connectivity or λ_1 are more likely to allow contagions along different paths to grow into epidemics, whereas in networks
 275 with low connectivity, contagions are more likely to die out.

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

280 Whereas we consider an SIR model for illustration, the results also hold for the other models. The model is parametrized by
 281 two rates: β is the probability of an infected node infecting a susceptible node in a given time period, and δ is the probability
 282 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
 283 infection and recovery. In either case, \mathcal{R}_0 is defined as $\boxed{\mathcal{R}_0 = \frac{\beta}{\delta}}$.

284 The epidemic threshold τ is defined as follows (30):

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

285 There are a few observations relevant here. First, the critical value of epidemic threshold is a function of the adjacency
 286 matrix E of the network topology (structure) \mathcal{G} . Second, a network topology with a higher epidemic threshold is less likely to
 287 have an epidemic. Third, interventions like immunizing nodes or reducing the number of connections (edges) can increase the
 288 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 (32)
SIR	Measles (33), Swine Flu H1N1 (34), Ebola (35)
SXIR	Chicken Pox (36), SARS (37), COVID-19 (38)
SIRD	COVID-19 ((39))

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

286 **Implementation of VPM.** We begin with a seed set of 1% of the nodes being infected, and evaluate epidemic outcomes using the
 287 SIR model. All the nodes in the network that are not infected or recovered are susceptible (S) to the infection. Each infected
 288 node can transmit an infection in each period probabilistically to each of its neighbors. The probability of an infection is
 289 $P_{\text{transmit}} = \beta$. Thus, a node can become infected (I) from contact with any of its neighbors. In each period, an infected node
 290 can be cured or recovered (R) probabilistically, with the likelihood $P_{\text{cure}} = \delta$. Recovered nodes cannot be reinfected and cannot
 291 transmit infections.

292 The process of immunizing (or vaccinating) a set of nodes involves choosing a proportion of nodes (5%, or 10% or 20%)
 293 and ensuring that these nodes do not transmit any disease. The nodes for immunization are chosen based on three strategies:
 294 random, ego-based, and alter-based. 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**)nfectious, (**R**)ecovered / (**R**)emoved

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

297 S.I. Epidemic Outcomes

298 In Figure S8, we examine the epidemic propagation characteristics on the Facebook network (40) using the same parameters as
 299 detailed in Table S5. The epidemic could be viewed as an *informational* epidemic propagating through Facebook. Alternatively,
 300 one might consider the Facebook network structure to serve as an approximation of contact network for the purposes of this
 301 evaluation.

302 We evaluate epidemics using the following metrics:

- 303 • **Proportion Infected at Peak** = $\frac{1}{N} \max_t (\sum_i I_{it})$: Since epidemics increase in intensity and eventually die down, an
 304 important characteristic is to measure the proportion of the population who are infected at the peak of the epidemic.
 305 This directly impacts important decisions like hospital capacity planning, etc.
- 306 • **Proportion Ever Infected** = $\frac{1}{N} \sum_i \max_t (I_{it})$: The proportion of the population that was ever infected by the disease
 307 is important since it represents the total spread of the disease in the population. It could also represent the number of
 308 people who might have immunity to future recurrences of the disease.

- 309 • **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
 310 the length of the infections. This represents the proportion of individual-period combinations with an infection.

311 For the Facebook network sample, we find that an epidemic's outcomes are better when using the ego-based strategy
 312 compared to the alter-based strategy, which in turn is better than the random strategy. This conclusion holds for all the
 313 metrics considered above.

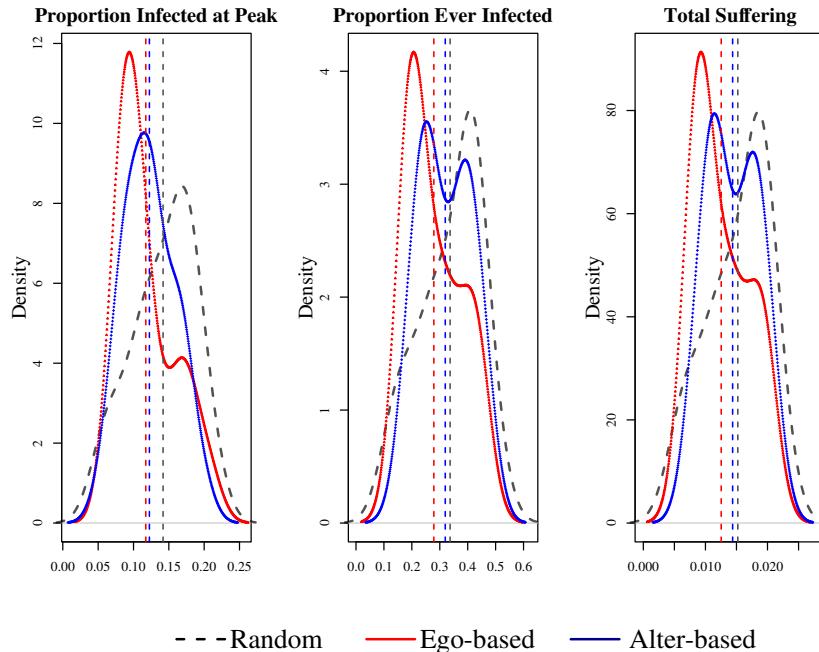


Fig. S8. Epidemic Outcomes with Immunization in Facebook Network. See Table S5 for parameters of simulation. All outcomes are density plots. We plot 3 outcomes: (a) the proportion of population infected at the peak, (b) proportion of population that was ever infected, and (c) total suffering. In each panel, the x -axis represents proportions and the y -axis represents density. We plot the outcomes for 3 strategies: (R)andom, (E)go-based, and (A)lter-based. The dashed vertical lines represent the means for the 3 strategies. We find that for the Facebook network, the ego-based strategy is better for all outcomes than the alter-based strategy, which in turn is better than the random strategy.

314 **References**

- 315 1. D Krackhardt, Structural leverage in marketing. *Networks marketing* pp. 50–59 (1996).
- 316 2. S Feld, Why your friends have more friends than you do. *Am. J. Sociol.* pp. 1464–1477 (1991).
- 317 3. V Kumar, D Krackhardt, S Feld, Why your friends are more popular than you are - a revisit in *Archives of the 33rd*
318 *Sunbelt Conference of International Network for Social Network Analysis.* (INSNA), (2013).
- 319 4. S Strogatz, Friends you can count on. *The New York Times* September 17 (2012).
- 320 5. MO Jackson, The friendship paradox and systematic biases in perceptions and social norms. *J. Polit. Econ.* **127**, 777–818
321 (2019).
- 322 6. J Kunegis, Konec: the Koblenz network collection in *Proceedings of the 22nd International Conference on World Wide*
323 *Web.* (ACM), pp. 1343–1350 (2013).
- 324 7. A Banerjee, AG Chandrasekhar, E Duflo, MO Jackson, The diffusion of microfinance. *Science* **341**, 1236498 (2013).
- 325 8. AL Barabási, R Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- 326 9. D Watts, S Strogatz, Collective dynamics of “small-world” networks. *nature* **393**, 440–442 (1998).
- 327 10. AL Barabási, et al., Evolution of the social network of scientific collaborations. *Phys. A: Stat. mechanics its applications*
328 **311**, 590–614 (2002).
- 329 11. ME Newman, Coauthorship networks and patterns of scientific collaboration. *Proc. national academy sciences* **101**,
330 5200–5205 (2004).
- 331 12. S Goyal, MJ Van Der Leij, JL Moraga-González, Economics: An emerging small world. *J. political economy* **114**, 403–412
332 (2006).
- 333 13. ME O’Kelly, HJ Miller, The hub network design problem: a review and synthesis. *J. Transp. Geogr.* **2**, 31–40 (1994).
- 334 14. AL Barabási, , et al., Scale-free networks: a decade and beyond. *science* **325**, 412 (2009).
- 335 15. MO Jackson, , et al., *Social and economic networks.* (Princeton university press Princeton) Vol. 3, (2008).
- 336 16. V Bala, S Goyal, A noncooperative model of network formation. *Econometrica* **68**, 1181–1229 (2000).
- 337 17. JK Goeree, A Riedl, A Ule, In search of stars: Network formation among heterogeneous agents. *Games Econ. Behav.* **67**,
338 445–466 (2009).
- 339 18. F Radicchi, C Castellano, F Cecconi, V Loreto, D Parisi, Defining and identifying communities in networks. *Proc. national*
340 *academy sciences* **101**, 2658–2663 (2004).
- 341 19. M McPherson, L Smith-Lovin, JM Cook, Birds of a feather: Homophily in social networks. *Annu. review sociology* **27**,
342 415–444 (2001).
- 343 20. P Dandekar, A Goel, DT Lee, Biased assimilation, homophily, and the dynamics of polarization. *Proc. Natl. Acad. Sci.*
344 **110**, 5791–5796 (2013).
- 345 21. SL Feld, The focused organization of social ties. *Am. journal sociology* **86**, 1015–1035 (1981).
- 346 22. DA McFarland, J Moody, D Diehl, JA Smith, RJ Thomas, Network ecology and adolescent social structure. *Am.*
347 *sociological review* **79**, 1088–1121 (2014).
- 348 23. ME Newman, Assortative mixing in networks. *Phys. review letters* **89**, 208701 (2002).
- 349 24. ME Newman, J Park, Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122 (2003).
- 350 25. I Sendiña-Nadal, MM Danziger, Z Wang, S Havlin, S Boccaletti, Assortativity and leadership emerge from anti-preferential
351 attachment in heterogeneous networks. *Sci. reports* **6**, 21297 (2016).
- 352 26. D Krackhardt, Structural leverage in marketing in *Networks in marketing*, ed. D Iacobucci. (Sage, Thousand Oaks), pp.
353 50–59 (1996).
- 354 27. A M’Kendrick, Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* **44**, 98–130 (1925).
- 355 28. F Brauer, The Kermack–McKendrick epidemic model revisited. *Math. biosciences* **198**, 119–131 (2005).
- 356 29. HW Hethcote, The mathematics of infectious diseases. *SIAM review* **42**, 599–653 (2000).
- 357 30. D Chakrabarti, Y Wang, C Wang, J Leskovec, C Faloutsos, Epidemic thresholds in real networks. *ACM Transactions on*
358 *Inf. Syst. Secur. (TISSEC)* **10**, 1–26 (2008).
- 359 31. BA Prakash, D Chakrabarti, M Faloutsos, N Valler, C Faloutsos, Got the flu (or mumps)? check the eigenvalue! *arXiv*
360 *preprint arXiv:1004.0060* (2010).
- 361 32. D Smith, J Dushoff, R Snow, S Hay, The entomological inoculation rate and plasmodium falciparum infection in african
362 children. *Nature* **438**, 492–495 (2005).
- 363 33. MJ Ferrari, et al., The dynamics of measles in sub-saharan africa. *Nature* **451**, 679–684 (2008).
- 364 34. M Small, C Tse, Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment.
365 *Phys. A: Stat. Mech. its Appl.* **351**, 499–511 (2005).
- 366 35. T Berge, JS Lubuma, G Moremedi, N Morris, R Kondera-Shava, A simple mathematical model for Ebola in africa. *J.*
367 *biological dynamics* **11**, 42–74 (2017).
- 368 36. S Deguen, G Thomas, NP Chau, Estimation of the contact rate in a seasonal SEIR model: application to chickenpox
369 incidence in france. *Stat. medicine* **19**, 1207–1216 (2000).
- 370 37. S Riley, et al., Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions.
371 *Science* **300**, 1961–1966 (2003).
- 372 38. K Prem, et al., The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in Wuhan,
373 China: a modelling study. *The Lancet Public Heal.* (2020).
- 374 39. D Caccavo, Chinese and Italian covid-19 outbreaks can be correctly described by a modified sird model. *medRxiv* (2020).

- 375 40. B Viswanath, A Mislove, M Cha, KP Gummadi, On the evolution of user interaction in Facebook in *Proc. Workshop on*
376 *Online Social Networks*. pp. 37–42 (2009).