Supporting information for "Estimating peer effects in networks with peer encouragement designs"

Dean Eckles, René F. Kizilcec, Eytan Bakshy

Contents

1 Description of experimental conditions

The experiment consisted of two design factors: encourage initiation and conversation salience (see details in the main text). Both factors only affected the user interface when users were viewing News Feed in the Web interface for Facebook (i.e., not interfaces for mobile phones). The encourage initiation factor has three levels that determine how often the existing feedback and textbox for making a comment are shown in News Feed by default, rather than requiring a click to see. The *always* and *never* levels correspond to either always or never automatically showing existing feedback when displaying the shared content. The *sometimes* condition shows existing feedback only when the shared content appears in the first position in News Feed.

2 Summaries by condition

We provide summary statistics for the peer encouragement conditions in Table S1, including the number of assigned egos, demographics, and a set of pre-experimental covariates. Analysis of variance for these pre-experiment covariates by condition were all non-significant, consistent with successful randomization.

A simplified version of the IV analysis of the effect of feedback received on ego behavior can be presented visually. Fig. S1 shows condition-level summaries of feedback received and content production.

Table S1: Peer encouragement conditions with Ns and summaries of pre-experiment covariates. Comparisons of means and quartiles (in brackets) of variables across conditions. Analysis of variance for these pre-experiment covariates by condition were all non-significant, consistent with successful randomization. Prior posts and prior feedback received are both skewed enough to have means larger than upper quartiles. Variation between the size of conditions is intentional and reflective of the default presentation at the time of the experiment.

Conversation salience	Encourage initiation	N	Female	Age	Active peers	Prior posts	Prior feedback received
high	always	2328489	0.50	30.42	189.25	0.85	5.40
					[46, 109, 233]	[0.06, 0.22, 0.78]	[0.22, 1.44, 5.11]
high	sometimes	25618796	0.50	30.42	189.35	0.85	5.41
					[46, 109, 233]	[0.06, 0.22, 0.78]	[0.22, 1.44, 5.11]
high	never	2328194	0.50	30.44	189.16	0.85	5.40
					[46, 110, 233]	[0.06, 0.22, 0.78]	[0.22, 1.44, 5.11]
low	always	4658871	0.50	30.42	189.37	0.85	5.41
					[46, 110, 233]	[0.06, 0.22, 0.78]	[0.22, 1.44, 5.11]
low	sometimes	9313677	0.50	30.43	189.41	0.85	5.41
					[46, 109, 233]	[0.06, 0.22, 0.78]	[0.22, 1.44, 5.11]
low	never	4656022	0.50	30.43	189.48	0.85	5.39
					[46, 110, 233]	[0.06, 0.22, 0.78]	[0.22, 1.44, 5.11]

3 Covariates

As described in Materials and Methods, we used relevant covariates to increase the precision of the estimates reported in the main text. In particular, we used dummies for strata defined by three discrete covariates:

1. Prior feedback received: Quartile buckets of the number of likes and comments on the ego's posts in the 18 days prior to the experiment.

2. Active peers: Quartile buckets of the number of peers (Facebook friends) who had used the Web interface to Facebook at least 7 of the 28 days prior to the experiment.

3. Prior sharing: Quartile buckets of the number of posts made by the ego in the 18 days prior to the experiment.

Note that since the peer encouragement only affected peers viewing a post in Facebook News Feed via the Web interface, the second variable only counts peers who are active via the Web interface.

Additionally, as discussed in Section 5.1 below, the main analysis allows for dependence within each of the 80,001 clusters defined by graph partitioning. Therefore, we also included dummies for each of these clusters to eliminate any mean dependence within clusters and potentially increase precision.

4 Model

This section provides additional information about the models and estimands that can motivate the design and analysis of the peer encouragement design. Let $\mathbb{P}_{\text{egos}}$ be the set of n egos and $\mathbb{P}_{\text{peers}}$ be the set of m_{p} peers. These sets are not disjoint: nearly all units in $\mathbb{P}_{\text{egos}}$ are in $\mathbb{P}_{\text{peers}}$. Let $\mathbb{P}_{\text{e}\cup\text{p}} = \mathbb{P}_{\text{egos}} \cup \mathbb{P}_{\text{peers}}$ be the union of $m_{\text{e}\cup\text{p}}$ units. We use bold letters for matrices and capital letters for random variables.

For the question of how receiving additional feedback affects ego behaviors, key quantities of interest are contrasts between ego behaviors with different amounts of received feedback. For all egos $i \in \mathbb{E}$, we parameterize received feedback as as a function of the sum of the feedback (likes L and comments C) on an ego's posts:

$$D_i \equiv g\Big(\sum_{j \in \mathbb{P}_{\text{peers}}} L_{ji} + C_{ji}\Big).$$

For substantive and data analytic reasons discussed in Section 9, we take $g(\cdot)$ to be a logarithmic function in our preferred analysis.

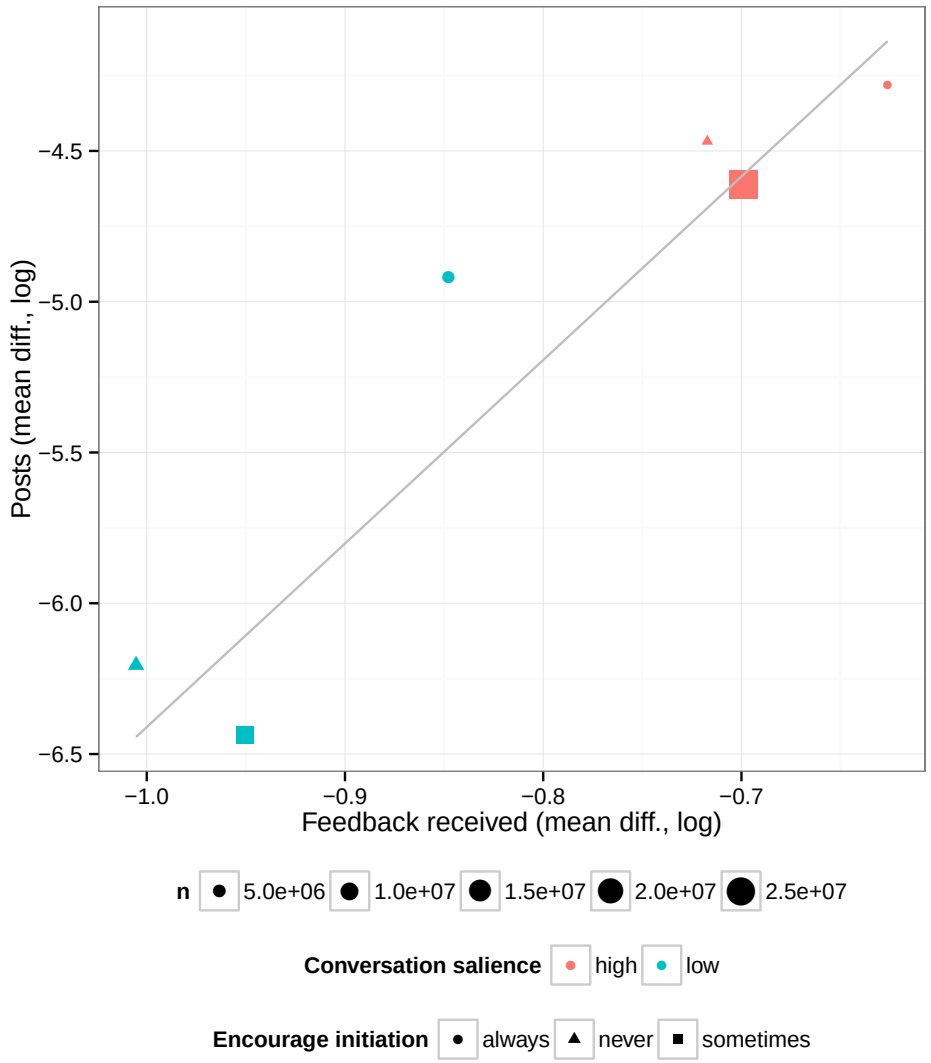


Figure S1: Summary of feedback received and posts by peer encouragement condition. Each point represents a single combination of the the two factors. We compute the mean (log) feedback received and (log) posts and subtract the same variable from the pre-experiment period. Each point is then the mean of this pre–post difference for that condition. A linear best fit line (weighted by n) is shown.

Let $Y_i(d_i, d_{-i}, z_i, \mathbf{z}_{-i})$ denote the potential outcome (e.g., posts shared during the experiment) for ego i if they received d_i feedback and were assigned to z_i and all other units receive feedback and have assignments according the rest of the $m_{\text{e}\cup\text{p}}$-vectors d and z.[1]

Without making further assumptions about other peer effects, these quantities of interest are contrasts between potential outcomes that may depend on vectors of all units' assignments and behaviors. For example,

$$Y_i(d_i, d_{-i}, z_i, \mathbf{z}_{-i}) - Y_i(d_i', d_{-i}, z_i, \mathbf{z}_{-i})$$

contrasts ego i's outcomes under two different levels of how much feedback i receives (d_i and d_i') while holding constant the other determinates of i's outcome, including their assignment, the assignment of all other units, and feedback received by other units. These two quantities are not simultaneously observable. Multiple sets of assumptions can be used to justify estimating summaries (e.g., averages) of these contrasts from data. We describe three sets of these assumptions.

4.1 Homogeneous effects

Standard treatments of instrumental variable methods and widespread practice in econometrics has, until recently, worked primarily with models in which the estimand is a coefficient in a linear model. In the present case, this model would have the form

$$Y_i(d_i) - Y_i(0) = \gamma d_i, \tag{1}$$

where we then observe the function $Y_i(\cdot)$ for the value of the observed level of the endogenous directed peer behavior $Y_i^{obs} = Y_i(D_i^{obs})$ (in this case, feedback received). The "dose" itself (feedback received) is function of the randomly assigned instrument, for which we observe $D_i^{obs} = D_i(Z_i^{obs})$.[2]

In this setting, for the finite population $\mathbb{P}_{\text{egos}}$, exact inference for γ is possible (17) by inverting a hypothesis test for $\gamma = \gamma_0$. We conduct this analysis in Section 5.2.

One important shortcoming of this model is that it implicitly makes assumptions about any interference. By writing the outcomes as a function only of a level of directed peer behaviors, this would typically require that outcomes are constant in many of the arguments previously posited (i.e. other behaviors and random assignment); that is, we have for all i that

$$Y_i(d_i) = Y_i(d_i, d_{-i}, z_i, \mathbf{z}_{-i}) = Y_i(d_i, d'_{-i}, z'_i, \mathbf{z}'_{-i})$$

[1] This specification of potential outcomes rules out certain kinds of simultaneity or feedback loops, such as the ego's outcome (e.g., posting) in an initial period causing them to receive more feedback subsequently, and then this causing the outcome in a later period.

[2] Writing $D_i(\cdot)$ as a function only of Z_i assumes that D_i has no other parents among our variables; this, along with the exclusion restriction, rules out some cases of simultaneity or aggregation of multiple time periods. For example, say an encouragement causes feedback to occur, which in turn causes an outcome, this in turn causes more feedback to occur, etc. Ogburn et al. (20) addresses the negative consequences for identification of endogeneous timing of the mediating behaviors.

for all $d \in \mathbb{D}^n$, $\mathbf{z}, \mathbf{z}' \in \mathbb{Z}^n$. This makes a hypothesis that $\gamma = \gamma_0$ a *sharp* hypothesis, in that all of the potential outcomes can be inferred from a single observation.

4.2 Heterogeneous effects

Without the assumption of Equation 1, effects may be heterogenous across units and across increments to the endogenous treatment. That is, the model might be linear but heterogeneous,

$$Y_i(d) - Y_i(0) = \gamma_i d_i, \tag{2}$$

or both heterogeneous and potentially non-linear.

In the absence of interference, the identification results from Angrist and Imbens (1) apply. That is, for each instrument there is a parameter γ that is a weighted average per-unit treatment effect, where the weights are determined by the shift in the distribution of D_i caused by the instrument. Angrist and Imbens (1) call this the *average causal response* (ACR). The weighting functions for all pairs of conditions are given by Fig. 3B in the main text. The weighting function for the main results are shown in Fig. S7.

4.3 With interference

The models above make a unit's potential outcomes invariant in changes to other units' assignments, but we expect units to be interacting.

In some other work, an explicit goal is to contrast very different treatment vectors. Hudgens and Halloran (16) consider a population average overall causal effect that compares two arbitrary distributions of treatment assignments ϕ and ϕ'. When ϕ is deterministic assignment to treatment and ϕ' is deterministic treatment assignment to control; Eckles et al. (15) call this the global ATE. This is not our goal here; rather, the present work aims to estimate effects on an individual under more-or-less the current regime, as this corresponds to questions about how egos are affected by marginal feedback and what effects small, targeted changes might have.

Following Hudgens and Halloran (16), one can define individual-level average potential outcomes in terms of a unit's assignment for some distribution ϕ of global assignment vectors. For the total effects of assignment, we have

$$\bar{Y}_i(z_i; \phi) \equiv E_\phi[Y_i(D_i(\mathbf{z}), d_{-i}, z_i, \mathbf{z}_{-i})].$$

Then define individual-level average effects by

$$\bar{\tau}_i(z_i, z_i'; \phi) \equiv \bar{Y}_i(z_i; \phi) - \bar{Y}_i(z_i'; \phi).$$

These can be summarized as the finite population average of individual-level effect for, e.g., assignment to treatment versus control:

$$\bar{\tau}(1, 0; \phi) \equiv \sum_{i \in \mathbb{P}_{\text{egos}}} \bar{\tau}_i(1, 0; \phi).$$

If units' probability of assignment under ϕ is constant in the population (as it is in our design), then the sample difference in means is an unbiased estimator. Thus, standard intent-to-treat estimators remain unbiased, but simply for a quantity that may depend on the distribution the assignment vector is drawn from. Similarly, by considering the analogous averages of the D_i, also for the first-stage effects. This is sufficient for both the numerator and denominator of the Wald estimator for instrumental variables to remain unbiased. Of course, as in the absence of interference, the estimator itself remains biased in finite samples (9).

However, interference can still affect the sampling distribution of these estimators and so affect the operating characteristics of standard methods for statistical inference; see Section 5 below. More specifically, Section 5.2 conducts a sensitivity analysis in the presence of interference; we find that this does not substantially affect our inferences.

4.3.1 Additive interference

More technically, Equation 1 does not exclude all forms of interference; in particular, if interference is additive, then this constant effects model could hold. More generally (i.e., allowing for heterogeneous effects), if there is additive interference then differences between outcomes for different peer behaviors are invariant in the other inputs; that is, that for all i there is some

$$\tau_i(d_i, d_i') = Y_i(d_i, d_{-i}, z_i, \mathbf{z}_{-i}) - Y_i(d_i', d_{-i}', z_i, \mathbf{z}_{-i})$$

for all $d, d' \in \mathbb{D}^n$, $\mathbf{z}, \mathbf{z}' \in \mathbb{Z}^n$. In this case, the individual-level average effects $\bar{\tau}_i(d_i, d_i'; \phi)$ do not depend on the choice of ϕ (as long as ϕ puts positive probability on the relevant values of D). As discussed in Section 5.2, this interference can affect inference, but in our sensitivity analysis, it does so only slightly.

5 Statistical inference

General statistical inference in networks remains a relatively open area of research, such that available methods involve strong substantive assumptions, un-scalable computation, low statistical power, and/or unclear asymptotics. Two distinct but related potential violations of standard independence assumptions apply to the present case: interference (i.e., spillovers, effects of other units' assignments) and network-correlated errors. We employ and combine multiple established methods with the aim of evaluating the robustness of the results to the specific assumptions of each. We expected that inference would not be substantially affected by accounting for potential interference and network-correlated errors since the peer encouragement is ego-specific and not correlated in the network.

5.1 Asymptotic inference with adjacency- and cluster-robust SEs

Following Conley (12), spatial econometrics has made use of estimators for variance–covariance matrices that are consistent in the presence of spatially correlated errors. These are Huber–White "sandwich" estimators of the form

$$\widehat{\mathrm{Var}}[\hat{\beta}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\hat{u}\hat{u}' \odot \mathbf{B})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{3}$$

where $\hat{u}$ is the vector of residuals, $\odot$ is element-wise multiplication, and $\mathbf{S}$ is a $n \times n$ matrix that selects and/or weights pairs of observations. In versions of this estimator robust to one-way clustering, $\mathbf{B}$ is a block diagonal matrix with $b_{ij} = 1$ if and only if i and j are in the same cluster. In the spatial case, $b_{ij} = K(s_{ij})$ where s_{ij} is the distance between units i and j and $K(\cdot)$ is a kernel.

We use this estimator with network distance, such that $\mathbf{B}^{\mathrm{adj}} = \mathbf{I} + \mathbf{A}$, where $\mathbf{A}$ is the adjacency matrix. That is, for each i, $B^{\mathrm{adj}}_{ii} = 1$ and for each $i \neq j$, $B^{\mathrm{adj}}_{ij} = A_{ij}$.

We note that while this method has been widely applied to networks and non-spatial measures of distance, including by Conley (12), the relevant asymptotics for networks are underdeveloped. The Conley (12) results use a metric space embedding, though other results use other sets of assumptions (18). This method also coincides with the use of multi-way clustering for dyadic data (3). To illustrate the performance of these methods in networks with local interference, we present some simple simulation results in Section 11 below.

The adjacency matrix for this analysis is prohibitively large. We use a sample of 1.7 million of its rows, including all columns, thus maintaining the full dependence structure for the egos included in the sample. In a smaller sample size, this would simply change the degrees of freedom for relevant t- and F-statistics, but this is without consequence at this sample size.

There may be some dependence between egos that are not each others' peers, but have mutual neighbors. The number of paths of length two is prohibitively large (i.e., over 5 billion such paths originating from the subsample of 1.7 million egos) to readily incorporate into the above estimator. However, we additionally computed estimators of the variance–covariance matrix using cluster-robust sandwich estimators. Here clusters are defined according to a conveniently available partitioning of the friendship graph for other purposes (e.g., for graph cluster randomized experiments, as in Ugander et al. (23) and Eckles et al. (15)) into 80,000 clusters using balanced label propagation (22); egos not in any of these clusters were assigned to another cluster. This analysis is partially motivated by recent results in spatial econometrics on the application of cluster-robust variance–covariance estimators to spatially dependent data (8). On its own, this cluster-robust variance–covariance estimator has the disadvantage that only approximately 30% of the edges between egos are within the same cluster, so clearly there is potential for between-cluster dependence.

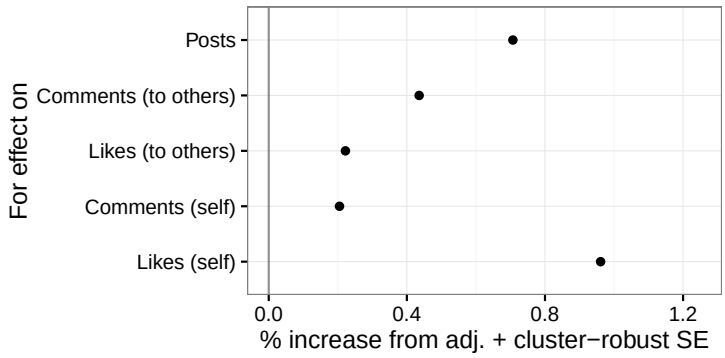


Figure S2: Comparison of adjacency- and spatial-robust SEs and standard heteroskedastic-robust SEs for the main estimates from Fig. 4 in the main text. The increases are less than 1%.

For these reasons, we use a variance–covariance estimator that combines both the adjacency- and cluster-robust estimators. The cluster-robust estimator consists of Equation 3 with a different "selector" matrix $\mathbf{B}^{\text{clu}}$ with $B_{ij}^{\text{clu}} = 1$ if and only if egos i and j are in the same cluster. This can be interpreted as arising from an alternative measurement of distance between i and j. Following, in the spatial literature, Conley and Molinari (13), Kelejian and Prucha (18), or, in work on multi-way clustering, Cameron et al. (11), one could use a weighting matrix $\mathbf{B}^{\text{both}}$ that is the element-wise maximum of the two, such that $B_{ij}^{\text{both}} = \max(B_{ij}^{\text{adj}}, B_{ij}^{\text{clu}})$. This estimator, call it $\widehat{\text{Var}}^{\text{both}}[\hat{\beta}]$, can be computed as a linear combination of estimators,

$$\widehat{\text{Var}}^{\text{m}}[\hat{\beta}] = \widehat{\text{Var}}^{\text{adj}}[\hat{\beta}] + \widehat{\text{Var}}^{\text{clu}}[\hat{\beta}] - \widehat{\text{Var}}^{\text{adj}\times\text{clu}}[\hat{\beta}],$$

where the last term is based on a selector matrix that is the element-wise product of the first two. Since $\widehat{\text{Var}}^{\text{adj}}[\hat{\beta}]$ is computed on a sample of rows and to avoid construction of the rest of these matrices, we replace $\widehat{\text{Var}}^{\text{adj}\times\text{clu}}[\hat{\beta}]$ with an estimator that is smaller in expectation, the standard heteroskedasticity-robust sandwich estimator for independent data, which is equivalent to using a selector matrix with ones on the diagonal.

The primary results in the main text make use of this combined adjacency- and cluster-robust sandwich estimator. Compared with not accounting for these potential sources of the dependence, this results in a quite small increase in SEs for the coefficients of interest. For the main IV estimates, the increase is less than 1% (Fig. S2). This was expected given that the instruments are not correlated in the network and interference was expected to be small.

5.2 Randomization inference with sensitivity analysis

We use Fisherian randomization inference to further examine the robustness of our results. First, we repeat the main analysis using the randomization inference method of Imbens and Rosenbaum (17). Second, we extend this method to conduct a sensitivity analysis to certain types of interference. We conduct this analysis for the effect of feedback received on content production (i.e., sharing posts), which is perhaps the most substantial outcome and the outcome with the largest p-value (and therefore expected to be most sensitive to alternative inferential methods).

Imbens and Rosenbaum (17) apply Fisherian randomization inference to an instrumental variables design, yielding exact tests of a constant effects model. Their simulations illustrate this method's robustness to weak instruments and higher power with long-tailed distributions. For the model in Equation 1, one can use an instrument to test hypotheses of the form $\gamma = \gamma_0$ as follows. Compute the residuals $r_0 = Y^{obs} - \gamma_0 D^{obs}$ and some statistic that is a function of the r_0 and instruments $\mathbf{Z}$. Then compare this observed statistic T^{obs} to the known null distribution of T under repeated sampling of $\mathbf{Z}$ from the distribution of treatment assignments ϕ. We construct a confidence set for γ by inverting this test. If this set is nonempty, then there are some values of γ for which this model is consistent with the data, at least with respect to the alternatives against which the choice of T has power. However, as is common to methods that invert Fisher's exact hypothesis tests of this kind, this confidence set could exclude some values of γ that are consistent with the data under a less restrictive model (i.e., one with interference).

We thus extend this method to allow us to examine the sensitivity of these results to a limited form of interference. Following Assumption 3 in the main text (i.e., direct-effect-bounded interference), we assume that the interference is smaller than the effect of some increment to D_i. In particular, we consider a model in which egos' outcomes depend linearly on the fraction of treated peers,

$$Y(d, \mathbf{Z}) = Y(0, 0) + \gamma d + \tilde{\mathbf{A}}\mathbf{Z}\zeta, \tag{4}$$

where d is an n-vector, $\mathbf{Z}$ is a $n \times 2$ matrix of indicators for whether each unit has conversation salience high and always encourage initiation, and $\tilde{\mathbf{A}}$ is the row-normalized adjacency matrix. It is possible to test joint hypotheses of the form $\gamma = \gamma_0$ and $\zeta = \zeta_0$ against some alternatives. As with the no-interference case, we compute the residuals $r_0 = Y^{obs} - \gamma_0 D^{obs} - \zeta_0 \tilde{\mathbf{A}}\mathbf{Z}^{obs}$ and some statistic that is a function of residuals r_0 and instruments $\mathbf{Z}$. This statistic is then compared with the known null distribution.

This general procedure is the same as in Bowers et al. (10), except that we apply it to inference with instruments and we treat the interference parameter as a nuisance, rather than trying to do inference for it. As we are primarily interested in γ, we can conduct sensitivity analysis by testing sets of hypotheses of the form $\gamma = \gamma_0$ and $\zeta_k \in (\zeta^-, \zeta^+)$ for $k \in \{1, 2\}$ for the two instruments we use.

We expect that any spillovers should be small compared with the direct effects (i.e.,

Assumption 3). To be conservative we allow the spillovers from each of the two columns of $\mathbf{Z}$ to be as large as γ. Since D_i is on a log scale, this corresponds to the assumption that the interference from each factor is less than the effect of a 172% increase in feedback received. Of course, we do not know γ so we examine sensitivity to spillovers as large as γ in two ways. In both cases, we set $\zeta^- = -\zeta^+$ for symmetry. First, we set $\zeta^+ = \gamma_0$, a hypothetical value of γ, such that as we test different values of γ we also test different values of ζ. This has the consequence that when testing $\gamma = 0$, we also have $\zeta = 0$, meaning that inference is not affected by interference at this point. So we also do a sensitivity analysis with $\zeta^+ = \hat{\gamma}$, such that there are the same levels of interference for all tested values of γ.

As our test statistic, we use the sum of ranks of r_0 within the four groups formed by the binary factors high conversation salience and always encourage initiation. We selected this test statistic because (a) the null distribution can be approximated without actually computing permutations (as this corresponds to the Kruskal–Wallis rank-sum test) and (b) it is expected to be sensitive to changes in γ.

We find that inference for γ is largely unaffected by these levels of linear interference; that is, the resulting confidence set is of a similar size and location as the confidence intervals from our asymptotic inference. Fig. S3 shows p-values as a function of γ. Across all settings of ζ for $\gamma = 0$, the largest p-value is 0.012, so we still reject $\gamma = 0$.

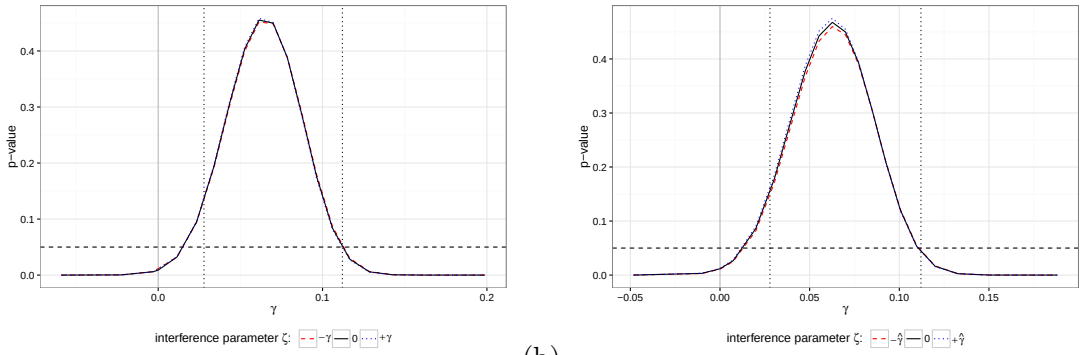
Note that because the confidence set for γ is non-empty when $\zeta = 0$, the data are consistent with there being no interference from the fraction of treated neighbors — at least to the extent deviations from this no interference model are detectable with these test statistics. That is, compared with the conditional randomization methods in Aronow (2) and Athey et al. (4), which condition on elements of the observed assignment vector, the procedure here tests a more specific model for effects of a unit's own treatment (and fails to reject it).

We also conducted the same sensitivity analysis, but with interactions between ego treatment and peer treatment. This allows for the effect of peers' assignment to have *their peers* encouraged to affect egos differently depending on each ego's assignment. This model is

$$Y_i(d_i, \mathbf{Z}) = Y_i(0,0) + \gamma d_i + \big(\tilde{A}_i \mathbf{Z} z_i - \frac{1}{n} \sum_{j \in \mathbb{P}_{\text{egos}}} \tilde{A}_j \mathbf{Z}\big)\zeta, \tag{5}$$

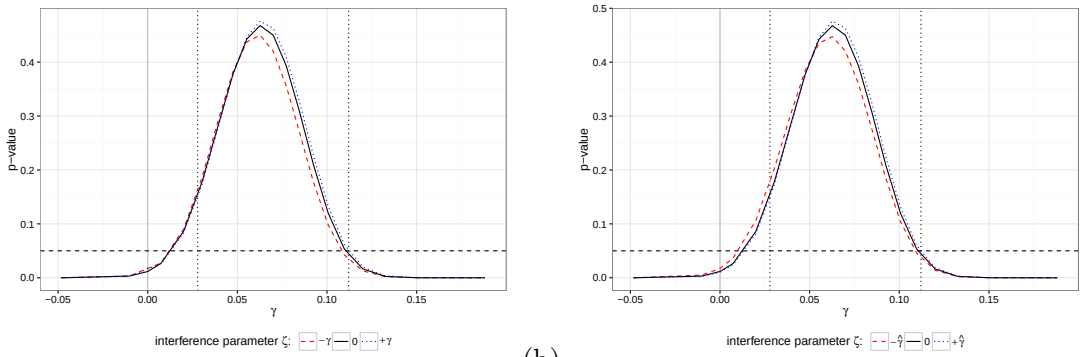
where $\tilde{\mathbf{A}}_i$ is the ith row of the row-normalized adjacency matrix $\tilde{\mathbf{A}}$, and ζ takes on the same values as before. Note that γ has the interpretation of being both the effect of feedback at the mean level of the fraction of ego-peers assigned and also, due to linearity, the average effect under a distribution of ego-peers assigned that is centered at this value. In this sense, in the framework of Hudgens and Halloran (16), inference for γ is inference for the average of average individual-level direct effects.

This interactive interference appears to affect inference somewhat more than the additive interference, but still does not substantially change the results (Fig. S4). Across all settings of ζ, the largest p-value for the hypothesis that $\gamma = 0$ is 0.017.



(a) (b)

Figure S3: Sensitivity analysis for additive spillovers using Fisherian randomization inference for the effect of feedback received on posting using the Kruskal–Wallis rank-sum test statistic. Values of γ with a p-value greater than 0.05 (dashed horizontal line) for all values of ζ are included in the 95% confidence set. Here, the confidence set is simply an interval defined by a start and end point. In (a), ζ takes on values that depend on the posited value of γ shown on the x-axis: $\zeta \in \{-\gamma, 0, \gamma\}$. In (b), ζ is constant for all tested values of γ based on our estimated value for γ: $\zeta \in \{-\hat{\gamma}, 0, \hat{\gamma}\}$. In neither case does this appreciably affect inference for the effect of feedback received on posting, γ. There is nonetheless some difference between the randomization inference confidence set and the 95% confidence interval from asymptotic adjacency- and cluster-robust inference (limits are shown as dotted vertical lines).



(a) (b)

Figure S4: Sensitivity analysis for interactive spillovers using Fisherian randomization inference for the effect of feedback received on posting, γ, using the Kruskal–Wallis rank-sum test statistic. Values of γ with a p-value greater than 0.05 (dashed horizontal line) for all values of ζ are included in the 95% confidence set. Here, the confidence set is simply an interval defined by a start and end point. In (a), ζ takes on values that depend on the posited value of γ shown on the x-axis: $\zeta \in \{-\gamma, 0, \gamma\}$. In (b), ζ is constant for all tested values of γ based on our estimated value for γ: $\zeta \in \{-\hat{\gamma}, 0, \hat{\gamma}\}$. In neither case does changing ζ substantially affect inference for the effect of feedback received on posting, γ. The limits of the 95% confidence interval from asymptotic adjacency- and cluster-robust inference are shown as dotted vertical lines.

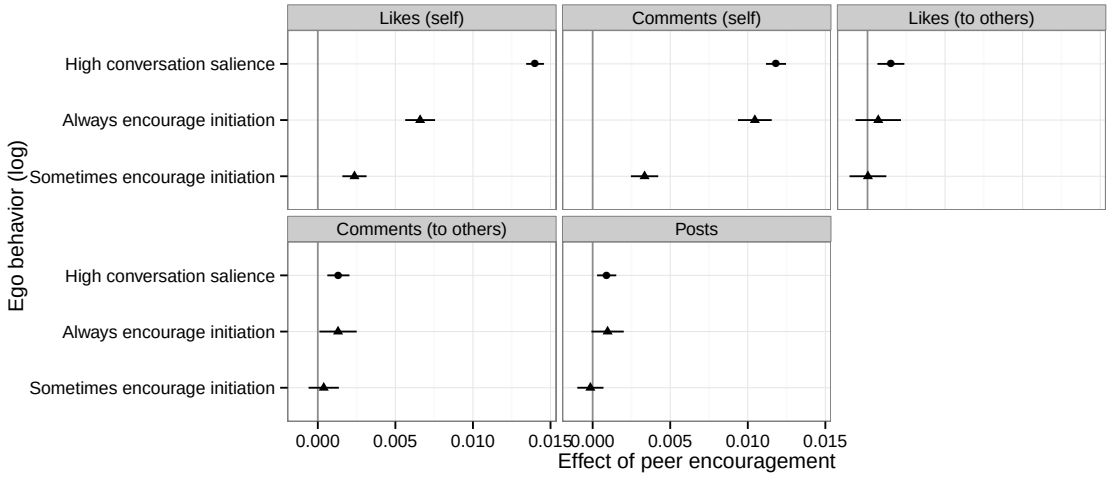


Figure S5: Effects of peer encouragements on (log) ego behaviors. These estimates are based on post-stratification on quartiles of prior feedback, friend activity, and prior sharing. Error bars are 95% heteroscedasticity-robust confidence intervals.

6 Intent-to-treat effects

When analyzing encouragement designs, it is common to report the total effects of random assignment to the encouragement on the outcomes. Figure S5 shows these "intent-to-treat" effects.

7 First-stage distributional effects

In addition to the effects on mean feedback received reported in the main text, we can compare the distributions of feedback received in different encouragement conditions. This shows what changes in feedback received are caused by the peer encouragement and thus what changes the TSLS analysis is averaging over. Figure 3 in the main text illustrates the difference in these distributions for all egos, and Figure S6 shows these distributional effects separately for egos who previously received differing levels of feedback; these are combined to produce the results in the main text.

In particular, in the case of exclusive, binary instruments, the differences in CDFs are the weights that define the ACR for that instrument or the weighted combination of ACRs. Using results extending Angrist and Imbens (1, Th. 2) to TSLS estimation with other TSLS specifications Lochner and Moretti (19, Prop. 2) and accounting for the log transformation, we compute the combined weights for the primary set of three binary

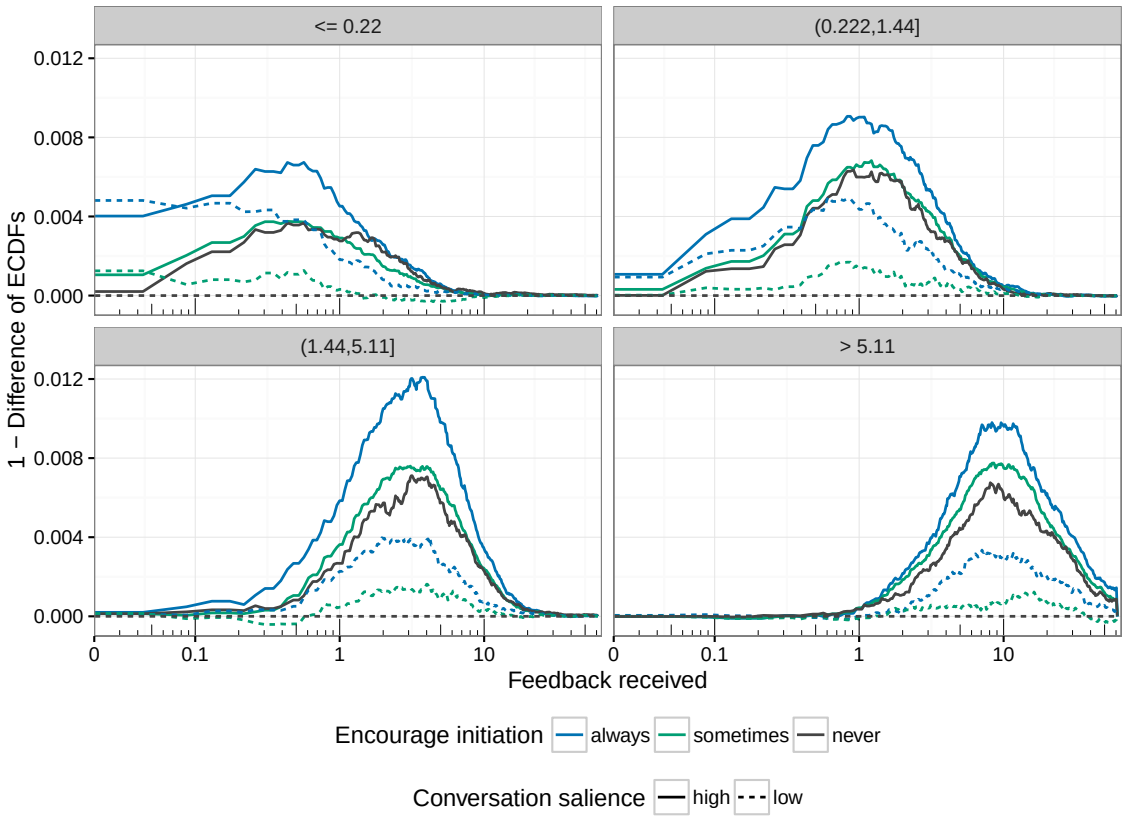


Figure S6: Effect of the encouragements on feedback received, by quartiles of prior feedback received. Using the lowest-feedback condition (never encourage initiation, low conversation salience) as the baseline, the lines represent the difference in probability that feedback received is at least the value on the x-axis. As expected, for egos who received less feedback prior to the experiment (top-left panel), the encourage initiation factor has larger effects relative to the conversation salience factor.

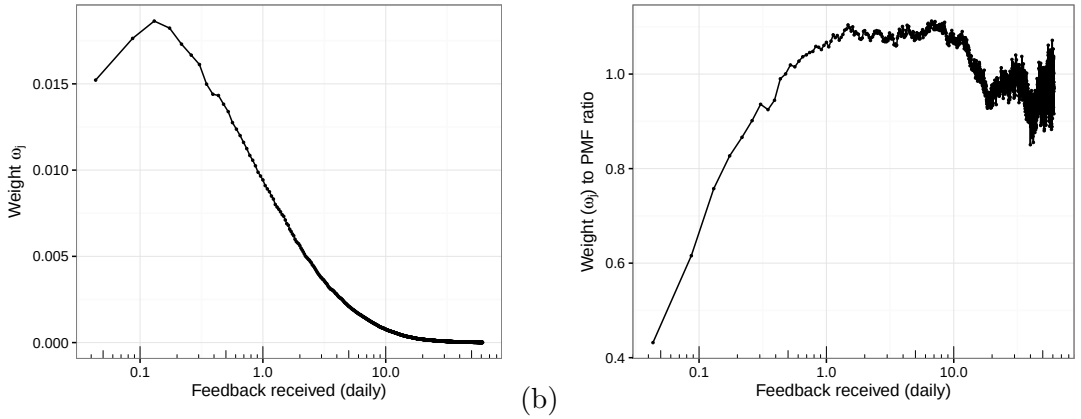


Figure S7: (a) Weighting function for TSLS with primary set of three instruments. Each point is the weight on changes from the $(k-1)$th value to the kth value. (b) Ratio of weighting function to probability mass function (PMF) for feedback received (excluding non-zero and maximum, winsorized value).

instruments (the main effects). These are displayed in Fig. S7. The weights are largest for small values of D_i, but this only partially reflects the greater probability mass on these smaller values; in fact, compared with the probability mass function, larger values of D_i are given more weight.

8 Alternative selection of instruments

We also produced estimates from multiple first-stage specifications: models with both factors and models with only the conversation salience factor and only the encourage initiation factor. We also include, following Belloni et al. (7), a lasso (i.e., L1 penalized, (21)) model. The matrix of potential instruments for this model has 325 columns with both factors (3 columns), interactions (2), and interactions with the strata-defining variables ($64 \times 5 = 320$). Note that this is intentionally overparameterized in that terms for all 64 strata (not 63) are included. The selected penalty $\lambda_s = 2.14 \times 10^{-5}$, which minimizes MSE in 10-fold cross-validation, results in a model with 23 of a possible 325 non-zero coefficients (Fig. S8), including the 3 main effects and 20 strata-specific terms.

Figure 4 in the main text and Table 8 present results from these four models. The estimates for most outcomes are statistically indistinguishable. For the "reply" behaviors, the estimates from the two factors are statistically significantly different. This could reflect that these encouragements may produce different types of feedback (e.g., comments vs. likes, or comments with different content), thus affecting the number of targets for these reply behaviors (note, e.g., that only comments, not likes, on the ego's post can themselves

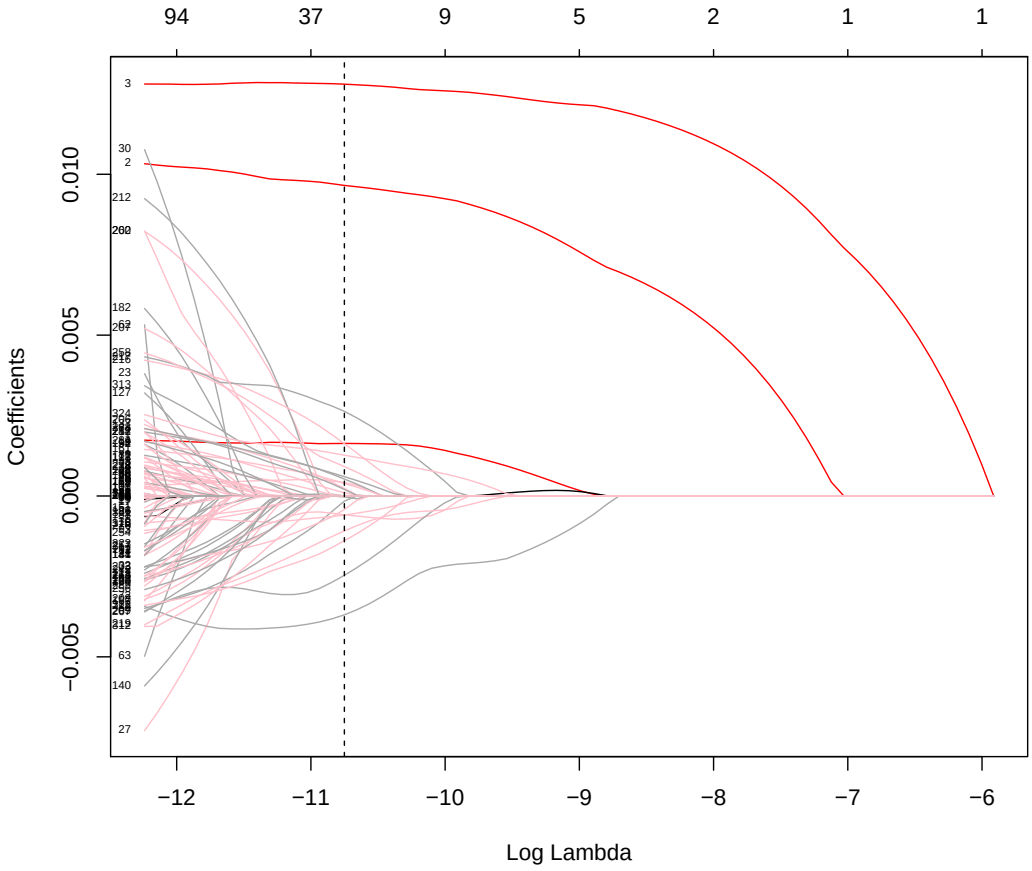


Figure S8: Regularization path for the lasso (L1 penalized) first-stage model. The selected value of the penalty λ is indicated with a dashed vertical line. The coefficient paths are numbered and colored according to: (1–3) main effects (red), (4–5) interactions of the factors (black), (6–325) stratum-specific effects, with main effects (pink) and interactions (grey). As λ is decreased, the first non-zero coefficient is for high conversation salience and the second is for always encourage initiation.

be liked by the ego). The lasso estimates do not significantly differ from the simpler model, suggesting that, at least on the log-transformed scale, there is little heterogeneity in the first stage between the 64 strata, at least to the extent that it is associated with heterogeneous effects of feedback.

Table S2: Effects of receiving feedback on five ego behaviors, as estimated using IV analysis of the peer encouragement design with four different first-stage specifications. These are coefficient estimates from a TSLS log–log model with network adjacency- and cluster-robust standard errors (in parentheses).

Outcome	Main effects	Salience only	Initiation only	Lasso
Likes (self)	1.046	1.184	0.800	1.058
	(0.026)	(0.032)	(0.048)	(0.025)
	z = 40.17	z = 36.47	z = 16.51	z = 41.50
	p < 1e-12	p < 1e-12	p < 1e-12	p < 1e-12
Comments (self)	0.964	0.968	1.060	0.961
	(0.019)	(0.022)	(0.045)	(0.018)
	z = 50.72	z = 44.72	z = 23.78	z = 52.22
	p < 1e-12	p < 1e-12	p < 1e-12	p < 1e-12
Likes (to others)	0.112	0.125	0.078	0.113
	(0.030)	(0.034)	(0.066)	(0.029)
	z = 3.78	z = 3.71	z = 1.18	z = 3.87
	p = 1.6e-04	p = 2.1e-04	p = 2.4e-01	p = 1.1e-04
Comments (to others)	0.105	0.099	0.125	0.106
	(0.024)	(0.028)	(0.053)	(0.024)
	z = 4.33	z = 3.59	z = 2.37	z = 4.46
	p = 1.5e-05	p = 3.3e-04	p = 1.8e-02	p = 8.1e-06
Posts	0.070	0.058	0.064	0.072
	(0.021)	(0.025)	(0.047)	(0.021)
	z = 3.26	z = 2.35	z = 1.36	z = 3.42
	p = 1.1e-03	p = 1.9e-02	p = 1.7e-01	p = 6.2e-04

9 Transformed and untransformed count variables

As with many behaviors in social media, counts of behaviors on Facebook are highly skewed. To guard against extreme values, all quantitative variables counting behaviors were win-

sorized: we computed the 99th percentile of the non-zero values of the variable, and all values above that were replaced with that value.

Besides the standard motivations for log-transforming thick-tailed count variables, the log-transformed variables lead to intuitively appealing models. We expected the data-generating process in the first stage to be better approximated by a multiplicative model, instead of an additive model. First, people who receive larger amounts of feedback will often have more peers who would be affected by encouragements and more posts to which it would apply. Other aspects of the News Feed system also suggest a multiplicative model. For example, the amount of feedback a story receives is among one of the top signals used to rank items in the News Feed (5), and it is well understood that content in high positions are more likely to be attended to (6, 14). Combined, it is easy to see how additional feedback could increases the likelihood that a post receives more feedback, thus introducing a multiplicative data generating process.

In the second stage model, similar general data analytic considerations apply. Furthermore, even if one expected that one additional like or comment would have the same effect for egos who receive different amounts of feedback, this is also consistent with the log-log model when the baseline levels of outcomes and feedback received are highly correlated. We therefore used a model with a log-log parameterization in our primary analyses.

We therefore used log-transformations of the quantitative variables counting behaviors in our primary analyses (we also provide estimates from linear models below). In particular, we transform these variables by adding one prior to dividing by the number of days[3] in the corresponding period and then take the natural log:

$$y = \log((y^* + 1)/n_{\text{days}}).$$

How the transformation of the outcome changes the estimand is straightforward. We can also state how the estimand of TSLS is changed by the transformation of the endogenous variable (feedback received). Let D^* be the untransformed endogenous variable. Following Angrist and Imbens (1, Th. 1), the TSLS estimand for a single binary instrument is

$$\frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{E}[D^*|Z=1] - \mathrm{E}[D^*|Z=0]} = \sum_{j=1}^{J} w_j \mathrm{E}[Y(j) - Y(j-1)|D^*(1) \geq j > D^*(0)]$$

where

$$w_j = \frac{\Pr(D^*(1) \geq j > D^*(0))}{\sum_{i=1}^{J} \Pr(D^*(0) \geq i > D^*(0))}$$

are weights that correspond to normalized differences in CDFs between $D^*(1)$ and $D^*(0)$. We define $D = g(D^*)$. D, like D^*, still takes on J values. The numerator is unchanged by working with D instead of D^*. However, the denominator changes so that the weights

[3]This rescaling is only of consequence for the non-log-transformed results in this section.

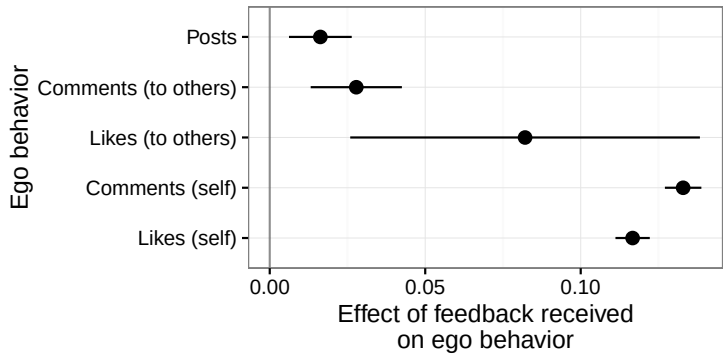


Figure S9: Effects of receiving feedback on five ego behaviors, as estimated using TSLS. Unlike in the main text, these results derive from variables that are not log-transformed. Error bars are 95% adjacency- and cluster-robust confidence intervals.

change:

$$\frac{\mathrm{E}[Y|Z=1]-\mathrm{E}[Y|Z=0]}{\mathrm{E}[D|Z=1]-\mathrm{E}[D|Z=0]}=\sum_{j=1}^{J} w_j \mathrm{E}[Y(j)-Y(j-1)|D(1)\geq g(j)>D(0)]$$

where

$$w_j=\frac{[(g(j)-g(j-1)]\Pr(D\geq g(j)>D(0))}{\sum_{i=1}^{J}[g(i)-g(i-1)]\Pr(D\geq g(i)>D(0))}.$$

As would be expected, with $g(x)=log(x+c)$, then $g(j)-g(j-1)$ decreases with j.

Similar modification apply to the other results in Angrist and Imbens (1), as this similarly affects how TSLS with multiple instruments combines the individual Wald estimates.

For comparison, Fig. S9 presents TSLS results with untransformed versions of these variables (i.e., with $y=y^*/n_{\text{days}}$ and $d=d^*/n_{\text{days}}$).

10 Simulations with ego-specific and general designs

We compare the statistical properties of the current ego-specific encouragement design with the properties of the more common general encouragement design. Based on simulated random graphs, we compute the true SE of the TSLS estimator under various conditions. To compare the two designs, we vary the random assignment and the specification of the first-stage model while keeping the second stage constant. In the ego-specific design, we randomly assign half of the nodes to be treated — implicitly assigning all of an assigned ego's peers. The ego-specific design is specified as

$$D_i=\beta Z_i+\eta_i$$

$$Y_i = \gamma D_i + 0.5\eta_i + \epsilon_i$$

with the first-stage effect size β, target parameter γ, ego-specific assignment indicator z_i (i.e., the instrument), and noise from a standard normal distribution in the first stage, η_i, and second stage, ϵ_i; the common error term η_i in the first and second stage results in confounding bias in the absence of the instrument.

Since the ego-specific design results in all peers being encouraged to a behavior directed at the ego, the equivalent in the general (non-ego-specific) design is when all peers happen to be assigned to be encouraged to that behavior (towards all of their neighbors). To achieve the same sized shock in the first stage in the general design with everyone assigned as in the specific design with $Z_i = 1$, we used the fraction of peers assigned to the encouragement as the instrument in the general design. The general design is specified as

$$D_i = \beta W_i + \eta_i$$

$$Y_i = \gamma D_i + 0.5\eta_i + \epsilon_i$$

with the proportion of an ego's assigned peers $W = \tilde{\mathbf{A}}Z$ as the instrument, but otherwise unchanged from the ego-specific specification.

We simulated 5,000 TSLS estimates based on the Watts–Strogatz small-world network model (24) for different numbers of units ($\log_2 n \in \{7, ..., 12\}$), rewiring probabilities $p_{\text{rw}} \in \{0.00, 0.01, 0.10\}$, neighborhood sizes (nei $\in \{1, 2, 5\}$, corresponding to average degree of 1, 4, and 10), and effect sizes ($\beta = 1$ and $\gamma \in \{0.0, 0.5, 1.0\}$). We use common random numbers for η and ϵ so that the randomness within the 5,000 replicates of each configuration arises from the random assignment of Z; that is, the potential outcomes are fixed.

Across all settings and as expected, the ego-specific design resulted in increased precision of $\hat{\gamma}$ and power to detect non-zero γ, compared with the general encouragement, as shown in Figs. S11 and S11.

11 Simulations with interference: Type I error rates of tests

Using simulations very similar to those in the previous section for ego-specific designs, we illustrate the performance of four methods for constructing tests for $\gamma = 0$. We use the same generative model as in the previous section, but add local interference, as in the model (Equation 4) posited by our sensitivity analysis in Section 5.2. The generative model is:

$$D_i = \beta Z_i + \eta_i$$

$$Y_i = \gamma D_i + (\tilde{\mathbf{A}}_i \mathbf{Z})\zeta + 0.5\eta_i + \epsilon_i.$$

The methods used are two tests that were expected to not be robust to interference and the two related methods we used in the main text and in the sensitivity analysis:

1. Heteroskedasticity-robust sandwich estimator for independent data,

2. Adjacency-robust sandwich estimator,

3. Wilcoxon rank-sum test,

4. Wilcoxon rank-sum test with the interference model of Equation 4.

We simulated 5,000 TSLS estimates and associated tests for each combination of $\log_2 n \in \{7, ..., 12\}$, $p_{\text{rw}} = 0.01$, neighborhood sizes (nei = 2), effect sizes ($\beta \in \{0.1, 0.5, 1.0\}$ and $\gamma = 0$), and interference ($\zeta \in \{0, 1, 2\}$). That is, when $\zeta > 0$, these simulations use very large interference, larger than even the first-stage effects.

For the Wilcoxon rank-sum test with the interference model, we use $\zeta_0 \in \{-\zeta, 0, \zeta\}$; that is, zero, the true ζ, and its negation. We select the maximum p-value for $\gamma = 0$, as in Section 5.2. In contrast to our sensitivity analysis, where we set $\zeta \in \{-\gamma, 0, \gamma\}$, we here set $\gamma = 0$.

The results are shown in Fig. S12. In the absence of interference ($\zeta = 0$), all tests have size (Type I error rates) close to the nominal size of $\alpha = 0.05$ for these settings. However, in the presence of interference, the two tests for independent data have larger-than-nominal size. On the other hand, the test using the adjacency-robust sandwich estimator and the Wilcoxon rank-sum test with the interference model both exhibit size $\leq \alpha$.

References

[1] Angrist, J. and Imbens, G. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.

[2] Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16.

[3] Aronow, P. M., Samii, C., and Assenova, V. A. (2015). Cluster-robust variance estimation for dyadic data. *Political Analysis*, 23(4):564–577.

[4] Athey, S., Eckles, D., and Imbens, G. W. (2015). Exact p-values for network interference. NBER Working Paper 21313. http://arxiv.org/abs/1506.02084.

[5] Backstrom, L. (2013). News feed fyi: A window into news feed. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed.

[6] Bakshy, E., Messing, S., and Adamic, L. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132.

[7] Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

[8] Bester, C. A., Conley, T. G., and Hansen, C. B. (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, 165(2):137–151.

[9] Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

[10] Bowers, J., Fredrickson, M. M., and Panagopoulos, C. (2013). Reasoning about interference between units: A general framework. *Political Analysis*, 21(1):97–124.

[11] Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2).

[12] Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45.

[13] Conley, T. G. and Molinari, F. (2007). Spatial correlation robust inference with errors in location or distance. *Journal of Econometrics*, 140(1):76–96.

[14] Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. (2008). An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 87–94. ACM.

[15] Eckles, D., Karrer, B., and Ugander, J. (2015). Design and analysis of experiments in networks: Reducing bias from interference. http://arxiv.org/abs/1404.7530.

[16] Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482).

[17] Imbens, G. W. and Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109–126.

[18] Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154.

[19] Lochner, L. and Moretti, E. (2015). Estimating and testing models with many treatment levels and limited instruments. *Review of Economics and Statistics*, 97(2):387–397.

[20] Ogburn, E. L., VanderWeele, T. J., et al. (2014). Causal diagrams for interference. *Statistical Science*, 29(4):559–578.

[21] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

[22] Ugander, J. and Backstrom, L. (2013). Balanced label propagation for partitioning massive graphs. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 507–516. ACM.

[23] Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. M. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proc. of KDD*. ACM.

[24] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2.

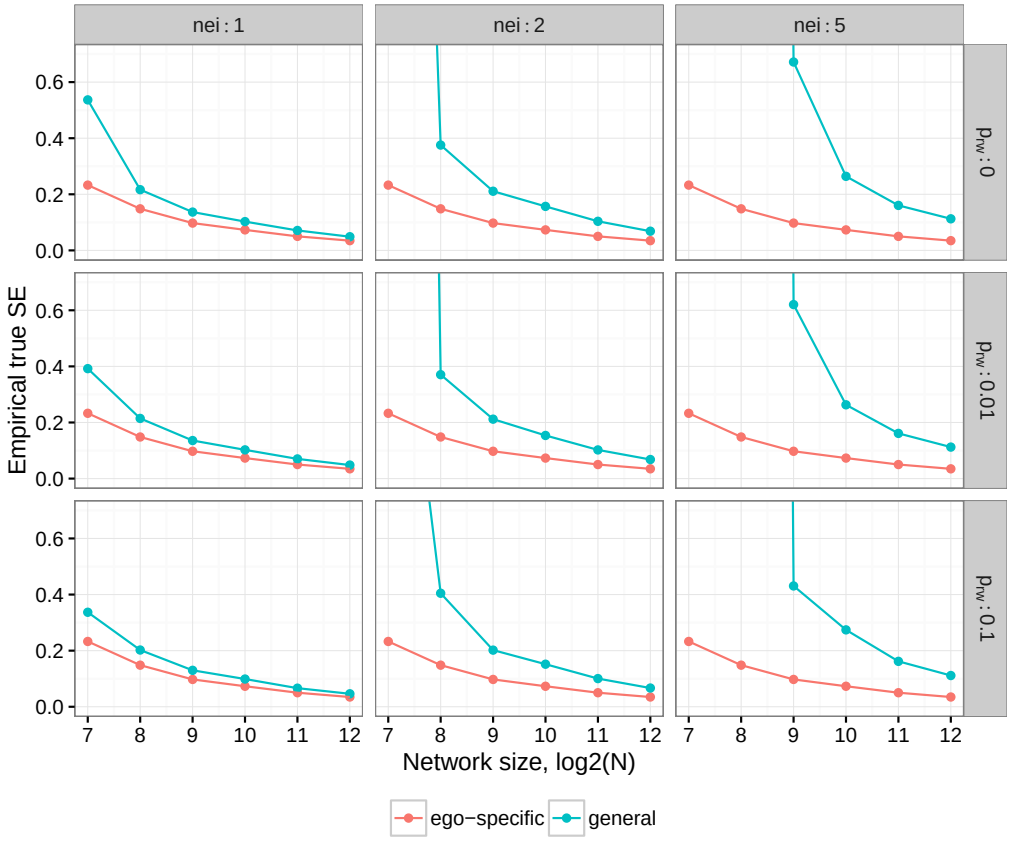


Figure S10: True SE for estimates of γ in ego-specific and general peer encouragement designs from simulations with small-world networks of different size (n), varying number of neighbors (nei), and different rewiring probabilities. The true standard error is estimated with the standard deviation of $\hat{\gamma}$ over 5,000 draws of Z. These results do not change with γ; results for $\gamma = 1$ are shown.

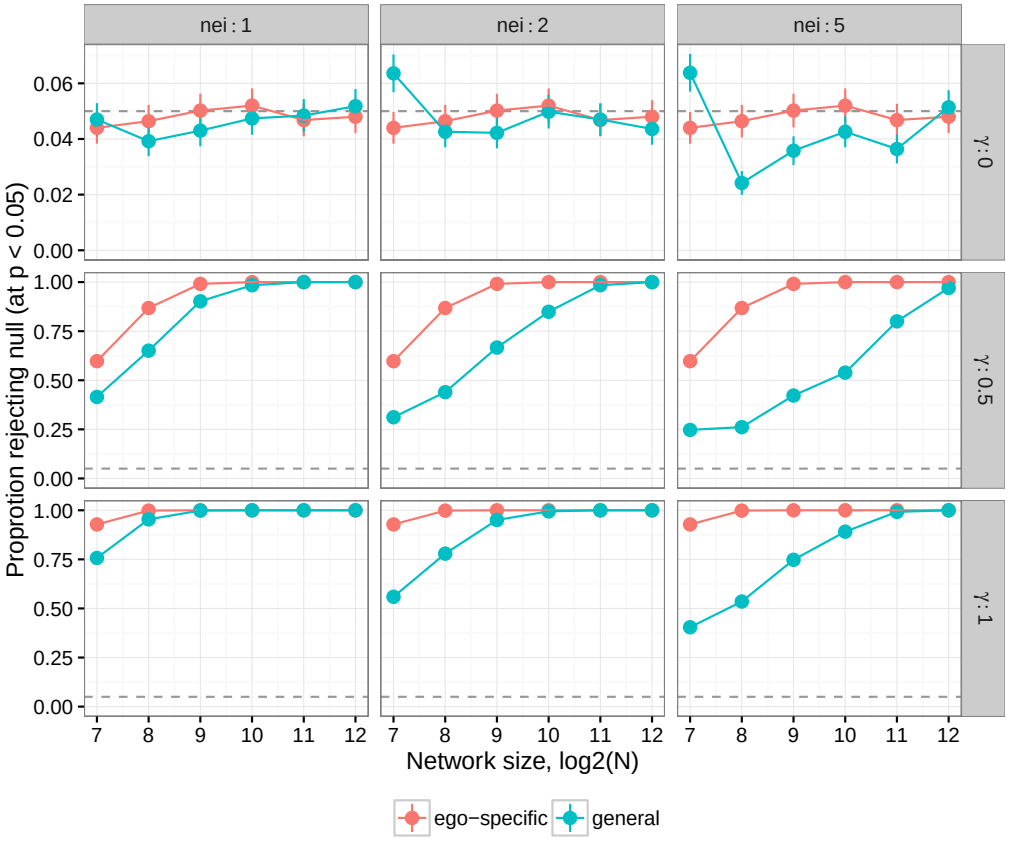


Figure S11: Rate of rejecting the null of $\gamma = 0$ in ego-specific and general peer encouragement designs from simulations with small-world networks of different size (n), varying number of neighbors (nei), and with different true effect sizes (γ). When $\gamma = 0$, this is the empirical size (Type I error rate) of the test, which appears to have size less than $\alpha = 0.05$, except for small n with the general design. When $\gamma \neq 0$, this is the power of the test. The p-values were computed using asymptotic adjacency-robust sandwich standard errors. Error bars are 95% confidence intervals for a proportion.

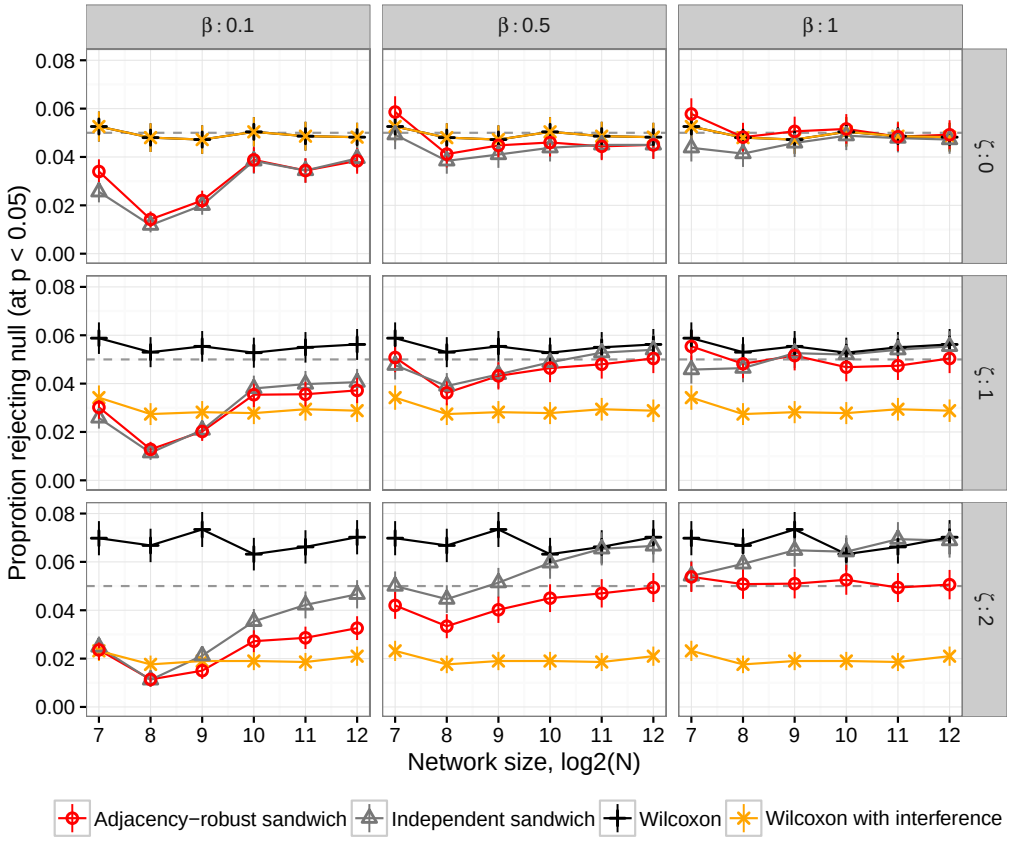


Figure S12: Size (Type I error rate) of tests for empirical size (Type I error rate) for tests of $\gamma = 0$ in the presence of varying levels of interference, ζ. In the case of $\zeta = 0$, the two Wilcoxon tests are identical by construction.