From Peer Effects to Network Evolution: Evidence from the Illinois Workplace Wellness Study*

Alden Cheng
University of Illinois and NBER

Damon Jones
University of Chicago

David Molitor
University of Illinois and NBER

Julian Reif
University of Illinois and NBER

September 2026

Abstract

We randomly assigned eligibility and financial incentives for a two-year workplace wellness program to 4,834 employees. Friendship network data were collected at baseline and again one and two years after randomization. Participation in health screenings rose by 4.2 percentage points for each additional coworker assigned to treatment, and by 6.4 percentage points for each additional coworker induced to participate as a result of treatment. The program also shaped network dynamics: eligibility increased the probability of being listed as a coworker by others in subsequent years by 9 percent, and the intervention increased network homophily along some dimensions. Behavioral spillovers amplified the effect of the intervention on screening participation by 6 percent, while dynamic spillovers through network evolution increased future participation by about 0.8 percent of baseline.

JEL Classification: D62, I1, C18

Keywords: Endogenous network formation, peer effects, health behavior

*This research was supported by the National Institute on Aging of the National Institutes of Health under award number R01AG050701; the National Science Foundation under Grant No. 1730546; the Abdul Latif Jameel Poverty Action Lab (J-PAL) North America U.S. Health Care Delivery Initiative; Evidence for Action (E4A), a program of the Robert Wood Johnson Foundation; and the W.E. Upjohn Institute for Employment Research. This study was pre-registered with the American Economics Association RCT Registry (AEARCTR-0001368). The findings and conclusions expressed are solely those of the authors and do not represent the views of the National Institutes of Health, any of our funders, or the University of Illinois. We thank participants at the Annual Conference of American Society of Health Economists, Advances with Field Experiments Conference, the CausNetS satellite at NetSci 2026, and Economic Science Association 2026 World Meetings for helpful comments and suggestions. Email: alden15@nber.org, damonjones@uchicago.edu, dmolitor@illinois.edu, and jreif@illinois.edu.

1 Introduction

Individuals' actions are often shaped by interactions with friends, yet these actions may in turn lead to changes in one's friendship network. As a result, policies implemented in social settings can generate two distinct externalities: behavior spillovers through peer effects, and changes in network structure in response to treatment. However, credibly estimating either externality is challenging. Identifying peer effects requires separating social influence from selection, common shocks, and the reflection problem (Manski, 1993), leading Angrist (2014) to argue that credible designs typically require experimental variation. Causal empirical evidence on network evolution is scarcer still—despite a vast theoretical literature on endogenous network formation—because it requires both exogenous variation and observation of the network post-treatment, data that is rarely collected in the field and often complicated by attrition.

In this paper, we provide one of the first pieces of causal evidence on both sources of externalities in a single empirical setting by designing and implementing a large-scale randomized controlled trial (RCT) on employees at a large public university. Eligibility to participate in health screenings and wellness activities were randomized among study participants, and information on their friendship networks were collected at baseline and in follow-up waves. At a high level, we use random variation in friends' treatment status to identify peer effects in health behavior. Having found that the baseline friendship network matters for health behavior, we then use randomization of treatment to identify effects on network evolution, showing that the intervention affected the future network both at the micro and macro levels (Jackson, Rogers and Zenou, 2017).

The Illinois Wellness Study lasted from 2016–2018 and involved 4,834 participants, all University of Illinois employees. We measure participants' baseline friendship network by asking them to list up to five coworkers with whom they discuss health-related behaviors with, and randomly assigned 3,300 participants to treatment, which conferred eligibility to participate in health screenings and wellness activities. Among individuals with the

same number of friends in the study, number of treated friends is random, thus allowing us to identify the effect of friends' treatment on one's health behavior, which we refer to as "reduced form" peer effects. Furthermore, under a plausible exclusion restriction—that friends' treatment affects one's health behavior only through its effect on friends' behavior—we are able to identify the effect of friends' health behavior on one's own behavior, i.e., endogenous peer effects in the terminology of Manski (1993).

We find strong evidence of peer effects in health behavior: each additional friend assigned to treatment increases an individual's own probability of participation in health screening by 4.2 percentage points, while each additional friend induced to participate in screenings by treatment increases own participation probability by 6.4 percentage points, results that are robust to state-of-the-art network inference methods (Leung, 2023). By contrast, the OLS estimate of this effect is more than twice as large, and remains so even after including high-dimensional controls for individual and friend characteristics using double machine-learning (Belloni, Chernozhukov and Hansen, 2014). Given the strong homophily in friend characteristics we observe, we interpret the large observational estimates as evidence of bias due to unobserved homophily, highlighting the importance of using experimental variation to obtain credible estimates of peer effects. A simple counterfactual simulation quantifies the magnitude of behavioral spillovers through peer effects in our setting: when almost all individuals are treated, participation in health screenings are about 5.2 percentage points higher relative to a counterfactual without peer effects, which is roughly 10% of the direct effect of treatment on screening participation.

Having shown that friendship networks generate behavioral spillovers in our setting, in the latter half of the paper we combine randomization of treatment at baseline (in 2016) with measurement of the network in follow-up waves (in 2017–2018) to study the second externality: changes in network structure at both the micro and macro levels.

Starting at the micro level, we estimate the effect of treatment on one's future friendship network. Although baseline treatment is randomized, lower response rates in the follow-up

surveys raise potential concerns over differential attrition. To address this, we exploit the directed nature of the network, the key insight being that comparisons within an individual of whether she lists participant(s) from the treatment or control group as (future) friends are uncontaminated by observable and unobservable determinants of the individual's decision about whether to respond to the survey. More concretely, we implement this by estimating dyadic regressions that include all possible pairs of respondents (i, j) where i responded to the follow-up survey, regressing an indicator for whether i lists j as a friend (in the follow-up survey) on j's treatment status and individual fixed effects for i. We estimate that treatment in 2016 increases the probability of being listed by others as a future friend by about 9 percent of baseline, both in 2017 and 2018. This effect is driven to a greater extent by higher rates of new link formation than better link retention, and is also concentrated among pairs that are more likely to be friends based on ex ante characteristics.

Building on our evidence that treatment increases the probability of forming new workplace friendships, we next characterize the links whose formation is induced by treatment. In the spirit of the instrumental variables framework, these links correspond to *compliers*: pairs (i, j) who become friends if and only if j is treated. We compare the similarity of these complier links to that of links that would form regardless of treatment. Under a monotonicity assumption on the effect of treatment on link formation, the similarity of complier links is identified using the complier-characterization approach of Abadie (2003). Using composite measures of similarity based on baseline demographic and health characteristics, we find that the links induced by treatment are more homophilous than links that would form regardless of treatment, suggesting that the friendships created by the intervention are disproportionately formed between more similar individuals, and that the intervention had an effect on a "macro" network characteristic: homophily (Jackson, Rogers and Zenou, 2017).

To quantify the implications of the spillover effects of our intervention on the network structure, we conduct counterfactual simulations that compare future screening rates with and without this dynamic externality. We find that when almost all individuals are treated,

future screening rates are about 0.35 percentage points higher when dynamic spillovers are accounted for relative to when they are not, an effect size that is roughly 7% that of behavioral spillovers from static peer effects.

This paper contributes to several strands of literature. First, we add to the literature on endogenous network formation. Despite a large body of theoretical work on this area over decades (Jackson and Wolinsky, 1996; Bala and Goyal, 2000), credible empirical evidence on causal determinants of network change is scarce, due to formidable identification and data challenges. Two notable exceptions are Carrell, Sacerdote and West (2013) and Banerjee et al. (2024), who observe many disjoint networks—classrooms and neighborhoods respectively—which they randomize into treatment and control to study effects on future network structure. We adopt a very different research design, instead randomizing treatment of nodes within a single network. Under this setup, we have two methodological innovations: (i) we demonstrate a simple and transparent way to eliminate potential individual-level biases from attrition under a directed definition of network links, and (ii) we show that despite only observing a single network, the treatment effect of an intervention on an aggregate network characteristic—specifically, homophily—is nonetheless identified under standard assumptions employed in reduced form work (namely, monotonicity).

Second, we contribute to a vast literature on peer effects. Identification challenges associated with peer effects have long been noted in the literature (Manski, 1993; Angrist, 2014), leading to the popularity of random assigned peer groups—particularly in education settings—as a research design to credibly estimate peer effects. We take a different approach, taking the observed friendship network as given and randomizing individuals to treatment and control, exploiting random variation in individuals' exposure to treated friends in order to identify peer effects. While the research design of random assignment to peer groups has the advantage of being directly policy-relevant in settings where peer groups are easy to control (e.g., classrooms or dorm rooms) given the optimal assignment literature (Carrell, Sacerdote and West, 2013; Bergeron et al., 2022), our strategy has two distinct advantages:

(i) endogenous peer effects are identified under our approach but not under random assignment to peer groups due to common shocks (Angrist, 2014), and (ii) peer effects estimates taking the friendship network as given may have greater external validity than estimates based on manipulated peer groups. Moreover, we find much weaker evidence of peer effects using organizational proximity (e.g., being in the same department) as proxies for peer groups (analogous to the use of classrooms for defining peers in education), highlighting the value of directly measuring friendship links and suggesting multiple dimensions of social connections (e.g., workplace friends versus workplace collaborators), with the influence of distinct dimensions varying across different domains of behavior (e.g., health behaviors versus work performance). At the same time, inference for peer effects estimates in networks (as opposed to peer groups that partition the sample) comes with additional challenges, and we contribute to literature employing state-of-the-art inference procedures that take these econometric issues seriously.

Third, and more narrowly, we add to a body of work on peers in the workplace. Most previous studies in this area focus on dynamics between managers and employees or coworkers working in close proximity, and the effects of these relationships on work-related outcomes (Bandiera, Barankay and Rasul, 2005, 2009, 2010; Mas and Moretti, 2009; Card et al., 2012; Glover, Pallais and Pariente, 2017; Cullen and Perez-Truglia, 2023). Our work adds to this literature in two ways: (i) we show that friendships at work also matter for non-work-related behavior, and (ii) we focus on self-reported workplace friends rather than peers as defined by organizational structure, showing that the former measure of social connections is more relevant for health-related behaviors than the latter.

This paper proceeds as follows. Section 2 describes our experimental design and data. In Section 3, we present our empirical strategy and results for peer effects in health behavior. Section 4 presents our empirical strategy and results for network evolution in response to the intervention, and Section 5 concludes.

2 Experimental Design and Data

The experimental design of The Illinois Workplace Wellness Study is shown in Appendix Figure D.1. We invited 12,459 benefits-eligible employees to participate in the study. This required first completing a baseline survey ($N = 4,834$) in July 2016. We randomly assigned 3,300 employees to the treatment group; the remaining 1,534 employees were assigned to the control group. Employees in the treatment group were all offered the opportunity to participate in iThrive, a workplace wellness program we designed in conjunction with the director of campus well-being services. In addition, they were further randomized into treatment arms which offered different financial incentives for participating in health screenings and wellness activities. Employees in both the treatment and the control group were invited to complete follow-up surveys in 2017 and 2018. See Jones, Molitor and Reif (2019) for additional details about the study.

The baseline and follow-up surveys asked respondents to list the "names of up to 5 friends, starting with the person you are most likely to discuss health-related behaviors with at work" (see Appendix Figure D.2). When respondents began typing in a name, a drop down menu appeared with names auto-populated from the university directory. This was done to reduce the number of typographic errors and to improve our abilities to match friends to other study datasets.

The left, blue bars in Figure 2 show the distribution of the number of friends listed on the survey.[1] About 80 percent of respondents listed at least one friend with a name we could match to the employee directory, and nearly 25 percent of respondents listed 5 friends. The right, orange bars in the same figure show the distribution when we limit it to friends who also participated in the Illinois Workplace Wellness Study. The social network can also be visualized using its graph, which is shown in Figure 1. We observe that a majority of 2,887 (out of 4,834) individuals are part of a single giant connected component.

[1] These results are based only on exact matches. We are still working on cleaning the friend names, which will increase the number of reported friends.

While balance in baseline characteristics with respect to treatment was established in Jones, Molitor and Reif (2019), more relevant to this study is balance with respect to the treatment status of one's peers. Hence, in Figure 3 we plot the predicted value of our main outcome—health screening—as a function of number of treated friends (controlling for number of friends in study).[2] We observe that predicted screening rate does not vary systematically with number of treated friends: if anything, there is a small decrease in predicted screening moving from zero treated friends to one treated friend, which would bias us against finding (positive) peer effects. In addition, Appendix Table 1 shows results from a regression of number of treated friends on more than 20 baseline demographic and health variables, as well as fixed effects for number of friends in the study. Only one coefficient is statistically significant at the five percent level, and the joint F-test of significance for all baseline characteristics is not statistically significant at conventional levels, providing further evidence of balance.

To illustrate homophily in our peer network, we present associations between observed characteristics of individuals and their friends. Specifically, we run regressions of the form:

$$X_i = \kappa_0 + \kappa_1 \bar{X}_{\mathcal{N}_i} + \nu_i, \tag{1}$$

where X_i is a baseline characteristic of individual i, and $\bar{X}_{\mathcal{N}_i}$ is the average of this characteristic among i's friends. Table 2 shows that for the vast majority of baseline characteristics, the estimate of κ_1 is positive and highly statistically significant, indicating strong homophily in the network.

[2]Predicted screening is obtained using an ensemble machine learning method combining logit, lasso, random forest, gradient boosting, and neural network, using baseline demographic and health variables as features.

3 Peer Effects Analysis

3.1 Notation

We denote the set of university employees by $\mathcal{N}^*$, a subset $\mathcal{N} \subset \mathcal{N}^*$ of whom are part of the study. In 2016, each study subject $i \in \mathcal{N}$ lists up to five friends that they discuss health-related behaviors with at work, and we let W_{ij} be an indicator for whether i lists j as one of these friends.[3] Note that we do not impose symmetry in these friend relationships, so W_{ij} may not be equal to W_{ji} in general.[4]

Denote the set of friends listed by individual i who are in the study by

$$\mathcal{N}_i \equiv \{j \in \mathcal{N} | W_{ij} = 1\},$$

and let $N_i \equiv |\mathcal{N}_i| = \sum_{j \in \mathcal{N}} W_{ij}$ be the number of friends listed by i. Denoting the treatment status of individual j by T_j, we can express total number of individual i's friends who are treated as $N_i^T = \sum_{j \in \mathcal{N}_i} T_j = \sum_{j \in \mathcal{N}} W_{ij} T_j$. For any outcome Y that is a dummy variable, we can similarly denote the number of i's friends who have $Y = 1$ by $N_i^Y \equiv \sum_{j \in \mathcal{N}_i} Y_j$.

We measure the friend network not only in 2016, but also subsequently in 2017 and 2018, so we can define similar variables based on the 2017 or 2018 networks. For these variables, we add a superscript t to make clear that they are defined with respect to the network in year $t \in \{2017, 2018\}$.

[3] As a convention, we set $W_{ii} = 0$ for all i.

[4] University administrative data (e.g., income, gender, occupation) are available for individuals not in the study ($i \in \mathcal{N}^* \backslash \mathcal{N}$), but these individuals were not included in the treatment randomization. Moreover, they did not provide us with a list of their friends and we lack data for several of their outcome, so we will not be using individuals who were not part of the study for our analysis, except potentially as controls.

3.2 Identification of Peer Effects

Due to randomization of treatment, the following regression provides a clean test for the presence of peer effects:

$$Y_i = \beta_0 + \beta_1 N_i^T + \delta_{N_i}^{\text{reduced}} + \epsilon_i \tag{2}$$

where Y_i is an outcome of interest such as whether i participates in health screenings or wellness activities in 2016, and $\delta_{N_i}^{\text{reduced}}$ are fixed effects for different values of N_i. Given that only individuals in the treatment group are eligible to participate in health screenings and wellness activities in 2016, we limit the sample to individuals in the treatment group: $\{i|T_i = 1\}$. The fixed effects $\delta_{N_i}^{\text{reduced}}$ are critical for identifying peer effects, since number of treated friends N_i^T is mechanically correlated with the number of friends that one lists who are in the study N_i. Nonetheless, conditional on N_i, potential outcomes are independent of N_i^T.

While equation (2) is agnostic about the mechanism through which peer effects operate, we can build on this to explore a leading explanation: endogenous peer effects, as defined by Manski (1993), which relates individuals' actions directly to the actions of their friends:

$$Y_i = \tau_0 + \tau_1 N_i^Y + \delta_{N_i} + u_i \tag{3}$$

Under the assumption that friends' treatment status affects one's behavior only through its effect on friends' own behavior, we can estimate the estimate the causal effect of friends' health behavior using an IV strategy. In this setup, equation (2) serves as the reduced form, and we instrument friends' behavior with number of treated friends in the first stage:

$$N_i^Y = \gamma_0 + \gamma_1 N_i^T + \delta_{N_i}^{\text{first}} + \nu_i. \tag{4}$$

While identification of β_1 is straightforward, inference is more nuanced. In many previous studies on peer effects, the peer network is not directly observed but rather assumed, e.g.,

at the college room or dorm level in Sacerdote (2001). In such situations, it is natural to cluster at the level at which peer effects is assumed to operate. In our setting however, we observe the actual network and as we saw in Figure 1, the largest connected component of the network consists of the vast majority of individuals in our study.

There is a tradeoff between having sufficiently many relatively balanced clusters, and the "quality" of the cluster (as defined by the ratio of links between individuals within the cluster relative to the total number of links involving individuals in the cluster). In particular, we can obtain high-quality clusters where all links are contained within each cluster e.g., by assigning the giant component to a single cluster. However, standard errors may be biased if there are only a small number of clusters (Cameron, Gelbach and Miller, 2008), and clusters with extremely unbalanced sample sizes also tend to have undesirable properties (MacKinnon and Webb, 2017). On the other hand, we can obtain many clusters, e.g., by arbitrarily assigning pairs of friends to a cluster. However, many of these clusters will be low-quality, in the sense that there are many links going out of each cluster.

Even if one has decided on an appropriate tradeoff (e.g., by fixing the number of clusters), there is also a computational challenge associated with forming clusters that maximize their quality. In fact, the exact solution to this optimization problem is NP-complete (Šíma and Schaeffer, 2006). Hence, to manage the tradeoff and to find appropriate clusters, we use the spectral clustering method introduced in Leung (2023). See Appendix Section A for implementation details.

3.3 Peer Effects Estimates

3.3.1 Reduced Form and Endogenous Peer Effects

Table 3 shows our main peer effects estimates, with standard errors obtained via spectral clustering. Reduced form and first stage results are shown in panel A, while IV and OLS results are shown in panel B.

Starting with panel A, the reduced form estimate in column 1 shows that an additional

friend being assigned to treatment increases the probability that an individual participates in a health screening by 4.2 pp, an effect that is statistically significant at the one percent level. Column 2 shows the first stage for peer effects on health screening is strong: in particular, number of treated friends is highly predictive of number of friends participating in health screenings, with a first stage F-statistic of 204.8. Columns 3 and 4 show analogous results for wellness activities: the reduced form estimate on the effect of treated friends on participation in wellness activities is positive (albeit marginally significant), and the first stage is strong (with an F-statistic of 95.8).

Turning to the estimates of endogenous peer effects in panel B, the IV estimate in column 1 indicates that each additional friend participating in health screenings increases one's own probability of participating by 6.4 pp, which is an 11 percent increase relative to the baseline mean. By contrast, the OLS estimate in column 2 of a 14.3 pp increase is more than twice as large, highlighting the potential for homophily to bias observational estimates of peer effects. Estimates of endogenous peer effects in wellness activities in columns 3 and 4 are qualitatively similar, with the OLS estimate being substantially larger than the IV estimate.

To explore whether selection in observables can explain the bias in the OLS estimates, Appendix Table D.1 presents further observational estimates that include more than 50 controls for baseline characteristics of individuals and their friends. Columns 1 and 2 repeat the baseline OLS and IV estimates from Table 3 for easier reference, while columns 3 and 4 show observational estimates that include controls using OLS and the post-double selection method described in Belloni, Chernozhukov and Hansen (2014) respectively (with panels A and B showing results for health screening and wellness activities). We observe in columns 3 and 4 that the inclusion of controls reduces the observational estimates of endogenous peer effects from the OLS estimate in column 2 without any controls. Yet, even controlling for these baseline characteristics, the observational estimates remain substantially larger than the IV estimates. Taken together, these results suggest that the bias of the observational peer effects estimates are likely due to selection on unobservables, underscoring the importance

of using experimental variation for identification.

3.3.2 Heterogeneity in Peer Effects

While analysis above treats all friends symmetrically, in reality we might expect some friends to matter more than others. In fact, our survey question gets at this by asking individuals to list friends in order, starting with the friend that one is most likely to discuss health behaviors with, to the least likely. Hence, we use this feature of our survey design to probe for peer effect heterogeneity by testing whether the treatment effect depends on the friend ordering. Specifically, we regress individual behavior on dummies for whether their first through fifth friend was treated (which we define to be zero if the individual did not list such a friend).

We plot the coefficient estimates in Appendix Figure D.3. While the point estimates tend to be larger for friends who are listed higher in individuals' friend lists, the 95 percent confidence intervals for the coefficient estimates overlap and an F-test of joint significance is unable to reject the null of homogeneous treatment effects.

We also test for heterogeneity along individual and friend characteristics. This is motivated by a literature on shared identities, for instance along racial (Fairlie, Hoffmann and Oreopoulos, 2014; Ye and Yi, 2023) and gender (Alsan et al., 2024) lines, which would predict larger effects for friends whose gender or race agrees with that of the individual. There is also work suggesting individuals from certain groups may be regarded as more knowledgeable, and are thus likely to be more persuasive (Sarsons, 2022). These views could either be due to accurate Bayesian inference, discrimination, or stereotypes.

To test whether there is heterogeneity in peer effects that supports these explanations, we estimate a version of equation (2) where we replace the number of friends treated with the number of friends treated that are from a particular group, interacted with dummies for the individual's own group. In addition, we replace the fixed effects for number of friends in the study with fixed effects for number of friends in the study from each group.

We explore heterogeneity by gender, race, age, and faculty status, and plot the coefficient estimates in Appendix Figures D.4a and D.4b. The estimates in some specifications are quite imprecise, but we see a consistent pattern whereby female friends seem to have a larger effect on individual behavior, regardless of the individual's gender.

3.3.3 Higher-Order Effects

Individuals' health behaviors may not be influenced only by their friends. In particular, friends-of-friends may also have an effect, since their direct effect on friends may also indirectly affect the individual. To test for these higher-order effects, we run similar regressions as equation (3), but replacing number of screened friends with number of screened friends-of-friends as the main endogenous variable, and similarly for the fixed effects. We estimate this using both OLS and IV (with the number of treated friends-of-friends as the instrument), and also include number of screened friends as an endogenous variable in some specifications (in which we include the corresponding fixed effects, and use number of treated friends as an additional instrument in the IV specification).

Results are presented in Appendix Table D.2. Comparing the IV estimates in columns 1 and 2 with the OLS estimates in columns 3 and 4, we observe a similar pattern to the main peer effects estimates: observational estimates of the effect of friends-of-friends' screening participation on own participation are about twice as large as the IV estimates. At the same time, while the IV estimates of the effect of friends-of-friends' screening participation on own participation are statistically insignificant at the five percent level, they remain positive.[5]

[5]Moreover, magnitudes of the IV point estimates are large. If friends-of-friends had no direct effects on individual behavior, we would expect the estimated effect of friends-of-friends on own participation without controlling for friends' participation to be roughly the square of the effect of friends' participation (e.g., $0.06^2 \approx 0.0036$ for screening), and zero conditional on friend participation. Instead, focusing on panel A for screening, the unconditional point estimate in column 1 is larger—at 0.023—and the conditional point estimate in column 2 is 0.019, with the difference of 0.004 interestingly being similar to the mediated effects of friends-of-friends' participation through friends. Results for Wellness Activities are qualitatively similar. Statistical significance notwithstanding, these IV estimates suggest that friends-of-friends may have a direct effect on individuals' behavior, that the measured peer network may understate social connections in the workplace, and hence, that full extent of behavioral spillovers arising from our intervention may exceed those captured by our main peer effects estimates.

3.4 Robustness Checks

3.4.1 Functional Form Assumptions

Equation (2) assumes that health behavior is linear in the number of friends treated. However, this parametric form may be inappropriate if individuals who listed fewer friends are influenced to a greater degree by each friend (in which case the share of treated friends may be a preferable specification), or if there are nonlinear effects.

Appendix Table D.3 shows that our results are robust to alternative functional forms, such as using the share of treated friends,[6,7] or dummies for each possible value of number of friends treated. Similarly, Appendix Table D.4 shows that our main results for peer effects are robust: individuals with a higher share of friends participating in screenings/wellness activities are more likely to participate in screening/wellness activities themselves, and with the OLS estimates being substantially inflated compared to the IV estimates.

In the specifications with dummies for each possible value of number of friends treated, we observe that the coefficient estimate for five friends treated is larger than the other coefficients (beyond what a linear extrapolation would predict). One possible explanation for this pattern is that these individuals with five treated friends may have had more treated friends, but could only list five friends in their survey response. Nonetheless, to check that our results are not entirely driven by the relatively small number of individuals who reported five friends in the study, we estimate equation (2) excluding these individuals and report the results in Appendix Table D.5. For the results on screening participation, we observe that the coefficient estimates on the effect of number of treated friends in the reduced form and on the effect of number of screened friends in the IV are smaller, but still statistically significant at a 5 or 10 percent level. While the corresponding coefficient estimates for Wellness activities

[6]The share is defined as zero for individuals who listed no friends.

[7]It is unclear whether the specification with number or share of friends treated is preferable. On the one hand, one does not need to include fixed effects for number of friends in the study when using the share as the right-hand-side variable. However, under certain asymptotics (e.g., if the number of friends grows large), the share of friends treated will tend towards the overall share treated under simple randomization, in which case there will be very limited variation in the share, which may lead to imprecise estimates.

are no longer statistically significant (which is not entirely surprising, given that they were only marginally statistically significant in Table 3), they remain positive.

3.4.2 Common Shocks

A potential threat to the exclusion restriction of the IV for endogenous peer effects is the potential for common shocks. In order for the IV results to be explained by common shocks in our setting, the shocks must vary systematically with the number or share of one's friends who are treated, which seems implausible if treatment was truly random. However, being in the same home college or home organization is a strong predictor of friendship links and we might expect shocks that operate at the college/organization level, so we may be concerned about common shocks as a finite-sample issue.

To test whether this is likely the case, in Appendix Table D.6, we present IV estimates of endogenous peer effects, including home college and/or home organization fixed effects. We observe that the inclusion of fixed effects reduces the IV estimates of the effect of number of screened friends on own screening by less than 10 percent, and the effect of number of friends participating in wellness activities on own participation by at most 25 percent. Given that these specifications leverage within-college/organization variation in the number of friends treated, this supports the interpretation of our IV results as being due to endogenous peer effects, rather than common shocks.

3.4.3 Alternative Peer Group Definitions

Our measure of friendship links are based on a survey question asking individuals to list up to five friends whom they discussed health-related behaviors with at work. In reality, there is likely to be measurement error, as individuals may forget to list some friends, or may have wanted to list more than five friends but could not do so.

In our main analysis, we adopted a directed measure of the network, i.e., if i lists j as a friend, we include j amongst i's friends but not necessarily the other way around (unless

j also lists i as a friend). One could imagine that in some cases where friendship links are not reciprocated, it could be that the person forgot or already listed five friends. Hence, to check the extent to which this is likely the case, we define a "reversed" friendship link, in which we count i as a friend of j's if and only if i listed j as a friend, and estimate equation (2) using this reversed friendship measure. If all unreciprocated friendship links are due to measurement error, then we would expect estimates that are very similar to those in Table 3, whereas if none are due to measurement error, we would expect estimates that are attenuated relative to Table 3.[8]

The estimates from regressions using the reversed friendship measure (shown in Appendix Table D.7) are somewhat smaller than when using our preferred friendship measure in Table 3 when screening is the outcome variable, but interestingly, larger when participation in wellness activities in the outcome. We view this as mixed evidence on whether a directed or undirected network measure is preferable.

An alternative definition of peers—which is commonly used in peer effects research where there is no direct measure of the network—is to assume that individuals in closer proximity to each other (e.g., roommates or classmates) are peers. In our setting, it seems natural to assume that individuals in the same home college or home organization are more likely to be peers, so Appendix Table D.8 presents OLS and IV results comparing estimates obtained using our survey-based measure of peers to estimates obtained when assuming that one's peer group consists of all individuals in the same home college or home organization. Given that home college and home organizations tend to be much larger than the number of friends listed by individuals in the survey, we focus on the share of the peer group participating in screening or wellness activities (rather than the number of individuals in the peer group), leaving out the individual herself. Similarly, the instrument is defined as the (leave-out) share of the peer group that is treated.

[8]The estimates will not be zero even if all unreciprocated links do not exist in reality, because in the regressions using the reversed friendship links, the estimates will still pick up peer effects coming from reciprocated links.

Appendix Table D.8 shows peer effects estimates using these alternative peer group definitions. Columns 1 and 2 of Appendix Table D.8 present IV and OLS estimates using our preferred measure of the peer network for easier comparison (repeating results already shown in Appendix Table D.4), while columns 3 and 4 (respectively, columns 5 and 6) show estimates using home college (home organization) as the measure of the peer group. The OLS estimates in columns 4 and 6 demonstrate a strong positive association between the leave-out share of one's home college or organization that is screened and own screening. While this association cannot be interpreted as causal, it does suggest that peer groups partially overlap with home college/organization, and/or that there are common shocks that operate at the college/organization level. IV estimates in columns 3 and 5 are both statistically insignificantly different from zero, with much larger standard errors than the preferred specification in column 1. Standard errors are especially large when defining the peer group as one's home college in column 3, since there are relatively few unique home colleges, so there is limited variation in the share of one's peer group that is treated as evidenced by the weaker first stage.[9]

These results show the value of having a direct measure of the peer network, without which one may lack power to detect peer effects even when they are present. In addition, it suggests that social networks may have multiple dimensions, with different implications for different outcome domains: while organizational proximity may be a reasonable measure for studying peer effects in work productivity, it may be less appropriate for measuring behavioral spillovers in health behaviors.

3.5 Behavioral Spillovers

An implication of our peer effects estimates is that behavioral spillovers may amplify the effect of our intervention. To quantify the magnitude of these externalities, we simulate the

[9] In fact, asymptotically as the size of each group tends to infinity, with simple randomization the variation in the instrument will vanish since the fraction treated in each home college will tend to the overall fraction treated.

share of individuals in the treatment group participating in health screenings as a function of the share of the population treated, comparing simulated screening rates under our IV estimates of endogenous peer effects in Table 3 with screening in the absence of peer effects. In addition, we compare effects under two modes of treatment assignment: randomized treatment (as is the case in our experiment), and targeted treatment based on a measure similar to Katz-Bonacich centrality but adapted to network settings with directed links (see Appendix Section C.3). Under the endogenous peer effects specification, each individual's propensity to participate in screenings depends on their friends' propensities to participate, so in each simulation we solve a fixed point in participation probability among all individuals in the network. Simulated screening behavior using the reduced form estimates of peer effects produce very similar results. Appendix Section C describes the simulations in more detail.

Results are shown in Figure 4. First, we observe that under randomized treatment, the screening rate among the treatment group increases with the share of the population treated, consistent with the positive behavioral spillovers generated by peer effects. In terms of magnitudes, the effect of treatment on screening probability when almost everyone is treated is about 0.592, compared to 0.540 when almost nobody is treated, a difference of about 5.2 percentage points, or roughly 10% of the baseline screening probability in the absence of peer effects. In our experiment, only 68% of individuals are treated, so the behavioral spillovers are somewhat smaller, at about 6% of the baseline.

The pattern looks somewhat different under targeted treatment. Screening rates among the treated rise much faster as a function of the share of the population treated than under randomized treatment, before decreasing, and eventually coinciding when all individuals are treated. Given that individuals who are most likely to generate the greatest positive externalities, the declining pattern when a relatively large share of the population is treated is unsurprising. The fact that screening rates are not monotonically decreasing under targeted treatment is slightly more surprising, but is due to individuals in the control group being ineligible for screening participation. Hence, for there to be behavioral spillovers from one

individual to the next, both must be assigned to the treatment group, and this is less likely when a relatively small proportion of the population is treated.

In the preceding analysis, we find causal evidence that health behavior propagates through links in the friendship network. Yet, these links themselves are not fixed: instead, they emerge through decentralized behavior and interactions between individuals. This raises the possibility that changes in behavior induced by the experiment may in turn precipitate changes in the network. Hence, in the next section, we provide causal evidence on the effect of our intervention on network evolution.

4 Endogenous Network Formation

The analysis in the previous section shows that social networks play a causal role in the diffusion of health-related behavior in the workplace at a point in time. On the other hand, how health behavior propagates through the population over time depends on potential changes to the network over time, which may be influenced by the intervention itself. This is particularly plausible in our setting, given that wellness programs may bring individuals together, or change habits or behavior that affect their attractiveness as a potential friend. Hence, in this section, we exploit experimental variation in treatment to study the causal effect of our intervention on the social network.

4.1 Network Evolution at the Micro Level

We start by estimating the effect of being treated on individuals' future friendships. Figure 5 presents graphical evidence. Specifically, consider all pairs (i, j) where i is an individual who answered the survey in year $t \in \{2017, 2018\}$ and j is any individual in the study. We compute the probability that i listed j as a friend in year t as a function of whether i and j are treated in 2016, and plot these probabilities (normalized by the friendship probability for the group where both i and j are untreated). Panel (a) shows that the probability that

i lists j in 2017 does not vary significantly whether i is assigned to treatment or not. On the other hand, the probability that i lists j is higher when j is treated compared to when j is not. Panel (b) shows that the same pattern holds qualitatively for the 2018 network. In other words, being treated does not increase the future number of friends one reports, but it increases the probability that others list the person as a friend in the future.

Despite treatment being randomly assigned, the evidence in Figure 5 is not necessarily causal due to the potential for two sources of differential attrition. We outline the intuition here, and provide mathematical details in Appendix section B.1.

First, there is the "classical differential attrition" concern that if i's treatment affects the probability that i responds to future surveys, treatment conditional on future survey response may no longer be random.

Classical differential attrition is less likely to pose a problem for the effect of j's treatment T_j on the probability that i lists j as a friend (i.e., the last two bars in Figure 5), which is where we see a larger effect. This is because all study participants j are included in the sample whether or not they responded to future surveys. However, this estimate may still be biased from a second source of differential attrition that is specific to network settings—if individuals' decision on whether to respond to the survey is influenced by whether their friends are treated.

To account for the potential for these sources of differential attrition, we include various sets of fixed effects in dyadic regressions. Specifically, we estimate regressions of the form:

$$W_{ij}^{t} = \varphi_0 + \varphi_1 T_j + \varphi_2 T_i + \gamma W_{ij}^{2016} + B_i'\psi + e_{ij}, \tag{5}$$

where W_{ij}^{t} is an indicator for whether i lists j as a friend in year t, and B_i are controls that address differential attrition. In our most rigorous specification, B_i consists of individual fixed effects δ_i, eliminating potential individual-level biases from differential attrition in our estimates of φ_1 at the cost of no longer being able to estimate φ_2 (given that T_i is absorbed

by the individual fixed effects). Alternatively, we can still obtain an unbiased estimate of φ_1 while still being able to estimate φ_2 if we are willing to impose more structure on the selection equation for survey response (although the estimate of φ_2 may still be biased by classical differential attrition). One such type of assumption is that the selection equation is common across individuals and depends only on T_i, N_i^T, and unobservables ω_i—in which case letting B_i be fixed effects for number of friends and number of treated friends δ_{N_i,N_i^T} suffices—but we can also relax this by allowing for heterogeneity in the selection equation based on observables, as well as for higher-order effects such as dependence on friends-of-friends and so forth. Appendix Section B.2 provides additional details.

Table 4 shows that an individual j being treated increases the probability of i listing j as a friend by about 0.0025 pp in 2017, and 0.0019 pp in 2018, consistently across specifications. This may seem like a small number, except that j can be any of 4,834 study participants; in particular, the effect size is roughly 9 percent of the dependent variable mean. In none of the specifications is the effect of own treatment statistically significant.

We also observe that our two approaches to controlling for differential attrition — including either the fixed effects δ_{N_i,N_i^T} or δ_i — do not seem to affect the coefficient estimate of φ_1. In fact, the largest change in the estimate of this coefficient comes when we include a control for i listing j as a friend in the baseline year of 2016. This suggests that differential attrition is unlikely to affect estimates of φ_1 to a large degree, which is unsurprising given the results in Appendix Table D.9 that number of treated friends has no statistically significant effect on own future response rates. Following Abadie et al. (2023), given that our main treatment is coworker treatment T_j, we should at least cluster our standard errors at the level of coworker j. To be conservative, we implement two-way clustering on both respondent i and coworker j, although our results are also robust to clustering only at the level of j.

To understand whether this effect is coming from formation of new links or retention of old links, we partition the sample into coworker pairs that were or were not friends at baseline (in 2016). Specifically, to study new link formation, we limit the sample to pairs that

were not friends in 2016, whereas to study retention of preexisting links, we limit to pairs who were already friends in 2016. To abstract from potential attrition issues, we consider specifications with individual fixed effects, and since it is no longer possible to control for baseline friendship (given that this is constant within each subsample we consider), we use predicted baseline friendship probability (based on a ridge regression at the pair level that predicts 2016 friendship baseline characteristics of the coworker pair) as a control.

Table 5 shows regression estimates of the treatment effect for new link formation in columns 1 and 2, and link retention in columns 3 and 4, both for the 2017 and 2018 networks (in panels A and B respectively). We find positive estimates for both formation and retention of links: estimates for link formation are statistically significant in both years, while estimates for link retention are only statistically significant for the 2017 network, potentially due to the much smaller sample size. In addition, we observe that even holding baseline friendship fixed, baseline friendship probability remains a strong predictor of future friendship, which motivates us to study whether the treatment effect is concentrated among coworker pairs with high baseline friendship probability.

To test this, we include baseline friendship probability and its interactions with the treatment status of the focal individual and their coworker as covariates in the regression. Table 6 presents the estimated effects on the 2017 and 2018 network in panels A and B respectively. Estimates for the full sample are shown in column 1, while columns 2 and 3 display results for the formation of new links and retention of preexisting links respectively. We observe that the interaction between coworker's treatment and baseline friendship probability is positive and statistically significant in the full sample (and also typically positive in the subsamples, albeit less precisely estimated), suggesting that the treatment effects on network evolution are driven to a greater extent by coworker pairs with high baseline friendship probability. The magnitude of this heterogeneity is also economically significant: for instance, the coefficient estimate on the interaction term for the full sample in panels A and B implies that the effect of being treated on being listed by others as a friend in the future is 0.110–0.112

percentage points (which is 4–5 times the dependent variable means of 0.028 and 0.022) larger for pairs that are 1 percentage point more likely to be friends at baseline.

4.2 Network Responses at the Macro Level

The finding that the treatment effect on network evolution is concentrated among pairs with high baseline friendship probability leads naturally to the related but distinct question of whether treatment increases network homophily. In particular, even if the treatment effect is concentrated among coworker pairs with more similar baseline characteristics, treatment may still increase or decrease homophily depending on whether homophily is stronger among marginal or inframarginal friendship pairs.

To formalize this, we adopt potential outcomes notation, using $W_{ij}^t(T_j)$ to denote whether i lists j as a friend at time t as a function of whether j is treated, and define marginal friendship pairs as $\left\{(i,j) : W_{ij}^t(1) > W_{ij}^t(0)\right\}$. Note that this formulation implicitly invokes a "link-SUTVA" type of assumption: more generally, potential outcomes for W_{ij}^t may depend on the entire treatment vector. For example, W_{ij}^t may depend not only on j's treatment but also on i's treatment, or the intervention may affect the extent to which the cap of five friends on the survey becomes binding. In practice, we view these violations as being relatively unlikely (or at least uncommon), given the evidence in columns 2 and 3 of Table 4 that T_i has little effect on $W_{ij}(t)$, and the relatively small number of individuals who report the maximum of five friends.

Letting S_{ij} be a measure of homophily between i and j, treatment of j, the question of whether treatment increases homophily is determined by the sign of the following quantity:

$$\theta \equiv \mathbb{E}\left[S_{ij}|W_{ij}^t(1) > W_{ij}^t(0)\right] - \mathbb{E}\left[S_{ij}|W_{ij}^t(1) = W_{ij}^t(0) = 1\right].$$

To estimate θ we use a method closely related to Abadie (2003) for characterizing compliers in IV. We start by making three assumptions:

1. *(Random Assignment of Treatment)* $T_j \perp \left(W_{ij}^t(1), W_{ij}^t(0)\right)$

2. *(Treatment Affects Friendships)* $Pr(W_{ij}^t|T_j = 1) > Pr(W_{ij}^t|T_j = 0)$

3. *(Monotonicity)* $W_{ij}^t(1) \geq W_{ij}^t(0)$ *a.s.*

Assumption 1 is satisfied given random random assignment, and assumption 2 is supported by the results in Table 4. The monotonicity assumption — that treatment destroys some potential friendship links — is fundamentally untestable, but quite intuitive. In particular, while there are numerous plausible channels through which treatment may increase an individual's friendship prospects — such as positive effects of treatment on approachability, visibility, or social status in the eyes of others — it is less clear how treatment may actively diminish them.

Under these assumptions, we can estimate $\mathbb{E}\left[S_{ij}|W_{ij}^t(1) = W_{ij}^t(0) = 1\right]$ using the mean of S_{ij} among coworker pairs with $W_{ij}^t(0) = 1$, and the mean among "marginal friends", $\mathbb{E}\left[S_{ij}|W_{ij}^t(1) > W_{ij}^t(0)\right]$ using Abadie's kappa method Abadie (2003). Specifically, to estimate the mean of S_{ij} among "marginal friends" while incorporating controls (e.g., individual fixed effects), we estimate an IV regression with $S_{ij}W_{ij}^t$ as the outcome, using T_j as an instrument for W_{ij}^t.

Figure 6 shows the estimated mean S_{ij} for three types of friend pairs in 2017 and 2018 respectively:

1. "Never Friends": $\left\{(i,j) : W_{ij}^t(1) = W_{ij}^t(0) = 0\right\}$ in light blue,

2. "Always Friends": $\left\{(i,j) : W_{ij}^t(1) = W_{ij}^t(0) = 1\right\}$ in medium blue,

3. "Marginal Friends": $\left\{(i,j) : W_{ij}^t(1) > W_{ij}^t(0)\right\}$ in dark blue.

Panel (a) shows homophily based on whether pairs share all of the following demographic characteristics—age bin, sex, race, faculty status, civil service, academic professional, and salary quartile—while panel (b) does the same based on the following baseline health characteristics—screening history, physical activity, gym use, marathon/10k/5k runs, and sick days. We ob-

serve that marginal friends are more similar on both demographic and health characteristics than inframarginal friends in both 2017 and 2018 (who are in turn more similar than never friends), suggesting that treatment increased network homophily.

4.3 Mechanisms

In this subsection, we outline simple model of one-sided link formation which incorporates elements from Currarini, Jackson and Pin (2009), to better organize our discussion on the mechanisms that are most likely to explain the effect of treatment on future friendship links. Consider a study participant i choosing which individuals to form/maintain a friendship link with at some time $t > 2016$ after the baseline year. Given the size of the university, study participant i likely only observes a subset $\mathcal{O}_i \subset \mathcal{N}$ of coworkers, and we assume that i can only form friendships with coworkers $j \in \mathcal{O}_i$. We model the formation of i's consideration set $\mathcal{O}_i$ as (Abaluck and Adams-Prassl, 2021):

$$Pr(j \in \mathcal{O}_i) = \begin{cases} \Lambda\Big(\zeta_1 T_j + \zeta_2 T_i T_j + \lambda(S_{ij})\Big), & W_{ij}^{2016} = 0, \\ 1, & W_{ij}^{2016} = 1, \end{cases}$$

where we assume coworkers who are friends at baseline are always in the future consideration set. The parameter ζ_1 captures chance meetings induced by j's treatment, while the ζ_2 captures the possibility of i and j meeting during health screenings or wellness activities. The function $\lambda(S_{ij})$ is increasing in the similarity between i and j, representing individuals of more similar background being more likely to meet (e.g., due to similarities in travel patterns), or more likely to notice each other.

Conditional on j being in i's consideration set, we assume the utility that i derives from forming a friendship with j is:

$$U_{ij} = \xi_1 T_j + \xi_2 T_i T_j + \xi_3 W_{ij}^{2016} + l(S_{ij}) + \upsilon_{ij},$$

and that i forms a friendship link with j if and only if U_{ij} exceeds some threshold c_i. The parameter ξ_1 captures the possibility that j being treated induces changes in traits or behavior that make it more likely for others to list them as a future friend,[10] while ξ_2 captures the possibility of shared experiences between i and j during wellness activities. The parameter ξ_3 denotes the effect of shared experiences as friends in the past, while $l(S_{ij})$ is an increasing function representing individuals' preferences for interacting with coworkers of a more similar background.

The fact that we find the lack of an interaction effect between T_i and T_j in Figure 5 on probability of future friendship suggests that interactions or shared experiences during wellness activities (ζ_2 and ξ_2) are unlikely to explain the effects of treatment on future friendships. The fact that we find some evidence that treatment leads to better friendship retention in Table 5 suggests that changes in traits or behavior induced by treatment play a role in explaining the effect of treatment on network evolution ($\xi_1 > 0$), given that baseline friends are always in the consideration set $\mathcal{O}_i$ regardless of treatment. Finally, our finding that the intervention increases homophily is consistent with chance meetings induced by treatment playing a role ($\zeta_1 > 0$), since if treatment had not effect on $\mathcal{O}_i$ we would expect marginal friends to be less alike than inframarginal friends.

4.4 Dynamic Spillovers from Network Evolution

The causal effect of treatment on future friendship links that we estimate implies that our intervention not only has behavioral spillovers in the baseline year, but also dynamic spillovers arising from its effect on network evolution. To quantify the implications of these dynamic spillovers for future screening rates, we proceed in two steps. For a given treatment assignment, we first simulate the future network using estimates of equation (5) for the formation and retention of friendship links, before using our IV estimates of endogenous peer effects in

[10] For instance, we find that treatment increases future energy levels, and decreases future smoking rates. Another possibility is that coworkers become more likely to ask treated individuals questions related to health, given their participation in the wellness activities.

Table 3 to simulate future screening rates (the main difference from the earlier simulations of behavioral spillovers being that individuals in the control group are eligible to participate in future screening, following our experimental design). As a benchmark for future screening rates without dynamic externalities, we simulate network evolution using the same equation but set the effect of treatment on link formation and retention to zero. For more details on the simulation procedure, see Appendix Section C.2.

Figure 7 plots simulated future screening rates as a function of the share of the population treated. The solid lines show future screening rates that incorporate dynamic spillovers from the intervention, whereas the dashed lines do not. The gray lines correspond to randomized treatment assignment, while the black lines assume targeted treatment (whereby individuals with the highest directed Katz-Bonacich centrality scores are treated first). First, we observe that under random treatment assignment, simulated future screening rates among the treated is increasing (roughly linearly) in the share of the population treated regardless of whether dynamic externalities are present, reflecting the behavioral spillovers. However, the rate of increase is greater when externality of the intervention on network evolution is accounted for. When almost the entire population is treated, the treatment effect on future screening is about 0.35 percentage points larger when dynamic externalities are accounted for relative to when they are not. We reach a similar conclusion with targeted treatment, although we observe that future screening rates among the treated is now monotonically decreasing as a function of the share of the population treated. This is because individuals who generate the most positive externalities are treated first, and unlike Figure 4, their ability to generate behavioral spillovers is not hampered by individuals in the control group being ineligible to participate in screening.

5 Conclusion

This paper provides causal evidence on peer effects and network evolution using a large-scale RCT on 4,834 employees at the University of Illinois. Combining measurements of the baseline friendship network with random assignment of treatment (which confers eligibility to participate in screenings), we find that each additional friend participating in health screenings increases own participation probability by 6.4 percentage points. These results are robust to state-of-the-art inference methods for network settings, and these behavioral spillovers amplify the direct effect of treatment on screening participation by about 6% (and potentially up to about 10% if the entire sample is treated).

In addition, leveraging subsequent measurements of the friendship network in future waves, we provide rare causal estimates of network evolution, a topic there is a large theoretical literature yet limited empirical evidence. We address potential concerns over differential attrition by including individual fixed effects in dyadic regressions (i.e., at the pair level). Intuitively, within-individual comparisons of whether she is more likely to list coworkers from the treatment or control group as friends in the future are uncontaminated by all observable and unobservable determinants of her decision whether to respond to future surveys. We find that treatment increases probability of being listed by coworkers as a friend in the future, with these effects coming from both new link formation, and better link retention. Simulations show that this dynamic externality of our intervention—by inducing additional future friendship links—amplifies future screening rates by about 7% the amount that behavioral spillovers do. Finally, by characterizing coworker pairs that are marginal and inframarginal friends—whose future friendship status does, or respectively does not depend on baseline treatment—we find that our intervention increased homophily of the friendship network.

This paper has several broader lessons that extend beyond our specific empirical setting. First, our finding that interventions can have dynamic externalities through their effects on the friendship network has implications for optimal targeting. Most targeted interventions focus on characteristics of the baseline network (Ballester, Calvó-Armengol and Zenou, 2006;

Galeotti and Goyal, 2010; Banerjee et al., 2019). Better knowledge of how interventions may influence the network opens the door to more sophisticated targeting policies, e.g., prioritizing treatment not solely based on individuals' current centrality in the network, but also based on the propensity of treatment to induce additional friendship links (for said individuals). Second, relatedly but on a more methodological point, these results imply that studies of peer effects should measure (and use) the baseline network, rather than the network at some future point in time (given the potential for interventions to influence network evolution).

Third, we develop econometric methods for estimating the causal effect of an intervention on network evolution and homophily while accounting for the potential for differential attrition. In particular, the directed nature of network links in our setting allow us to address differential attrition by estimating dyadic regressions with individual fixed effects. Moreover, under a monotonicity assumption, we show that methods in Abadie (2003) can be used to estimate the causal effect on homophily, by characterizing coworker pairs that are marginal and inframarginal friends.

Taken together, our work shows that in addition to behavioral spillovers, interventions can also generate dynamic externalities by causally influencing the network structure. This opens up a rich set of research questions, including better understanding the underlying mechanisms generating these dynamic spillovers, designing optimal targeting rules that account for network evolution, and the development of econometric and structural modeling tools for studying the long-run effects of interventions on network structure and outcomes.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics 113*, 231–263.

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Abaluck, J. and A. Adams-Prassl (2021). What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Economics 136*(3), 1611–1663.

Alsan, M., M. Durvasula, H. Gupta, J. Schwartzstein, and H. Williams (2024). Representation and extrapolation: Evidence from clinical trials. *The Quarterly Journal of Economics 139*(1), 575–635.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1229.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. wanted: The key player. *Econometrica 74*(5), 1403–1417.

Bandiera, O., I. Barankay, and I. Rasul (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics 120*(3), 917–962.

Bandiera, O., I. Barankay, and I. Rasul (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica 77*(4), 1047–1094.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies 77*(2), 417–458.

Banerjee, A., E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan (2024). Changes in social network structure in response to exposure to formal credit markets. *The Review of Economic Studies 91*(3), 1331–1372.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies 86*(6), 2453–2490.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Bergeron, A., P. Bessone, J. K. Kabeya, G. Z. Tourek, and J. L. Weigel (2022). Optimal assignment of bureaucrats: Evidence from randomly assigned tax collectors in the drc. Technical report, NBER Working Paper W30413.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology 92*(5), 1170–1182.

Bramoullé, Y., R. Kranton, and M. D'Amours (2014). Strategic interaction and networks. *American Economic Review 104*(3), 898–930.

Cai, Y. (2024). A modified randomization test for the level of clustering. *Journal of Business & Economic Statistics 42*(3), 933–945.

Cameron, C. A., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review 102*(6), 2981–3003.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Cullen, Z. and R. Perez-Truglia (2023). The old boys' club: Schmoozing and the gender gap. *American Economic Review 113*(7), 1703–1740.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Fairlie, R. W., F. Hoffmann, and P. Oreopoulos (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review 104*(8), 2567–2591.

Galeotti, A., B. Golub, and S. Goyal (2020). Targeting interventions in networks. *Econometrica 88*(6), 2445–2471.

Galeotti, A. and S. Goyal (2010). The law of the few. *American Economic Review 100*(4), 1468–1492.

Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics 132*(3), 1219–1260.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature 55*(1), 49–95.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*, 44–74.

Jones, D., D. Molitor, and J. Reif (2019). What do workplace wellness programs do? evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics 134*(4), 1747–1791.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika 18*(1), 39–43.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies 76*(3), 1071–1102.

Leung, M. P. (2023). Network cluster-robust inference. *Econometrica 91*(2), 641–667.

MacKinnon, J. G. and M. D. Webb (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics 32*(2), 233–254.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–145.

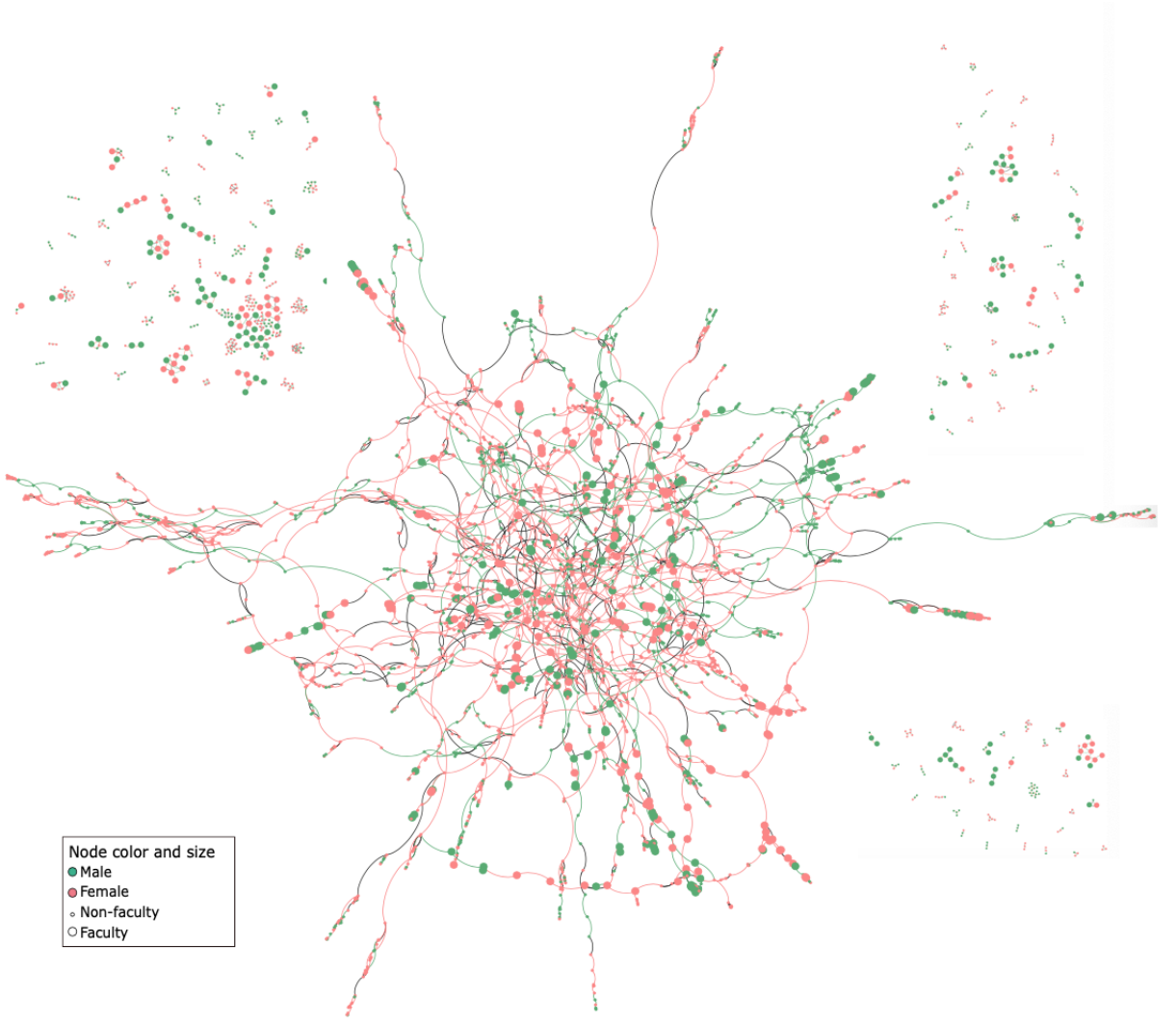
Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Sarsons, H. (2022). Interpreting signals in the labor market: evidence from medical referrals. *Working paper*.

Šíma, J. and S. E. Schaeffer (2006). On the np-completeness of some graph cluster measures. In *International Conference on Current Trends in Theory and Practice of Computer Science*, pp. 530–537. Springer.

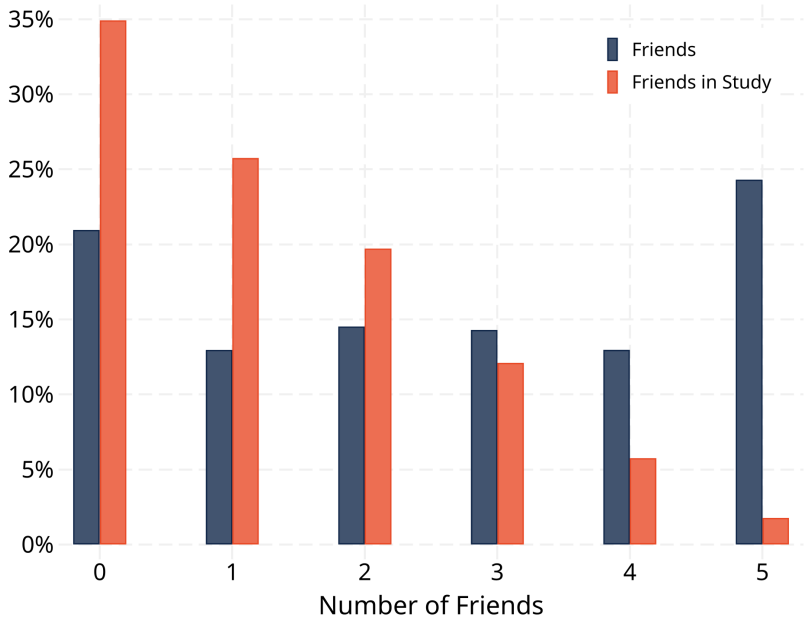
Ye, H. and J. Yi (2023). Patient-physician race concordance, physician decisions, and patient outcomes. *Review of Economics and Statistics 105*(4), 766–779.

Figure 1: Graph of Social Network



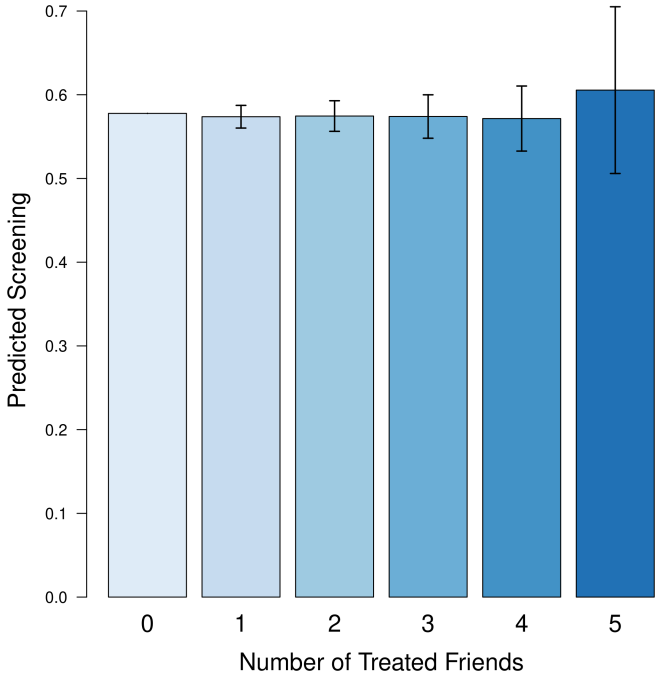
Notes: This figure shows a graph of the friendship network in 2016. Male study participants are represented by green circles while female study participants are represented by pink circles. Large circles correspond to study participants who are faculty, whereas small circles correspond to study participants who are non-faculty.

Figure 2: Number of Friends Reported on the 2016 Baseline Survey



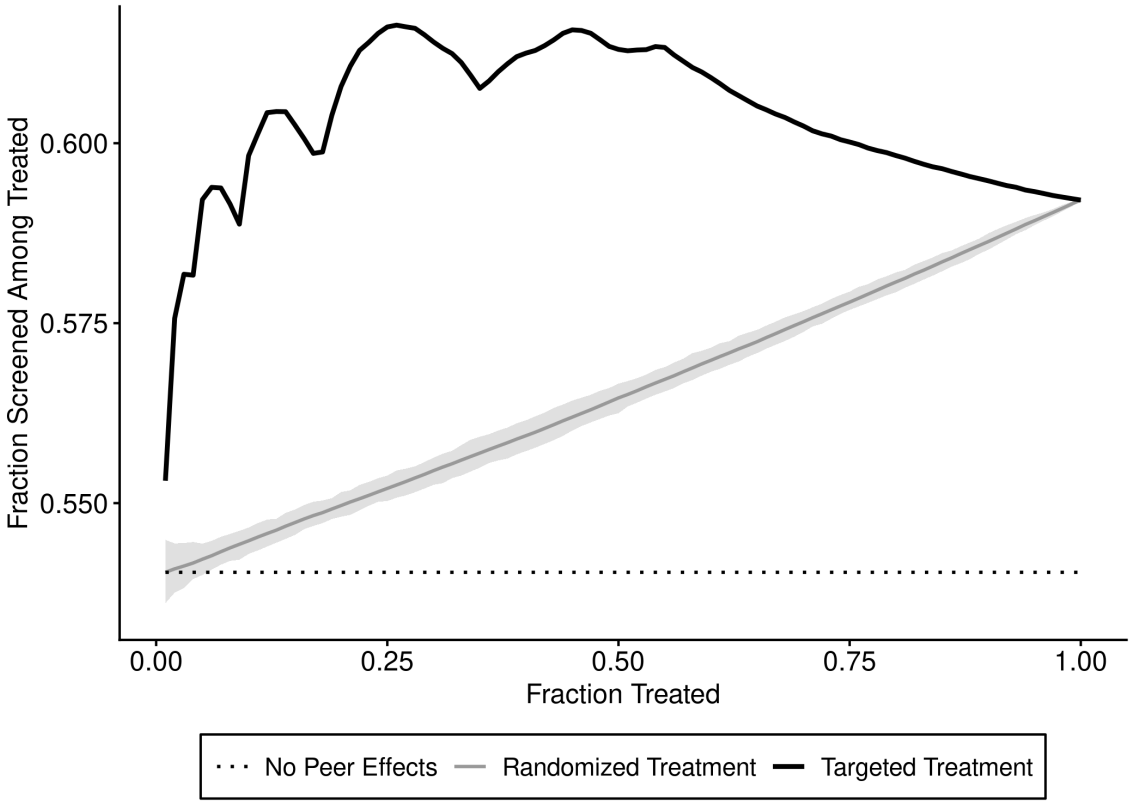
Notes: This figure shows a histogram of the number of friends (in dark blue) and the number of friends in the study (in orange) listed by each study participant in 2016.

Figure 3: Predicted Screening Rate by Number of Friends Treated



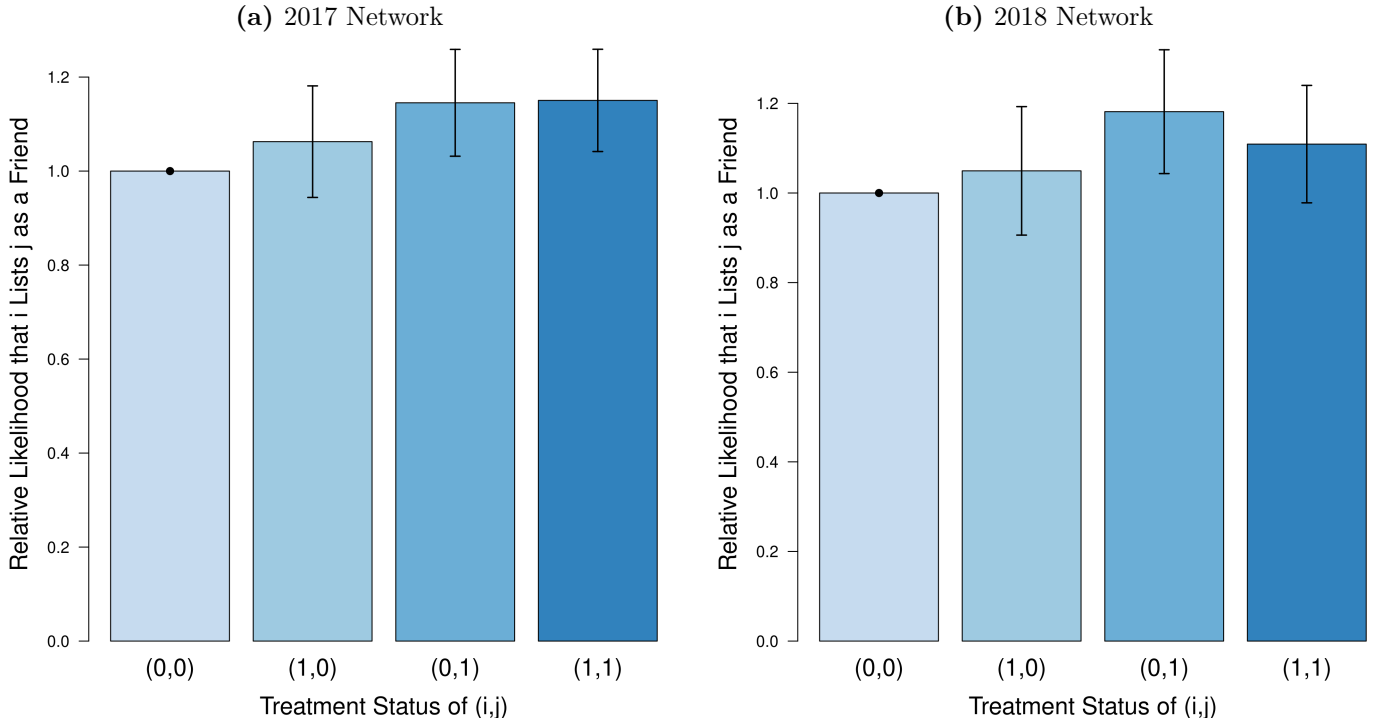
Notes: This figure shows the predicted screening rate as a function of number of treated friends for treated individuals, controlling for the number of friends in the study. Predicted screening rate is defined as the fitted values from a regression of screening in 2016 on baseline stratification and health variables, as well as financial incentives for screening.

Figure 4: Simulated Screening Rates in the Presence/Absence of Peer Effects



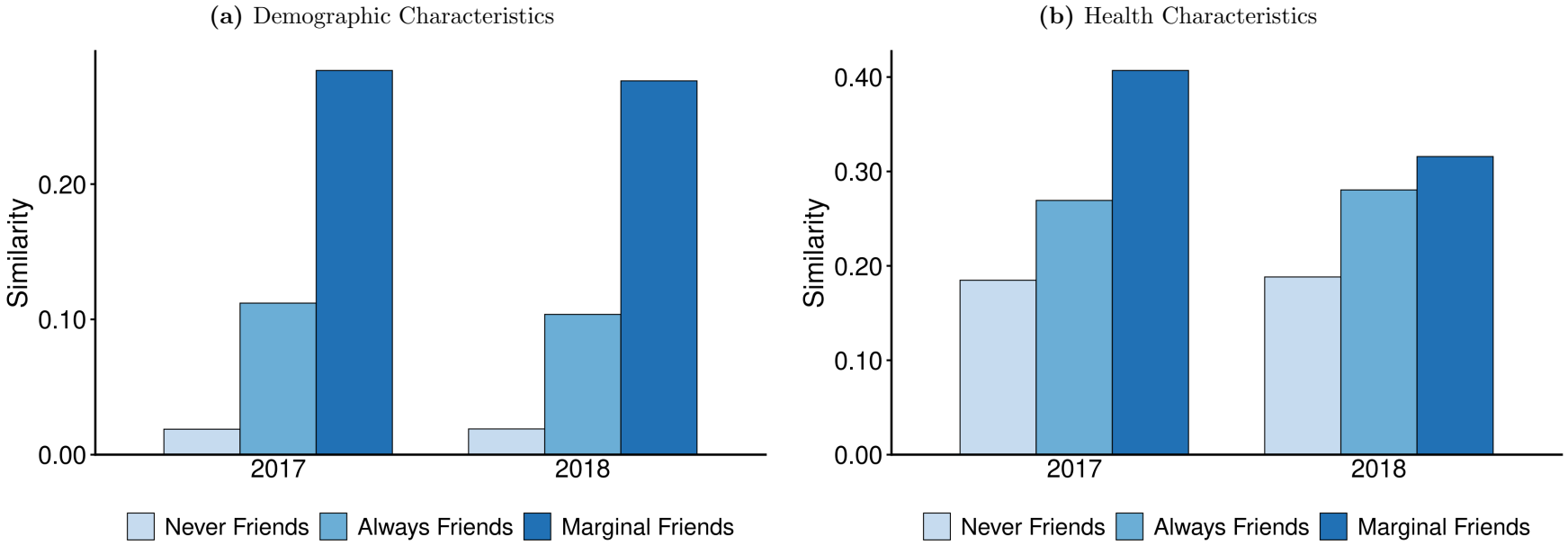
Notes: This figure shows simulated screening rates among individuals in the treatment group as a function of the share of the population treated. The dotted horizontal line shows simulated screening rates in the absence of peer effects, while the gray and black lines show simulated screening rates under the main IV estimates of peer effects estimates in Table 3, where treatment is assigned randomly, or where individuals who are most central to the network (as defined by a modification of the Katz-Bonacich centrality measure for networks with directed links). The shaded bands represent 95% pointwise confidence intervals based on 100 replications of treatment assignment (for each share of the population treated).

Figure 5: Future Probability of Friendship Link by Treatment Status



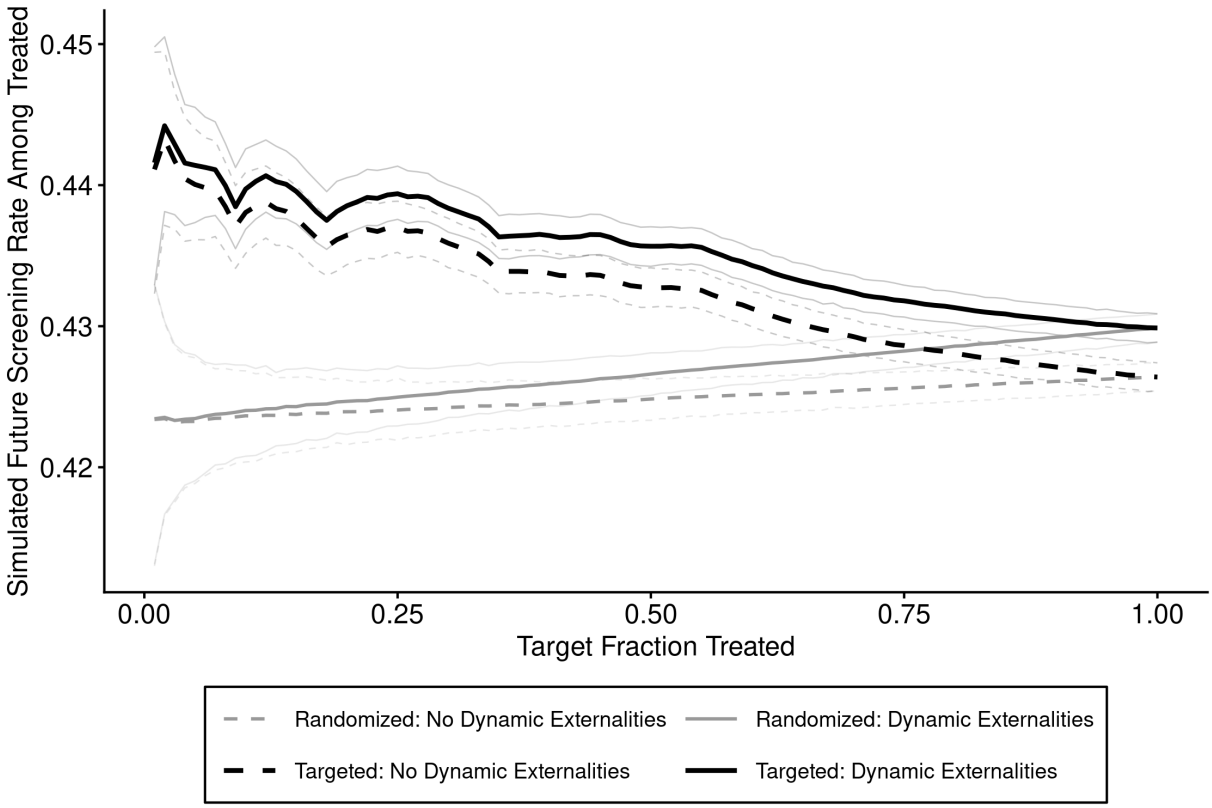
Notes: This figure shows the probability of a future friendship link for friend pairs (i, j) by the treatment status of i and j, relative to the group where neither i nor j are treated. The unit of analysis is a friend pair (i, j), and the samples for panels (a) and (b) are limited to pairs where i completed the 2017 and 2018 surveys respectively. The probability of a friendship link for the group where neither i nor j are treated is 0.00028 in 2017, and 0.00022 in 2018.

Figure 6: Homophily Among Never Friends, Inframarginal Friends, and Marginal Friends in 2017 and 2018



Notes: This figure shows the probability that pairs (i, j) share the exact same demographic or health characteristics, based on whether i and j are never friends, inframarginal friends, or marginal friends. Never friends in year t are defined as pairs (i, j) where i would not list j as a friend in year t regardless of j's treatment (in 2016). Inframarginal friends in year t are defined as pairs (i, j) where i would list j as a friend in year t regardless of j's treatment. Marginal friends in year t are defined as pairs (i, j) where i would list j as a friend if and only if j was treated. The demographic characteristics included in panel (a) include indicators for age bin, gender, race (White , Black, and other), faculty, academic professional, civil service, and salary quartiles. The health characteristics included in panel (b) include indicators for whether individuals were ever screened prior at baseline, whether they were physically active (at baseline), whether they went to the gym in the year before the start of the study, whether they participated in the IL marathon/10k/5k runs in any of the three years leading up to the study, and whether they took any sick leave in the year leading up to the study.

Figure 7: Simulated Future Screening Rates in the Presence/Absence of Dynamic Spillovers



Notes: This figure shows simulated future screening rates among individuals in the treatment group as a function of the share of the population treated. Evolution of the network from the baseline year to the next that incorporate dynamic spillovers from the intervention are shown in the (black and gray) solid lines, whereas network evolution that does not incorporate these spillovers are shown using dashed lines. To simulate network evolution while accounting for dynamic spillovers, we use estimates from dyadic regressions separately for new link formation and old link retention (based on the subsamples of coworker pairs who were not friends in the baseline year, or who were friends in the baseline year respectively), whereas to simulate evolution without dynamic spillovers, we set the coefficient on treatment to zero in these regressions. After simulating the future network, we use this new network in conjunction with the main IV estimates of peer effects in Table 3 to simulate screening rates. The gray (solid and dashed lines) assume that treatment is assigned randomly, whereas the black lines assume treatment is assigned first to individuals with the highest "directed Katz-Bonacich centrality measure". The thin solid and dashed lines represent 95% pointwise confidence intervals based on 100 replications of treatment assignment (for each share of the population treated).

Table 1: Balance in Baseline Characteristics

	Conditional Association with Number of Treated Friends (1)	p-Value (2)
A. Demographic variables		
Male	0.005	0.738
Age 50+	–0.007	0.629
Age 37–49	–0.015	0.328
White	–0.011	0.334
Salary Q1 (bottom quartile)	0.025	0.076
Salary Q2	0.008	0.568
Salary Q3	–0.010	0.480
Faculty	–0.026	0.018
Academic staff	0.023	0.150
B. Survey variables		
Ever screened	–0.019	0.033
Physically active	–0.026	0.092
Trying to be active	–0.011	0.373
Chronic condition	–0.023	0.104
Excellent or very good health	–0.007	0.662
Not poor health	0.001	0.531
Physical problems	0.020	0.202
Lots of energy	–0.016	0.270
Bad emotional health	0.003	0.851
Overweight	0.013	0.431
High BP/cholesterol/glucose	–0.017	0.252
IL Marathon/10K/5K (2014–2016)	0.003	0.814
Any campus gym visit	0.004	0.730
Any sick leave	0.023	0.071
Sample size	3,300	
Joint test for all variables (p-value)		0.167

Each row reports the coefficient estimate and corresponding p-value from a regression of a baseline characteristic of treatment-group participants on the number of treated friends, controlling for fixed effects for the number of friends in the study. The final row reports the p-value from an F-test based on a regression of the number of treated friends among treatment-group participants on all the characteristics listed in the table, controlling for fixed effects for the number of friends in the study.

Table 2: Homophily in Friend Characteristics

	Association with Average Friend Characteristics (1)	p-Value (2)
A. Demographic variables		
Male	0.570	0.000
Age 50+	0.341	0.000
Age 37–49	0.171	0.000
White	0.512	0.000
Salary Q1 (bottom quartile)	0.555	0.000
Salary Q2	0.277	0.000
Salary Q3	0.196	0.000
Faculty	0.858	0.000
Academic staff	0.668	0.000
B. Survey variables		
Ever screened	0.178	0.000
Physically active	0.147	0.000
Trying to be active	0.054	0.023
Chronic condition	0.082	0.000
Excellent or very good health	0.071	0.002
Not poor health	–0.018	0.000
Physical problems	0.026	0.246
Lots of energy	0.081	0.000
Bad emotional health	0.027	0.257
Overweight	0.130	0.000
High BP/cholesterol/glucose	0.064	0.005
IL Marathon/10K/5K (2014–2016)	0.213	0.000
Any campus gym visit	0.245	0.000
Any sick leave	0.501	0.000
Sample size	3,146	

Each row reports the coefficient estimate and p-value from a regression of a study participant's baseline characteristic on the average value of the same characteristic among their friends. Standard errors are robust to heteroskedasticity.

Table 3: Peer Effects on Health Screening and Wellness Activities

A. Reduced-form and first-stage estimates	Health Screening (Reduced Form) (1)	Friends Screened (First Stage) (2)	Wellness Activities (Reduced Form) (3)	Friends in Activities (First Stage) (4)
Number of Treated Friends	0.042***	0.667***	0.026	0.420***
	(0.014)	(0.047)	(0.016)	(0.043)
Dependent Variable Mean	0.576	0.559	0.314	0.327
FE for Number of Friends in Study	X	X	X	X
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.009	0.616	0.007	0.402
B. IV and OLS estimates	Health Screening (IV) (1)	Health Screening (OLS) (2)	Wellness Activities (IV) (3)	Wellness Activities (OLS) (4)
Number of Screened Friends	0.064***	0.143***		
	(0.019)	(0.019)		
Number of Friends in Wellness Activities			0.062*	0.115***
			(0.034)	(0.017)
Dependent Variable Mean	0.576	0.576	0.314	0.314
FE for Number of Friends in Study	X	X	X	X
First-Stage F-Statistic	204.8	—	95.8	—
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.029	0.040	0.020	0.024

This table reports estimates of the effect of the number of screened friends on health screening and the effect of the number of friends participating in wellness activities on participation in wellness activities. Panel A reports reduced-form and first-stage estimates; panel B reports IV and OLS estimates. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. Friends Screened and Friends in Activities denote the number of friends participating in health screenings and wellness activities, respectively. The sample is restricted to individuals treated in the baseline year. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 4: Effect of Treatment on Future Network

A. Effects on 2017 network

	Friend in Follow-up Year (×100)			
	(1)	(2)	(3)	(4)
Coworker Treated	0.00268**	0.00252***	0.00252***	0.00252***
	(0.00106)	(0.000820)	(0.000820)	(0.000820)
Treated	0.000588	0.000327	0.000341	
	(0.000935)	(0.000792)	(0.000794)	
Friend in 2016		0.530***	0.530***	0.530***
		(0.00828)	(0.00828)	(0.00828)
Dependent Variable Mean	0.028	0.028	0.028	0.028
Fixed Effects for Number of Friends and Number of Treated Friends			X	
Individual Fixed Effects				X
Number of Observations	17,239,311	17,239,311	17,239,311	17,239,311
R^2	0.000	0.284	0.284	0.285

B. Effects on 2018 network

	(1)	(2)	(3)	(4)
Coworker Treated	0.00197*	0.00191**	0.00191**	0.00191**
	(0.00101)	(0.000854)	(0.000854)	(0.000854)
Treated	–0.000667	–0.00114	–0.00126	
	(0.000909)	(0.000824)	(0.000825)	
Friend in 2016		0.380***	0.380***	0.380***
		(0.00892)	(0.00892)	(0.00892)
Dependent Variable Mean	0.022	0.022	0.022	0.022
Fixed Effects for Number of Friends and Number of Treated Friends			X	
Individual Fixed Effects				X
Number of Observations	14,595,660	14,595,660	14,595,660	14,595,660
R^2	0.000	0.193	0.193	0.193

This table reports regression estimates of the effect of treatment on future network structure. The dependent variable is an indicator for the individual listing the coworker as a friend in the follow-up year. The 2017 and 2018 samples are restricted to individuals who responded to the corresponding follow-up survey, although the set of coworkers is not restricted to survey respondents. The dependent variables and the baseline probability of a friendship link are multiplied by 100, so coefficients on the treatment indicators are expressed in percentage points. Observations are at the individual-coworker-pair level. Standard errors two-way clustered by respondent and coworker are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 5: Effect of Treatment on New Link Formation and Link Retention

A. Effects on 2017 network	New Coworker (×100)		Retained Coworker (×100)	
	(1)	(2)	(3)	(4)
Coworker Treated	0.00123*	0.00141**	5.04**	5.36***
	(0.000640)	(0.000679)	(2.05)	(2.06)
Baseline Probability of Friendship Link (×100)		0.386***		1.78***
		(0.0173)		(0.362)
Dependent Variable Mean	0.013	0.013	53.0	53.0
Individual Fixed Effects	X	X	X	X
Number of Observations	17,234,404	17,234,404	3,982	3,982
R^2	0.000	0.010	0.430	0.437
B. Effects on 2018 network				
	(1)	(2)	(3)	(4)
Coworker Treated	0.00181***	0.00194***	1.03	1.19
	(0.000633)	(0.000659)	(2.02)	(2.02)
Baseline Probability of Friendship Link (×100)		0.300***		0.796**
		(0.0156)		(0.342)
Dependent Variable Mean	0.011	0.011	38.0	38.0
Individual Fixed Effects	X	X	X	X
Number of Observations	14,591,422	14,591,422	3,443	3,443
R^2	0.000	0.007	0.457	0.458

This table reports regression estimates of the effect of a coworker's treatment assignment on new friendship-link formation (columns 1 and 2, pairs not linked at baseline) and the retention of existing friendship links (columns 3 and 4, pairs linked at baseline). The 2017 and 2018 samples are restricted to individuals who responded to the corresponding follow-up survey, although the set of coworkers is not restricted to survey respondents. The dependent variables and the baseline probability of a friendship link are multiplied by 100, so coefficients on the treatment indicators are expressed in percentage points. All specifications include individual fixed effects. Observations are at the individual-coworker-pair level. Standard errors two-way clustered by respondent and coworker are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 6: Effect of Treatment on Future Network: Treatment Effect Heterogeneity

A. Effects on 2017 network			
	Friend (×100)	New Friend (×100)	Retained Friend (×100)
	(1)	(2)	(3)
Coworker Treated	–0.000335	–0.0000421	3.53
	(0.000811)	(0.000604)	(2.73)
Friend in 2016	0.515***		
	(0.00825)		
Baseline Probability of Friendship Link (×100)	0.330***	0.343***	1.72**
	(0.0395)	(0.0330)	(0.678)
Treated × Baseline Probability (×100)	0.0367	0.00947	–0.422
	(0.0380)	(0.0326)	(0.714)
Coworker Treated × Baseline Probability (×100)	0.112***	0.0549*	0.531
	(0.0396)	(0.0299)	(0.563)
Dependent Variable Mean	0.028	0.013	53.0
Individual Fixed Effects	X	X	X
Number of Observations	17,239,311	17,234,404	3,982
R^2	0.291	0.010	0.437
B. Effects on 2018 network			
	(1)	(2)	(3)
Coworker Treated	–0.000966	0.0000665	0.438
	(0.000792)	(0.000615)	(2.52)
Friend in 2016	0.369***		
	(0.00886)		
Baseline Probability of Friendship Link (×100)	0.277***	0.292***	0.242
	(0.0420)	(0.0332)	(0.675)
Treated × Baseline Probability (×100)	–0.0486	–0.0594*	0.594
	(0.0420)	(0.0330)	(0.737)
Coworker Treated × Baseline Probability (×100)	0.110***	0.0713**	0.242
	(0.0390)	(0.0284)	(0.472)
Dependent Variable Mean	0.022	0.011	38.0
Individual Fixed Effects	X	X	X
Number of Observations	14,595,660	14,591,422	3,443
R^2	0.197	0.007	0.459

This table reports pair-level regressions examining heterogeneity in the effects of treatment on subsequent friendship links. Friend in 2016 is an indicator for a friendship link between the individual and coworker in the baseline network. The 2017 and 2018 samples are restricted to individuals who responded to the corresponding follow-up survey, although the set of coworkers is not restricted to survey respondents. The dependent variables and the baseline probability of a friendship link are multiplied by 100, so coefficients on the treatment indicators are expressed in percentage points. All specifications include individual fixed effects. Observations are at the individual-coworker-pair level. Standard errors two-way clustered by respondent and coworker are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Online Appendix

A Network Cluster-Robust Inference

A.1 Defining and Assessing Quality of Clusters

The worry for using robust standard errors for inference in the peer effects setting is that errors may be correlated between individuals by virtue of peer effects. A natural solution is to use clustered standard errors: for example, if individuals working in different departments interact exclusively with other individuals in the same department, clustering standard errors by department would likely be appropriate. Unfortunately, there are no such "high-quality" clusters based on contextual knowledge in our setting, so we consider clustering theory and methods that explicitly leverage the network structure as described in Leung (2023).

The theory in Leung (2023) assumes an undirected network structure, so for inference purposes, we do the same. Specifically, given the initial directed adjacency matrix $\boldsymbol{W}$ (whose (i,j) entry is the friendship indicator W_{ij} defined in Section 3), we symmetrize it to obtain an undirected adjacency matrix $\boldsymbol{A}$, where $A_{ij} \equiv \max\{W_{ij}, W_{ji}\}$. Intuitively, if the true network structure is directed but we consider an undirected structure for inference, our standard errors may be conservative, since we are allowing for potentially more dependence between observations than there actually is.

For a particular cluster $S \subseteq \{1, ..., N\}$, we measure its "quality" of a cluster (with respect to the network structure) based on its *conductance*, defined as:

$$\phi_{\boldsymbol{A}}(S) \equiv \frac{|\partial_{\boldsymbol{A}}(S)|}{vol_{\boldsymbol{A}}(S)},$$

where the numerator is the *edge boundary size*, defined as $|\partial_{\boldsymbol{A}}(S)| \equiv \sum_{i \in S} \sum_{j \notin S} A_{ij}$, and the denominator is the volume of S, defined as $vol_{\boldsymbol{A}}(S) \equiv \sum_{i \in S} \sum_{j=1}^{N} A_{ij}$. By definition, $\phi_{\boldsymbol{A}}(S)$ lies between zero and one, and the smaller the value it takes, the higher-quality the cluster is.

Suppose we have L clusters $C_1, ..., C_L$, and consider the GMM formulation of the estimation problem, denoting the sample moment function for a particular cluster C_l by $\hat{G}_l(\theta)$, where θ is the parameter of interest (with true value θ_0). Under weak network dependence and standard regularity conditions, we have:

$$\frac{1}{\sqrt{N}} \begin{pmatrix} N_1 \hat{G}_1(\theta_0) \\ \vdots \\ N_L \hat{G}_L(\theta_0) \end{pmatrix} \to^d N(\mathbf{0}, \boldsymbol{\Sigma}^*), \boldsymbol{\Sigma}^* = \begin{pmatrix} \rho_1 \Sigma_{11} & \cdots & \sqrt{\rho_1 \rho_L} \Sigma_{1L} \\ \vdots & \ddots & \vdots \\ \sqrt{\rho_1 \rho_L} \Sigma_{1L} & \cdots & \rho_L \Sigma_{LL} \end{pmatrix},$$

where $\rho_l = \lim_{N \to \infty} N_l / N$, and $\Sigma_{lm} = \lim_{N \to \infty} Cov\left(\sqrt{N_l} \hat{G}_l(\theta_0), \sqrt{N_m} \hat{G}_m(\theta_0)\right)$. The main requirement for validity of small-L clustered standard errors is *asymptotic independence*, i.e.,

$$\Sigma_{lm} = 0 \; \forall l \neq m.$$

It turns out that the main condition required for this to hold is that the maximal conductance of the

clusters tend to zero:

$$\max_{l=1,...,L} \phi_{\boldsymbol{A}}(C_l) \to 0 \text{ as } N \to \infty.$$

This is clearly an asymptotic condition, but based on simulations, Leung (2023) suggests using clustered standard errors if the maximal conductance is less than 0.1, with a minimum number of clusters $L \geq 5$ (Cameron, Gelbach and Miller (2008); Cai (2024)).

Do the clusters chosen based on contextual knowledge satisfy this condition? It turns out that the maximal conductances if we cluster by home college, home organization, race-by-age interactions, and faculty-by-age interactions are respectively, 0.6, 1, 1, and 0.64, which suggests that these clusters are not of high enough quality.

A.2 Constructing High-Quality Clusters

A useful concept for considering optimal clusters is the *kth-order Cheeger constant of* $\boldsymbol{A}$, defined as:

$$h_k(\boldsymbol{A}) = \min\{\max_{1\leq l\leq k} \phi_{\boldsymbol{A}}(S_l) : S_1, ..., S_k \text{ partitions}\{1, ..., N\}\},$$

for $k > 1$. A necessary condition for asymptotic independence is:

$$h_L(\boldsymbol{A}) \to 0,$$

otherwise there exists no partition of the data that yields high-quality clusters. Unfortunately, the problem of finding the optimal partition and thus computing the Cheeger constant is NP-complete (ma and Schaeffer 2006). So, instead we rely on results from spectral graph theory to find approximately optimal clusters.

A key concept in spectral graph theory is the graph Laplacian, which is informative about the magnitude of the Cheeger constants and extremely useful for finding approximate optimal clusters. There exists several variants of the Laplacian, and Leung (2023) considers a version of the normed graph Laplacian, defined as:

$$\boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2},$$

where $\boldsymbol{D}$ is the degree matrix, i.e., the diagonal matrix with $D_{ii} \equiv \sum_{j=1}^{n} A_{ij}$. A practical issue with computing this in our setting is that $\boldsymbol{D}$ is singular because some individuals have no friends (or other individuals listing them as friends), so instead we use the generalized inverse.

If we order the eigenvalues of the normalized Laplacian by:

$$\lambda_1 \leq ... \leq \lambda_N,$$

we have the following relationship between the Cheeger constants and the spectrum of the Laplacian:

$$\frac{\lambda_k}{2} \leq h_k(\boldsymbol{A}) \leq C_k \lambda_k^{1/2},$$

where C_k does not depend on N and is $O(k^2)$. Hence, the eigenvalues λ_k are informative about the magnitude of the Cheeger constants.

Choosing the number of clusters L entails a tradeoff: we may lack power with small L,[1] but cluster quality typically falls as we increase L. Based on simulations, Leung (2023) suggests choosing the largest value of L such that $\lambda_L \leq 0.05$, and then verifying that the resulting clusters have maximal conductance of at most 0.1. In our network, however, the eigenvalue rule is not informative: the baseline study network has 299 connected components among respondents with at least one link, so λ_L remains below 0.05 for values of L well beyond the range for which the spectral clusters described below have low conductance. We therefore choose L directly from the conductance criterion: we compute spectral clusters for each L on a grid, and use the largest L for which the maximal conductance of the clusters is below 0.1.

Next, to find high-quality clusters for a given L, we use spectral clustering. This entails applying k-means clustering to the first L eigenvectors $\{\boldsymbol{V}_l\}_{l=1}^{L}$ of the Laplacian (where the eigenvectors are ordered such that $\boldsymbol{V}_l$ is the eigenvector associated with λ_l). Specifically, we embed individuals in $\mathbb{R}^L$ space by associating with each individual i, the position:[2]

$$\boldsymbol{X}_i = \left(\frac{V_{1i}}{\left(\sum_{l=1}^{L} V_{li}^2\right)^{1/2}}, ..., \frac{V_{Li}}{\left(\sum_{l=1}^{L} V_{li}^2\right)^{1/2}} \right).$$

We then apply k-means clustering using these positions to obtain the clusters $C_1, ..., C_L$, except that for individuals listing no friend with no others listing them as a friend, we randomly assigned them to clusters of size ten.

A.3 Inference

The conductance threshold of 0.1 is based on simulations in Leung (2023), which may differ from our setting in many ways. Hence, as a robustness check, we also compute clustered standard errors using spectral clusters with different values of L, and plot the t-statistic of our main estimate, to see if the statistical significance is sensitive to this choice. Appendix Figure D.5 shows that for all of the values of L considered, the t-statistic is above the 1.96 threshold for statistical significance at the five percent level. In the paper, the number of clusters we used corresponds to the vertical dashed line in this figure, which is the largest L for which the maximal conductance of the spectral clusters is below 0.1.

B Details on Differential Attrition

The problem with differential attrition arises due to the fact that the sample of individuals who replied to the survey in 2017 are a selected sample. Hence, while treatment is independent of potential outcomes (including future links), it may not be independent of potential outcomes once one conditions on completing the 2017 survey, especially if treatment has an effect on response rates.

To describe our approach, first consider our regression specification without any fixed effects (ignoring

[1] Cameron, Gelbach and Miller (2008) and Cai (2024) suggest at least five clusters.

[2] In the definition below, if the denominator is zero, we simply define the ratio as zero.

controls other than baseline friendship for notational simplicity):

$$W_{ij}^{t} = \varphi_0 + \varphi_1 T_j + \varphi_2 T_i + \gamma W_{ij}^{2016} + e_{ij}.$$

Due to differential attrition, the coefficients φ may be different from the coefficients φ^* if we were able to estimate the regression using the full sample, including individuals who did not complete the screening:

$$W_{ij}^{t*} = \varphi_0^* + \varphi_1^* T_j + \varphi_2^* T_i + \gamma^* W_{ij}^{2016} + e_{ij}^*,$$

where we use asterisks to label quantities that can only be known if we observe the full network in 2017/2018. To reduce clutter, we omit the baseline friendship term γW_{ij}^{2016} and treat conditioning on W_{ij}^{2016} as implicit in the conditional expectations below. This control matters for the arguments in Section B.2: if survey response depends on the number of treated baseline friends, then among respondents treated coworkers are overrepresented among baseline friends, so the comparison of treated and control coworkers must be made holding baseline friendship fixed.

Our general approach to addressing attrition is related to the framework for the Lee (2009) bounds. However, the Lee bounds cannot be applied in a straightforward manner in this setting because the regressions are at the pair level, whereas the selection equation is at the individual level.

In the most general case, we can write the selection equation as:

$$Z_i^{t*} = f_i^*(T_1, ..., T_N, \omega_i)$$

for some function $f_i^* : \{0,1\}^N \times \mathbb{R} \to \mathbb{R}$, and assume that the individual completes the survey in year t if and only if:

$$Z_i \equiv \mathbb{I}[Z_i^{t*} \geq 0] = 1.$$

B.1 Regression without Fixed Effects

To illustrate intuitively the potential biases in the regression without fixed effects, suppose that f_i^* is linear in its arguments:

$$f_i^*(T_1, ..., T_N, \omega_i) = b_{i,0} + \sum_{j=1}^{N} b_{i,j} T_j + \omega_i.$$

First, consider the potential effect of attrition on the "own" treatment effect estimate (i.e., φ_2, the coefficient on T_i):

$$\begin{aligned}
\mathbb{E}[W_{ij}^{t*} | Z_i^{t*} \geq 0, T_i = 1, T_j] &= \varphi_0^* + \varphi_1^* T_j + \varphi_2^* + \mathbb{E}[e_{ij}^* | Z_i^{t*} \geq 0, T_i = 1, T_j] \\
&= \varphi_0^* + \varphi_1^* T_j + \varphi_2^* + \mathbb{E}[e_{ij}^* | b_{i,0} + b_{i,i} + \sum_{j' \neq i} b_{i,j'} T_{j'} + \omega_i \geq 0, T_j] \\
&= \varphi_0^* + \varphi_1^* T_j + \varphi_2^* + \mathbb{E}[e_{ij}^* | \omega_i \geq -b_{i,0} - b_{i,i} - \sum_{j' \neq i} b_{i,j'} T_{j'}, T_j],
\end{aligned}$$

$$\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_i = 0, T_j] = \varphi_0^* + \varphi_1^* T_j + \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - \sum_{j' \neq i} b_{i,j'} T_{j'}, T_j],$$

$$\begin{aligned}\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_i = 1, T_j] - \mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_i = 0, T_{\sim i}] = \varphi_2^* + \\ \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - b_{i,j} - \sum_{j \neq i} b_{i,j'} T_{j'}, T_j] - \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - \sum_{j' \neq i} b_{i,j'} T_{j'}, T_j].\end{aligned}$$

So, φ_2 will be biased due to "classical differential attrition" if:

- Unobserved future propensity to list others as friends is correlated with willingness to complete future surveys: $Cov(e_{ij}^*, \omega_i) \neq 0$, and;
- Treatment affects willingness to complete future surveys: $b_{i,i} \neq 0$.

Next, consider the potential effect of attrition on the "other" treatment effect estimate (i.e., φ_1, the coefficient on T_j):

$$\begin{aligned}\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_j = 1, T_i] &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|Z_i^{t*} \geq 0, T_j = 1, T_i] \\ &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \varphi_2^* + \mathbb{E}[e_{ij}^*|b_{i,0} + b_{i,i} T_i + b_{i,j} + \sum_{j' \neq i,j} b_{i,j'} T_{j'} + \omega_i \geq 0, T_i] \\ &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \varphi_2^* + \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - b_{i,i} T_i - b_{i,j} - \sum_{j' \neq i,j} b_{i,j'} T_{j'}, T_i],\end{aligned}$$

$$\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_j = 0, T_i] = \varphi_0^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - b_{i,i} T_i - \sum_{j' \neq i,j} b_{i,j'} T_{j'}, T_i].$$

$$\begin{aligned}\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_j = 1, T_i] - \mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_j = 0, T_i] = \varphi_1^* + \\ \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - b_{i,i} T_i - b_{i,j} - \sum_{j' \neq i,j} b_{i,j'} T_j, T_j] - \mathbb{E}[e_{ij}^*|\omega_i \geq -b_{i,0} - b_{i,i} T_i - \sum_{j' \neq i,j} b_{i,j'} T_{j'}, T_j].\end{aligned}$$

From this expression, we see that φ_1 will be biased due to "differential attrition in peer effects settings" if:

- Unobserved future propensity to list others as friends is correlated with willingness to complete future surveys: $Cov(e_{ij}^*, \omega_i) \neq 0$, and;
- Treatment of other individuals affects willingness to complete future surveys: $b_{i,j} \neq 0$.

B.2 Regressions with Fixed Effects

It seems hard to make progress without making some restrictions on f_i^*. One intuitive restriction we can make is to assume that f_i^* does not differ across individuals and depends on the entire vector of treatments $T_1, ..., T_n$ only through the treatment effect of i, T_i, and the number of friends of i that are treated N_i^T.

Assumption 1. $f_i^*(T_1, ..., T_n, \omega_i) = f(T_i, N_i^T, \omega_i)$ for some function f, for all i.

Under Assumption 1, if we include fixed effects for number of friends N_i and number of treated friends N_i^T (conditioning on N_i is left implicit below), we have:

$$\begin{aligned}\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_j = 1, T_i, N_i^T] &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|Z_i^{t*} \geq 0, T_j = 1, T_i, N_i^T] \\ &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|f(T_i, N_i^T, \omega_i) \geq 0, T_i, N_i^T]\end{aligned}$$

$$\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_j = 1, T_i, N_i^T] = \varphi_0^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|f(T_i, N_i^T, \omega_i) \geq 0, T_i, N_i^T]$$

Taking the difference, we see that the regression estimate of φ_1^* is unbiased.

On the other hand, if we consider the estimate of the own effect (T_i), we have:

$$\begin{aligned}\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_i = 1, T_j, N_i^T] &= \varphi_0^* + \varphi_1^* T_j + \varphi_2^* + \mathbb{E}[e_{ij}^*|Z_i^{t*} \geq 0, T_i = 1, T_j, N_i^T] \\ &= \varphi_0^* + \varphi_1^* T_j + \varphi_2^* + \mathbb{E}[e_{ij}^*|f(1, N_i^T, \omega_i) \geq 0, T_j, N_i^T].\end{aligned}$$

$$\mathbb{E}[W_{ij}^{t*}|Z_i^{t*} \geq 0, T_i = 0, T_j, N_i^T] = \varphi_0^* + \varphi_1^* T_j + \varphi_2^* + \mathbb{E}[e_{ij}^*|f(0, N_i^T, \omega_i) \geq 0, T_j, N_i^T].$$

so the regression estimate of φ_2^* may still be biased if unobserved future propensity to list others as friends (e_{ij}^*) is still correlated with willingness to complete future surveys (and treatment affects willingness to complete future surveys) conditional on number of treated friends.

One concern is that Assumption 1 is too strong. For instance, it would be violated if there are higher-order interactions (i.e., that individuals may also be affected by the behavior of friends-of-friends, or friends-of-friends-of-friends and so forth). Hence, a simpler way to identify φ_1^* is to simply include individual fixed effects, since:

$$\begin{aligned}\mathbb{E}[W_{ij}^{t}|T_j = 1, i] &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|T_j = 1, i] \\ &= \varphi_0^* + \varphi_1^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}^*|i]\end{aligned}$$

$$\mathbb{E}[W_{ij}^{t}|T_j = 0, i] = \varphi_0^* + \varphi_2^* T_i + \mathbb{E}[e_{ij}|i].$$

The second equality holds provided that, conditional on i and W_{ij}^{2016}, f_i^* does not depend on T_j in a way that is correlated with e_{ij}^*. This is satisfied whenever survey response depends on others' treatment only through the treatment status of i's baseline friends, in any functional form and allowing for heterogeneity across i, which is considerably weaker than Assumption 1. It would be violated if i's response depended on the treatment of coworkers whom i would list as friends in the future but did not list at baseline. However, there is no variation in T_i within individual i, so φ_2^* is not identified.

C Simulation Details

C.1 Simulating Behavioral Externalities Using Baseline Network

We simulate peer effects using our IV estimates of endogenous peer effects:

$$Y_i = \alpha_0 + \alpha_1 N_i^Y + \kappa_{N_i} + u_i.$$

Under the endogenous peer effects specification, all individuals' outcomes are simultaneously determined, and in the baseline year only treated individuals are eligible for screening. So, for any vector of counterfactual treatment assignment $\tilde{T}$, we need to find a fixed point in the vector of screening probabilities $\tilde{Y}(\tilde{T})$ such that for each individual i, we have:

$$\tilde{Y}_i(\tilde{T}) = \tilde{T}_i \cdot (\alpha_0 + \alpha_1 N_i^{\tilde{Y}(\tilde{T})} + \kappa_{N_i}),$$

where $N_i^{\tilde{Y}(\tilde{T})} \equiv \sum_j W_{ij}\tilde{Y}_j(\tilde{T})$ is the expected number of i's friends who participate in screening under the treatment assignment $\tilde{T}$.

We solve this system using fixed point iteration, starting from the benchmark without peer effects:

$$\tilde{Y}_i^{(0)}(\tilde{T}) = \tilde{T}_i \cdot (\alpha_0 + \kappa_{N_i}),$$

and updating according to:

$$\tilde{Y}_i^{(k+1)}(\tilde{T}) = \tilde{T}_i \cdot (\alpha_0 + \alpha_1 N_i^{\tilde{Y}^{(k)}(\tilde{T})} + \kappa_{N_i}),$$

until convergence.

C.2 Dynamic Spillovers: Simulating Network Evolution and Behavioral Externalities

To illustrate the effect of dynamic externalities on future outcomes, we compare simulated outcomes in 2017 under the endogenous peer effects specification, using counterfactual friendship networks that do or do not incorporate the effect of treatment on network evolution.

Specifically, for any counterfactual treatment assignment $\tilde{T}$, we simulate the counterfactual future network $\tilde{W}^{2017}$ that incorporates the effect of treatment on network evolution as follows. For pairs where i did not list j as a friend in 2016 (i.e., $W_{ij} = 0$), we assume that $Pr(\tilde{W}_{ij}^{2017} = 1) = \varphi_0^{new} + \varphi_1^{new}\tilde{T}_j$ based on our estimates of equation (5) with individual fixed effects on the sample where $W_{ij} = 0$. Similarly, for pairs where $W_{ij} = 1$, we assume that $Pr(\tilde{W}_{ij}^{2017} = 1) = \varphi_0^{retention} + \varphi_1^{retention}\tilde{T}_j$ based on our estimates of equation (5) with individual fixed effects on the sample where $W_{ij} = 1$. We simulate the future network that does not incorporate the effect of treatment on network evolution in the same way, but setting $\varphi_1^{new} = 0$ and $\varphi_1^{retention} = 0$.

We then simulate screening probabilities using the endogenous peer effect specification largely following the previous subsection, but with two changes. First, we use the counterfactual friendship network $\tilde{W}_{ij}^{2017}$ to define friends. Second, we allow all individuals to participate in screening regardless of whether they are

treated, following the experimental design in 2017. For simplicity, we assume that the same individuals get treated in both years, i.e., the counterfactual treatment $\tilde{T}$ that affects network evolution is also used for simulating future outcomes.

We seek to find a fixed point in the vector of screening probabilities $\tilde{Y}^{2017}(\tilde{T})$ such that for each individual i, we have:

$$\tilde{Y}_i^{2017} = \alpha_0^{2017} + \alpha_1 \tilde{N}_i^{\tilde{Y}^{2017}(\tilde{T})} + a\tilde{T}_i + \kappa_{N_i}^{2017},$$

where $\tilde{N}_i^{\tilde{Y}^{2017}} \equiv \sum_j \tilde{W}_{ij}^{2017} \tilde{Y}_j^{2017}$ is the expected number of i's friends in the counterfactual friendship network $\tilde{W}^{2017}$ that participate in screening (in 2017), and aT_i assumes that the effect of own baseline treatment (in 2016) is a level shift on the probability of participating in screening in 2017.[3]

We solve this system using fixed point iteration, starting from the benchmark without peer effects:

$$\tilde{Y}_i^{2017,(0)}(\tilde{T}) = \alpha_0^{2017} + a\tilde{T}_i + \kappa_{N_i}^{2017},$$

and updating according to:

$$\tilde{Y}_i^{2017,(k+1)}(\tilde{T}) = \alpha_0^{2017} + a\tilde{T}_i + \alpha_1 N_i^{\tilde{Y}^{2017,(k)}(\tilde{T})} + \kappa_{N_i}^{2017},$$

until convergence.

C.3 Modification of Katz-Bonacich Centrality in a Directed Network

To determine the order in which to treat individuals in order to maximize average screening rates in the presence of behavioral spillovers, we derive a simple modification of the Katz-Bonacich Centrality measure for a network with directed links (Katz, 1953; Bonacich, 1987). For simplicity, consider a linear endogenous peer effects model with constant treatment effects:

$$\boldsymbol{y} = \boldsymbol{a} + \tau \boldsymbol{t} + \beta \boldsymbol{W} \boldsymbol{y}, \tag{C.1}$$

where $\boldsymbol{y}$ is the vector of outcomes, $\boldsymbol{t}$ is the vector of treatment assignments, $\boldsymbol{a}$ includes other determinants of individuals' outcomes, and $\boldsymbol{W}$ is the adjacency matrix whose (i, j) entry is W_{ij}, the indicator for i listing j as a friend defined in Section 3.

Adopting the standard assumption in the literature that $\beta\rho(\boldsymbol{W}) < 1$ where $\rho(\boldsymbol{W})$ is the spectral radius of $\boldsymbol{W}$ (Ballester, Calvó-Armengol and Zenou, 2006; Bramoullé, Kranton and D'Amours, 2014; Galeotti, Golub and Goyal, 2020), equation (C.1) implies that the outcome vector can be written as:

$$\boldsymbol{y} = \boldsymbol{M}(\boldsymbol{a} + \tau \boldsymbol{t}), \tag{C.2}$$

where $\boldsymbol{M} \equiv (\boldsymbol{I} - \beta \boldsymbol{W})^{-1}$ and $\boldsymbol{I}$ is the identity matrix. Note that the condition $\beta\rho(\boldsymbol{W}) < 1$ is sufficient

[3] We estimate a by imposing the 2016 IV estimate of peer effects, regressing 2017 screening net of that term on treatment and fixed effects for number of friends, and then taking the intercept, treatment shift, and fixed effects from that regression.

for invertibility of $\boldsymbol{I} - \beta \boldsymbol{W}$, and also that this is satisfied in our setting given that:

$$\rho(\boldsymbol{W}) \leq \max_i \sum_j W_{ij} = \max_i N_i \leq 5,$$

which implies that $\beta\rho(\boldsymbol{W}) \leq (0.064)(5) = 0.32 < 1$.

Consider a treatment vector $\boldsymbol{t}$, and the incremental effect of adding an individual i to treatment: $\boldsymbol{t}' = \boldsymbol{t} + \boldsymbol{e}_i$. Writing outcomes $\boldsymbol{y}$ as a function of treatment, the effect on outcomes of this marginal treated individual is equal to:

$$\boldsymbol{y}(\boldsymbol{t}') - \boldsymbol{y}(\boldsymbol{t}) = \tau \boldsymbol{M}(\boldsymbol{t}' - \boldsymbol{t}) = \tau \boldsymbol{M} \boldsymbol{e}_i,$$

and hence, the effect on aggregate treatment is:

$$\sum \boldsymbol{y}(\boldsymbol{t}') - \sum \boldsymbol{y}(\boldsymbol{t}) = \tau \mathbf{1}' \boldsymbol{M} \boldsymbol{e}_i,$$

where $\mathbf{1}$ is a vector of ones. Given that τ is constant by assumption, its value does not matter as long as $\tau > 0$. Therefore, the vector which measures the incremental effect of each individual to aggregate outcomes (up to a constant factor τ) is given by:

$$\boldsymbol{s} = \boldsymbol{M}' \mathbf{1} = (\boldsymbol{I} - \beta \boldsymbol{W}')^{-1} \mathbf{1}.$$

Under the targeted treatment counterfactual, we prioritize treatment to individuals with the highest values of $\boldsymbol{s}$.

Note that strictly speaking, this is not exactly the optimal targeted treatment for maximizing outcomes in the baseline year, since individuals in the control group cannot participate in screenings or wellness activities. A model that explicitly incorporates this restriction, e.g., with the following outcomes equation:

$$\boldsymbol{y} = diag(\boldsymbol{t})(\boldsymbol{c} + \beta \boldsymbol{W} \boldsymbol{y}),$$

would be substantially more complicated given that the marginal effect of an individual being treated depends on who is already treated, making it a set selection problem. Nonetheless, Appendix Figure D.6 shows that our targeted treatment outperforms a treatment assignment heuristic that prioritizes treating individuals who are listed by the most coworkers as a friend in terms of maximizing screening rates among the treated.

In addition, note that the "oscillating pattern" for the screening rate among the treated as a function of the share of the population treated under the targeted treatment counterfactual is a feature of our setting, and can arise even under the optimal assignment rule. To see this, consider a stylized example with 4 coworkers, where coworkers A and B list each other as friends, coworkers C and D list each other as friends, and there are no additional friendship links. Assuming no additional heterogeneity (i.e., $\boldsymbol{a} = a\mathbf{1}$ in the notation above), an optimal targeted treatment order is: A, B, C, D. In this case, screening rate among the treated is lowest when only A is treated, since there are no spillovers with other coworkers still being ineligible for screening. The screening rate then rises as B is assigned to treatment, due to behavioral spillovers between A and B. Next, the average screening rate falls when C is added to treatment, since

there are no behavioral spillovers between C and any other individual, with D still being ineligible; at the same time, note that the screening rate at this point is still higher than with only A being treated, since there are currently still behavioral spillovers between A and B. Finally, the screening rate rises again when D is treated, as a result of behavioral spillovers between C and D.

While individuals from the control group can also participate in screening in 2017, we still see the same oscillating pattern for a similar reason. In the 4-coworker example above, ignoring dynamic externalities, when C is added to treatment (with A and B already treated), this generates behavioral spillovers for D's 2017 screening (unlike for 2016 screening, where D was ineligible). However, since D is untreated, behavioral spillovers onto C are muted, so C's screening propensity is lower than that of A and B (who generate mutual spillovers). Only when D is also treated is C's screening propensity the same as that of A and B, which explains the oscillating pattern. Nonetheless, the oscillations are smaller in 2017 because (when only A, B, and C are treated) D screens at the control-group rate rather than not at all, which generates some spillovers onto C.

"Peer Effects and Network Formation: Evidence from the Illinois Workplace Wellness Study"

Alden Cheng, University of Illinois and NBER
Damon Jones, University of Chicago
David Molitor, University of Illinois and NBER
Julian Reif, University of Illinois and NBER

D Appendix Figures and Tables

Figure D.1: Experimental Design

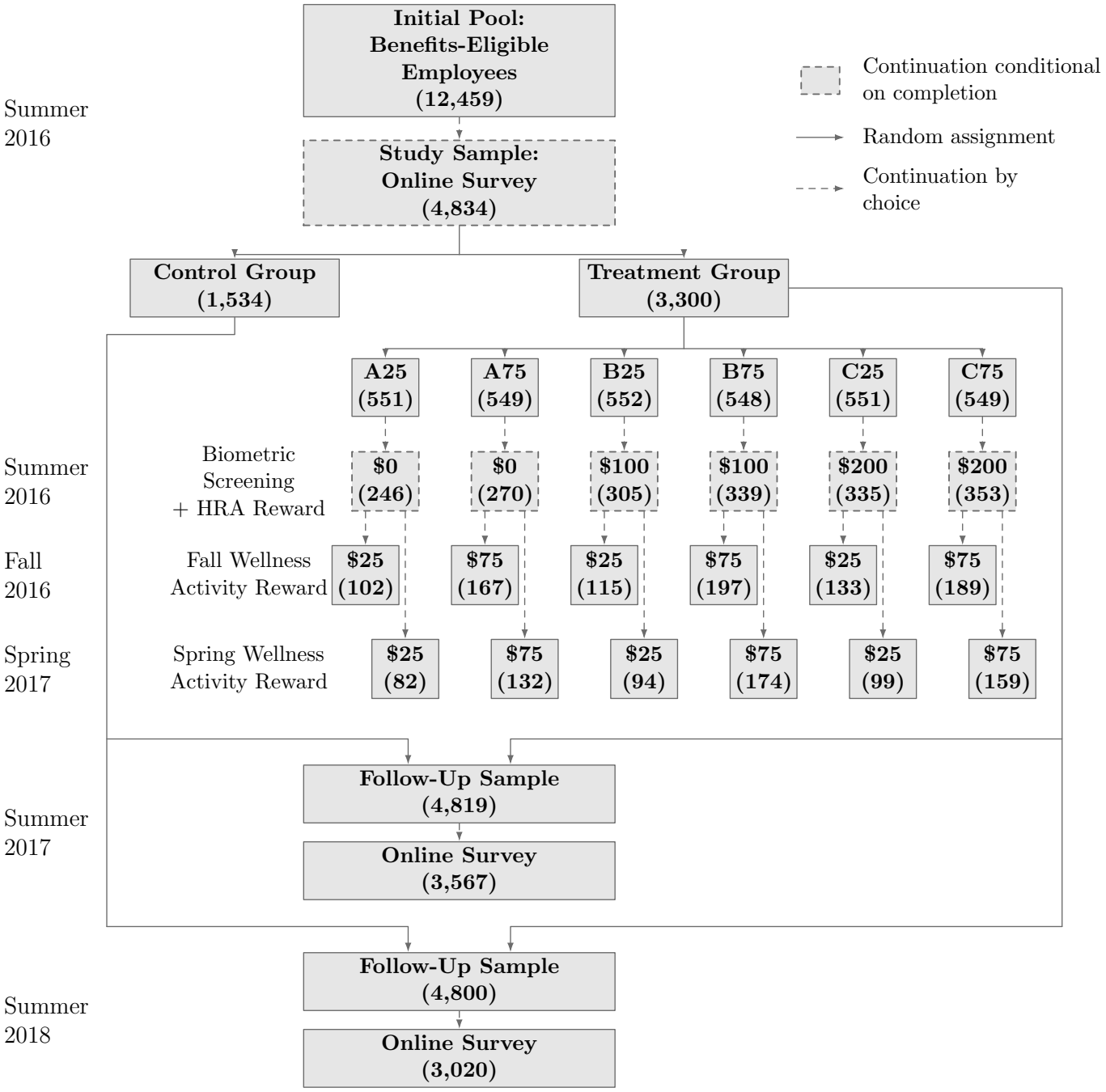


Figure D.2: Survey Question Used to Construct the Social Network

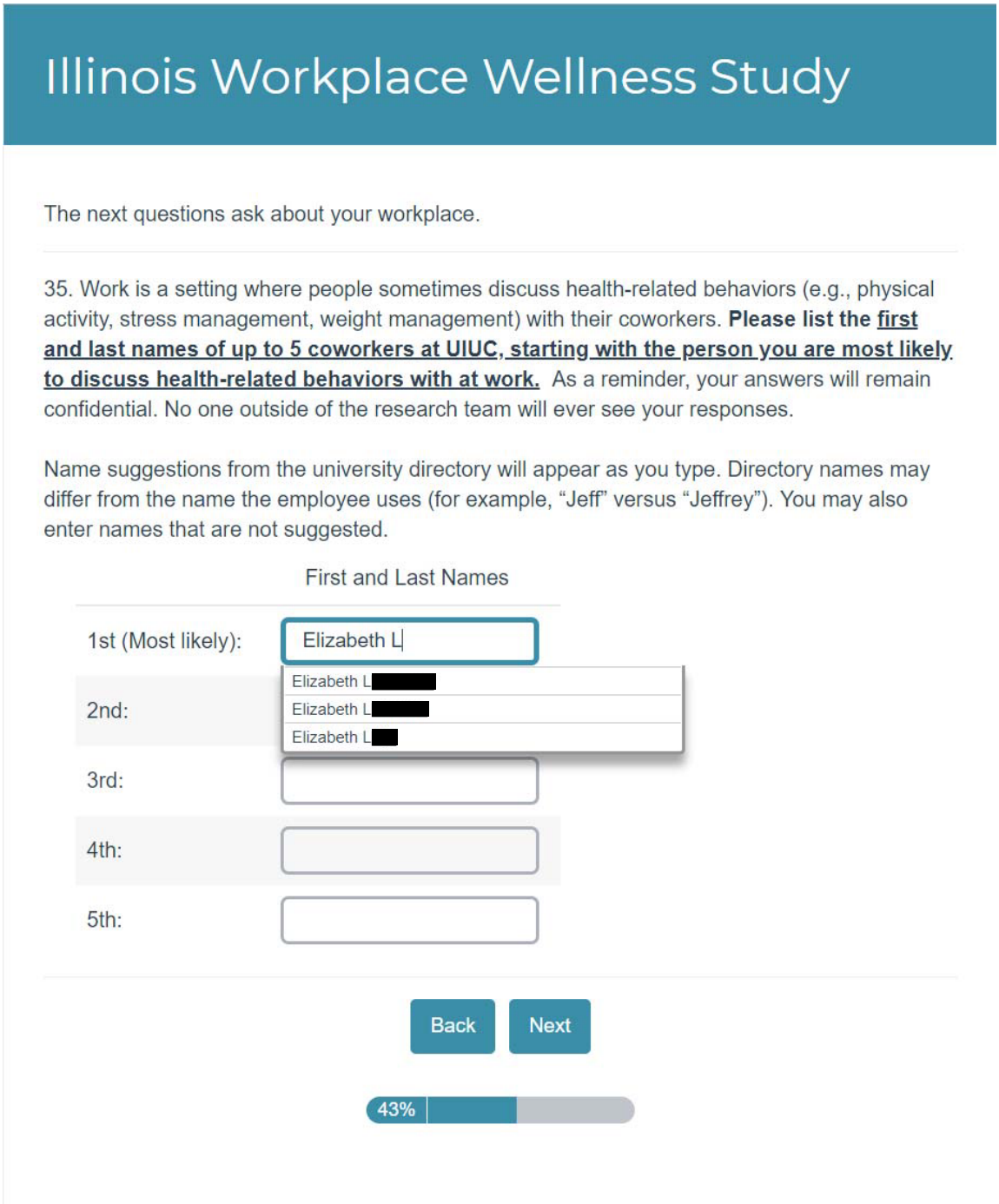
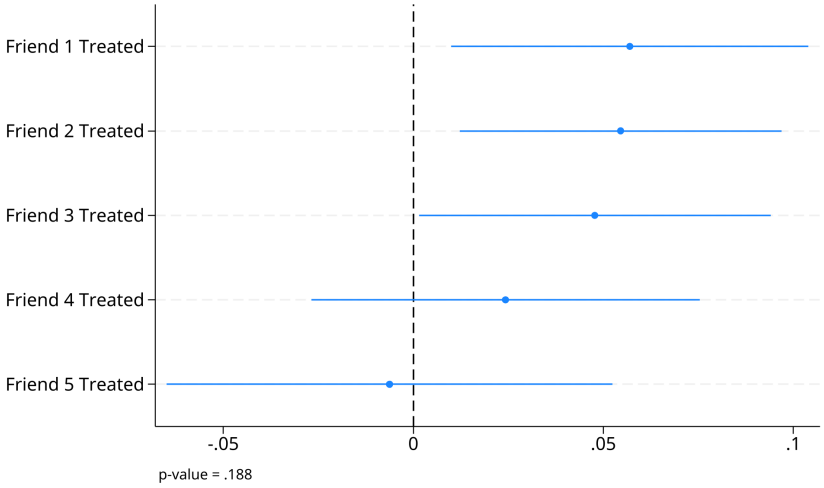


Figure D.3: Peer Effect Heterogeneity by Friend Order

(a) Health Screening



(b) Wellness Activities

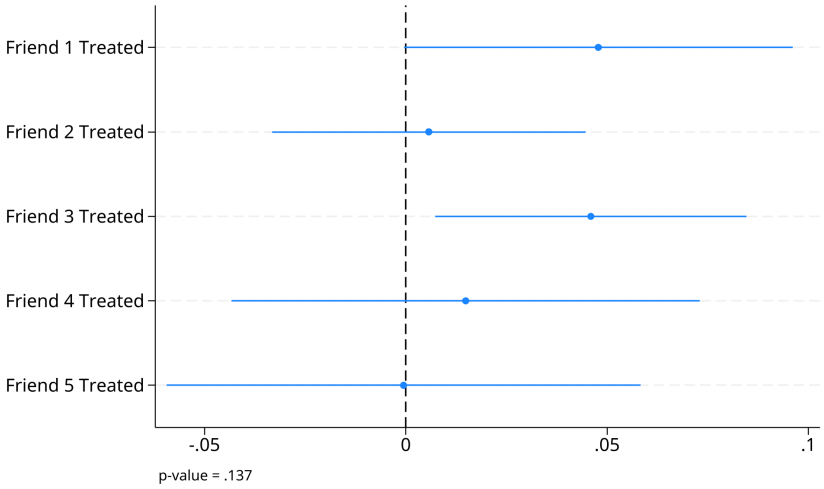
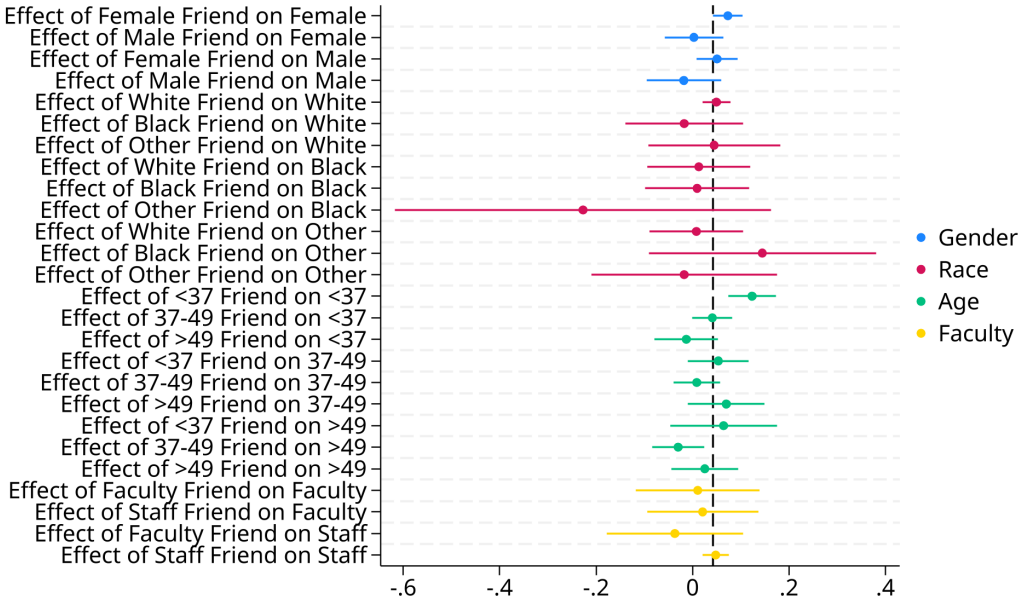


Figure D.4: Peer Effect Heterogeneity by Individual Characteristics

(a) Health Screening



(b) Wellness Activities

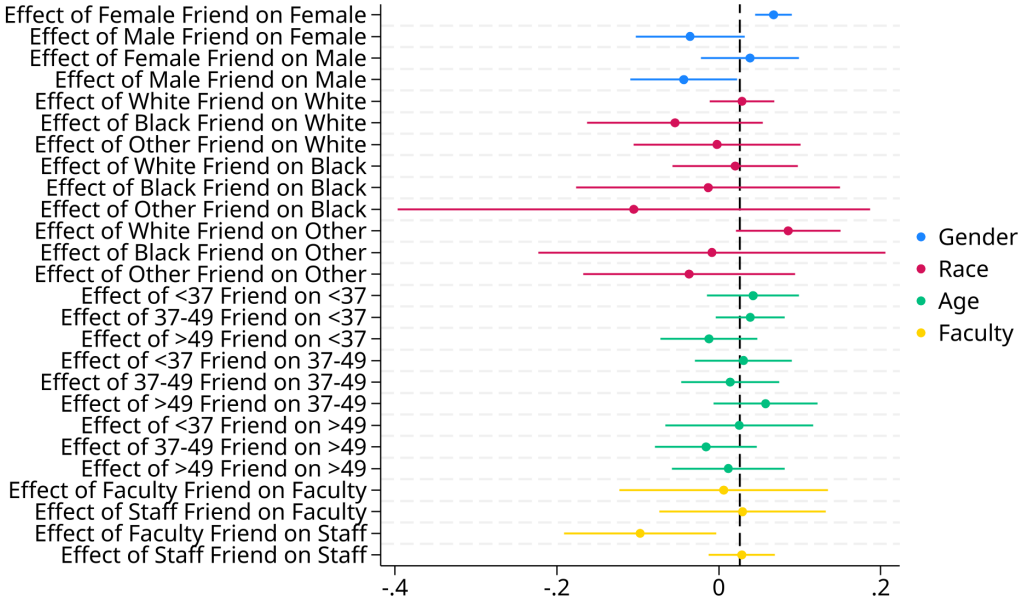


Figure D.5: t-Statistic of Main Estimate as a Function of Different Values for the Number of Clusters Used for Spectral Clustering

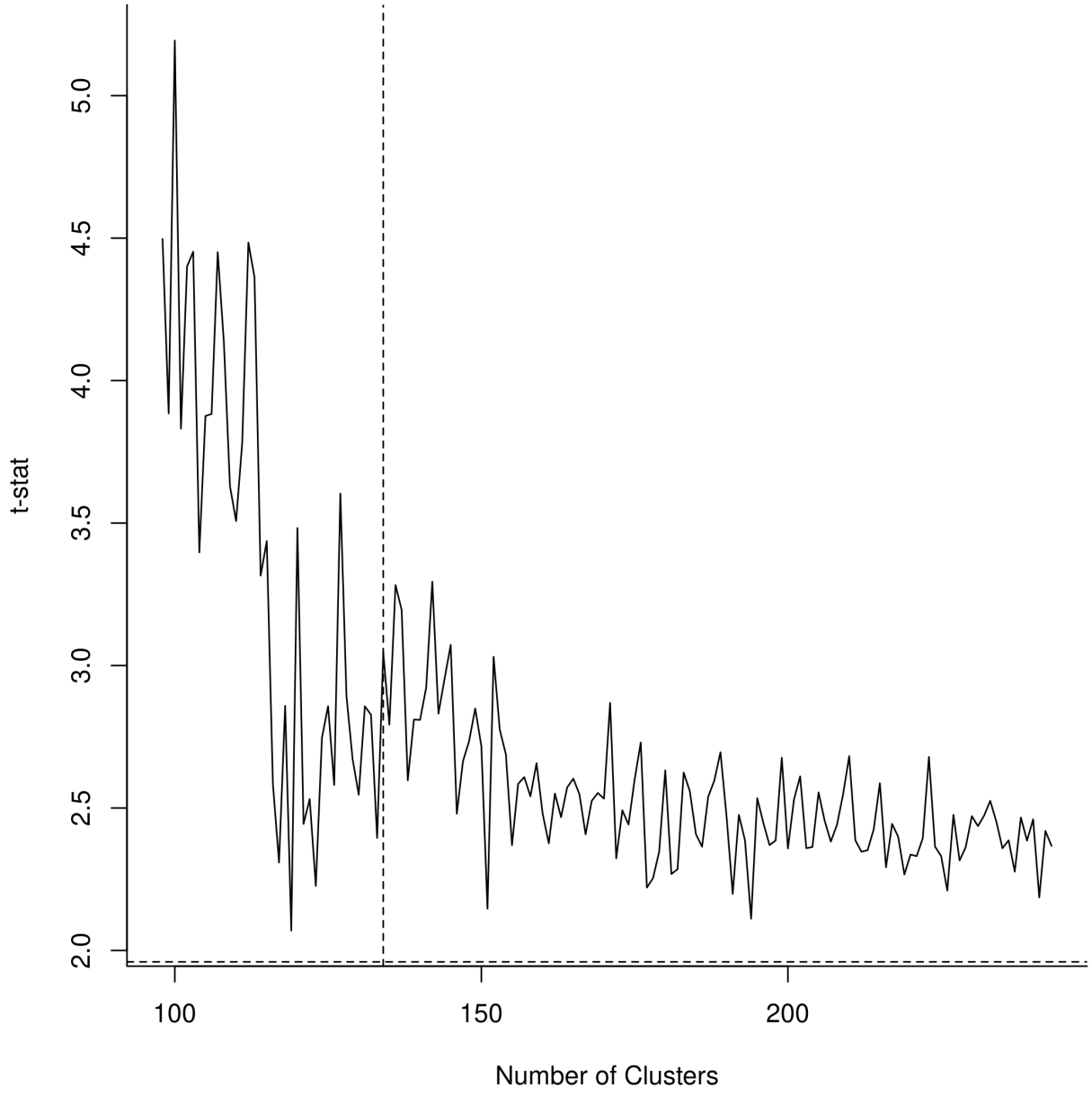
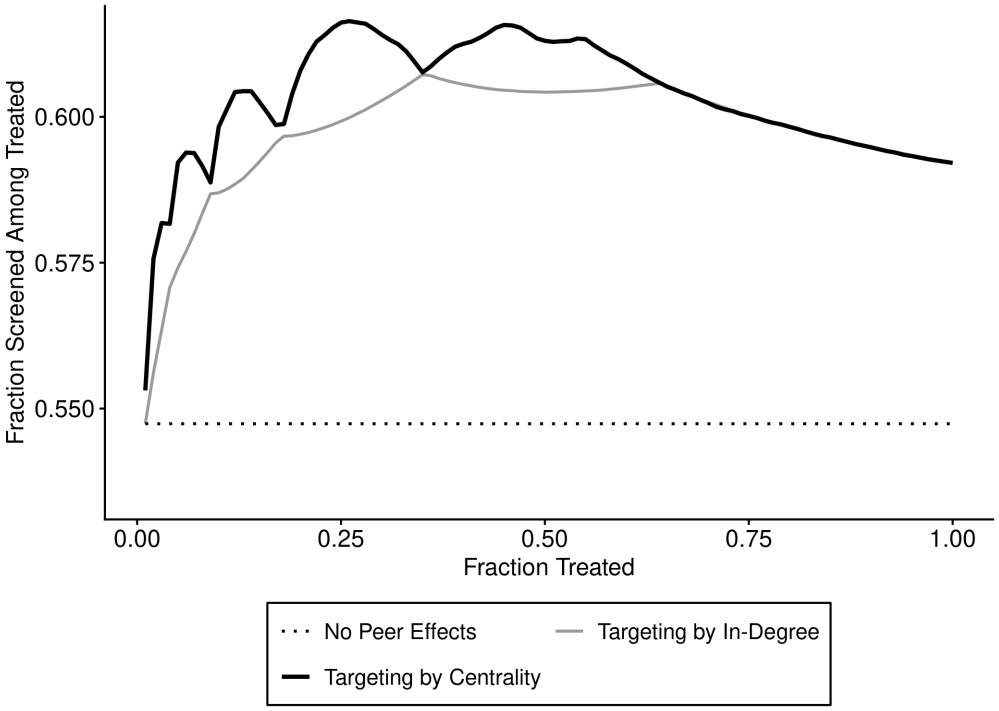
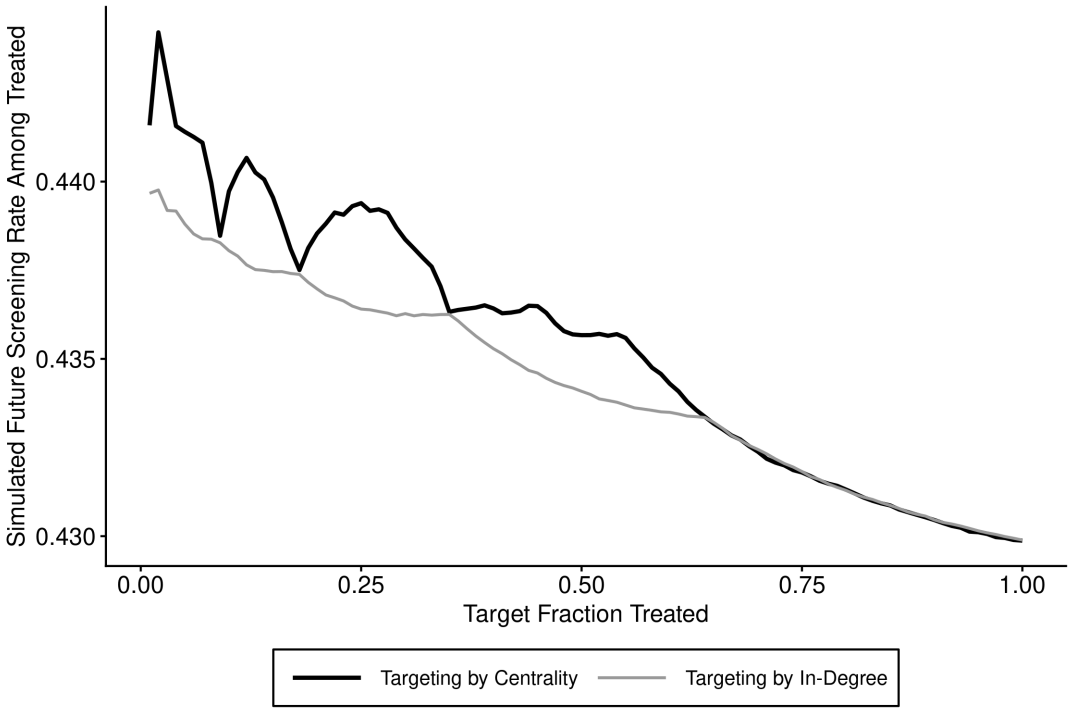


Figure D.6: Comparison of Screening Rate Among Treated Using In-Degree or Centrality Targeting

(a) Simulated Screening in 2016



(b) Simulated Screening in 2017 with Dynamic Externalities



Notes: Panel (a) shows the simulated 2016 screening rate among the treated, as a function of the share of the population treated under no peer effects, under peer effects when treatment is prioritized for individuals listed by the most coworkers as friends, and under peer effects when treatment is prioritized for individuals with the highest "directed Katz-Bonacich centrality". Panel (b) shows the simulated 2017 screening rate among the treated, as a function of the share of the population treated (in 2016) under peer effects and endogenous network evolution, when treatment is prioritized for individuals listed by the most coworkers as friends, and when treatment is prioritized for individuals with the highest "directed Katz-Bonacich centrality".

Table D.1: Observational Peer Effects with Additional Controls

A. Health screening	IV	OLS	OLS (Controls)	Post-Double Selection
	(1)	(2)	(3)	(4)
Number of Screened Friends	0.064***	0.143***	0.112***	0.134***
	(0.019)	(0.019)	(0.015)	(0.018)
Dependent Variable Mean	0.576	0.576	0.576	0.576
First-Stage F-Statistic	204.8	—	—	—
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.029	0.040	0.090	—
B. Wellness activities				
	(1)	(2)	(3)	(4)
Number of Friends in Wellness Activities	0.062*	0.115***	0.083***	0.086***
	(0.034)	(0.017)	(0.020)	(0.021)
Dependent Variable Mean	0.314	0.314	0.314	0.314
First-Stage F-Statistic	95.8	—	—	—
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.020	0.024	0.085	—

Panel A reports estimates of the effect of the number of screened friends on health screening. Panel B reports estimates of the effect of the number of friends participating in wellness activities on participation in wellness activities. The IV specifications in column 1 and the OLS specifications in column 2 control for the number of friends in the study. The OLS specifications with additional controls in column 3 include the total number of friends, the demographic and baseline health characteristics listed in the balance table, and the numbers of friends in the study with each of these characteristics. The post-double-selection specifications in column 4 select controls from the same set following Belloni, Chernozhukov, and Hansen (2014), with fixed effects for the number of friends in the study partialled out. In the IV specifications, the numbers of screened friends and friends participating in wellness activities are instrumented by the number of treated friends. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.2: Peer Effects Estimates for Friends-of-Friends

A. Health screening				
	IV	IV	OLS	OLS
	(1)	(2)	(3)	(4)
Number of Screened Friends-of-Friends	0.023	0.019	0.052***	0.035**
	(0.018)	(0.017)	(0.019)	(0.014)
Number of Screened Friends		0.060***		0.127***
		(0.018)		(0.017)
Dependent Variable Mean	0.576	0.576	0.576	0.576
First-Stage F-Statistic	468.7	106.7	—	—
FE for Number of Friends-of-Friends in Study	X	X	X	X
FE for Number of Friends in Study		X		X
Number of Observations	3,300	3,300	3,300	3,299
R^2	0.049	0.069	0.050	0.076
B. Wellness activities				
	(1)	(2)	(3)	(4)
Number of Friends-of-Friends in Wellness Activities	0.039*	0.036	0.047***	0.037***
	(0.023)	(0.023)	(0.014)	(0.013)
Number of Friends in Wellness Activities		0.054		0.097***
		(0.033)		(0.018)
Dependent Variable Mean	0.314	0.314	0.314	0.314
First-Stage F-Statistic	173.0	57.5	—	—
FE for Number of Friends-of-Friends in Study	X	X	X	X
FE for Number of Friends in Study		X		X
Number of Observations	3,300	3,300	3,300	3,299
R^2	0.059	0.069	0.059	0.071

Panel A reports estimates of the effect of the number of screened friends-of-friends on participation in health screening. Panel B reports estimates of the effect of the number of friends-of-friends participating in wellness activities on participation in wellness activities. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. In column 1, the number of participating friends-of-friends is instrumented by the number of treated friends-of-friends. In column 2, the numbers of participating friends-of-friends and participating friends are instrumented by the numbers of treated friends-of-friends and treated friends, respectively. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.3: Reduced Form Peer Effects: Other Functional Forms

	Screening			Wellness Activities		
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Treated Friends	0.042***			0.026		
	(0.014)			(0.016)		
Share of Friends Treated		0.071***			0.046*	
		(0.026)			(0.027)	
1 Friend Treated			0.039			0.040
			(0.027)			(0.033)
2 Friends Treated			0.085**			0.059
			(0.036)			(0.039)
3 Friends Treated			0.089*			0.083
			(0.047)			(0.052)
4 Friends Treated			0.184**			0.042
			(0.076)			(0.093)
5 Friends Treated			0.493***			0.282
			(0.083)			(0.198)
Dependent Variable Mean	0.576	0.576	0.576	0.314	0.314	0.314
FE for Number of Friends in Study	X	X	X	X	X	X
Number of Observations	3,300	3,300	3,300	3,300	3,300	3,300
R^2	0.009	0.008	0.010	0.007	0.007	0.007

This table reports reduced-form estimates of the effect of treated friends on health screening and participation in wellness activities. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. Share of Friends Treated is defined as the number of treated friends divided by the number of friends in the study and is set to zero for individuals with no friends in the study. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.4: Endogenous Peer Effects: Other Functional Forms

	Screening		Wellness Activities	
	IV	OLS	IV	OLS
	(1)	(2)	(3)	(4)
Share of Friends Screened	0.113***	0.242***		
	(0.040)	(0.033)		
Share of Friends Participating in Wellness Activities			0.126*	0.184***
			(0.068)	(0.030)
Dependent Variable Mean	0.576	0.576	0.314	0.314
First-Stage F-Statistic	499.1	—	182.3	—
FE for Number of Friends in Study	X	X	X	X
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.025	0.032	0.017	0.018

This table reports estimates of the effect of the share of screened friends on health screening and the effect of the share of friends participating in wellness activities on participation in wellness activities. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. Each participation share is defined as the number of participating friends divided by the number of friends in the study and is set to zero for individuals with no friends in the study. In the IV specifications, the participation share is instrumented by the share of friends in the study who were assigned to treatment. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.5: Peer Effects Estimates: Individuals Listing At Most Four Friends

A. Reduced-form and first-stage estimates	Health Screening (Reduced Form) (1)	Friends Screened (First Stage) (2)	Wellness Activities (Reduced Form) (3)	Friends in Activities (First Stage) (4)
Number of Treated Friends	0.029**	0.631***	0.016	0.389***
	(0.013)	(0.049)	(0.017)	(0.044)
Dependent Variable Mean	0.574	0.530	0.313	0.309
FE for Number of Friends in Study	X	X	X	X
Number of Observations	3,246	3,246	3,246	3,246
R^2	0.007	0.591	0.006	0.373
B. IV and OLS estimates	Health Screening (IV) (1)	Health Screening (OLS) (2)	Wellness Activities (IV) (3)	Wellness Activities (OLS) (4)
Number of Screened Friends	0.047**	0.145***		
	(0.020)	(0.023)		
Number of Friends in Wellness Activities			0.042	0.118***
			(0.041)	(0.018)
Dependent Variable Mean	0.574	0.574	0.313	0.313
FE for Number of Friends in Study	X	X	X	X
First-Stage F-Statistic	162.5	—	78.0	—
Number of Observations	3,246	3,246	3,246	3,246
R^2	0.022	0.036	0.016	0.023

This table reports estimates of the effect of the number of screened friends on health screening and the effect of the number of friends participating in wellness activities on participation in wellness activities. Panel A reports reduced-form and first-stage estimates; panel B reports IV and OLS estimates. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. Friends Screened and Friends in Activities denote the number of friends participating in health screenings and wellness activities, respectively. The sample is restricted to individuals treated in the baseline year who listed at most four friends in the study. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.6: Endogenous Peer Effects: Within-Home College/Home Organization Variation

A. Health screening				
	(1)	(2)	(3)	(4)
Number of Screened Friends	0.064***	0.065***	0.059***	0.060***
	(0.019)	(0.018)	(0.022)	(0.022)
Dependent Variable Mean	0.576	0.576	0.576	0.576
First-Stage F-Statistic	204.8	235.2	158.8	156.6
FE for Number of Friends in Study	X	X	X	X
Home-College Fixed Effects		X		X
Home-Organization Fixed Effects			X	X
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.029	0.064	0.219	0.220
B. Wellness activities				
	(1)	(2)	(3)	(4)
Number of Friends in Wellness Activities	0.062*	0.062*	0.047	0.049
	(0.034)	(0.034)	(0.041)	(0.042)
Dependent Variable Mean	0.314	0.314	0.314	0.314
First-Stage F-Statistic	95.8	95.2	63.6	61.0
FE for Number of Friends in Study	X	X	X	X
Home-College Fixed Effects		X		X
Home-Organization Fixed Effects			X	X
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.020	0.049	0.207	0.209

Panel A reports IV estimates of the effect of the number of screened friends on participation in health screening. Panel B reports IV estimates of the effect of the number of friends participating in wellness activities on participation in wellness activities. The endogenous peer participation measure is instrumented by the number of treated friends. All regressions include fixed effects for the number of friends in the study. Columns 2 and 4 include home-college fixed effects, while columns 3 and 4 include home-organization fixed effects. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.7: Peer Effects Estimates: Reversed Friendship Links

A. Reduced-form and first-stage estimates	Health Screening (Reduced Form) (1)	Friends Screened (First Stage) (2)	Wellness Activities (Reduced Form) (3)	Friends in Activities (First Stage) (4)
Number of Treated Friends	0.032**	0.652***	0.044***	0.384***
	(0.015)	(0.041)	(0.014)	(0.034)
Dependent Variable Mean	0.576	0.567	0.314	0.323
FE for Number of Friends in Study	X	X	X	X
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.010	0.676	0.010	0.468
B. IV and OLS estimates	Health Screening (IV) (1)	Health Screening (OLS) (2)	Wellness Activities (IV) (3)	Wellness Activities (OLS) (4)
Number of Screened Friends	0.049**	0.140***		
	(0.022)	(0.019)		
Number of Friends in Wellness Activities			0.115***	0.124***
			(0.029)	(0.016)
Dependent Variable Mean	0.576	0.576	0.314	0.314
FE for Number of Friends in Study	X	X	X	X
First-Stage F-Statistic	255.6	—	129.5	—
Number of Observations	3,300	3,300	3,300	3,300
R^2	0.026	0.039	0.027	0.028

This table reports estimates of the effect of the number of screened friends on health screening and the effect of the number of friends participating in wellness activities on participation in wellness activities, where friends are defined using the reverse of the reported friendship link. Panel A reports reduced-form and first-stage estimates; panel B reports IV and OLS estimates. Health Screening is an indicator for whether the individual underwent a program health screening in 2016. Wellness Activities is an indicator for whether the individual participated in any program wellness activity in fall 2016 or spring 2017. Friends Screened and Friends in Activities denote the number of friends participating in health screenings and wellness activities, respectively. The sample is restricted to individuals treated in the baseline year. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.8: Endogenous Peer Effects: Alternative Definitions of Peer Groups

A. Health screening						
	IV	OLS	IV	OLS	IV	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Friends Screened	0.113***	0.242***				
	(0.040)	(0.033)				
Share of Home College Screened (leave-out)			0.193	0.828***		
			(0.374)	(0.202)		
Share of Home Organization Screened (leave-out)					–0.053	0.300***
					(0.074)	(0.080)
Dependent Variable Mean	0.576	0.576	0.576	0.576	0.576	0.576
First-Stage F-Statistic	499.1	—	90.8	—	428.4	—
Number of Observations	3,300	3,300	3,297	3,297	3,216	3,216
R^2	0.025	0.032	0.012	0.021	0.001	0.020
B. Wellness activities						
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Friends Participating in Wellness Activities	0.126*	0.184***				
	(0.068)	(0.030)				
Share of Home College Participating in Wellness Activities (leave-out)			0.189	0.700***		
			(0.493)	(0.151)		
Share of Home Organization Participating in Wellness Activities (leave-out)					–0.042	0.281***
					(0.162)	(0.080)
Dependent Variable Mean	0.314	0.314	0.314	0.314	0.314	0.314
First-Stage F-Statistic	182.3	—	36.2	—	135.9	—
Number of Observations	3,300	3,300	3,297	3,297	3,216	3,216
R^2	0.017	0.018	0.010	0.015	0.003	0.016

This table reports estimates of the effect of the share of peers participating in health screening on own screening (Panel A) and the effect of the share of peers participating in wellness activities on own participation in wellness activities (Panel B), using alternative definitions of the peer group. In the IV specifications, the peer participation share is instrumented by the share of the corresponding peer group assigned to treatment. Columns 1 and 2 define peers using the survey question asking individuals to identify their peers. Columns 3 and 4 define peers as members of the home college, and columns 5 and 6 define peers as members of the home organization. The home-college and home-organization measures exclude the individual. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table D.9: Effect of Treatment on Follow-up Survey Completion

A. Completed 2017 follow-up survey		
	(1)	(2)
Treated	–0.025**	–0.025**
	(0.011)	(0.011)
Number of Treated Friends		–0.001
		(0.009)
Dependent Variable Mean	0.738	0.738
FE for Number of Friends in Study	X	X
Number of Observations	4,834	4,834
R^2	0.004	0.004
B. Completed 2018 follow-up survey		
	(1)	(2)
Treated	–0.032**	–0.032**
	(0.012)	(0.012)
Number of Treated Friends		–0.005
		(0.010)
Dependent Variable Mean	0.625	0.625
FE for Number of Friends in Study	X	X
Number of Observations	4,834	4,834
R^2	0.007	0.007

This table reports estimates of the effect of treatment on completion of the follow-up surveys. The dependent variable in panel A is an indicator for completing the 2017 follow-up survey, and in panel B an indicator for completing the 2018 follow-up survey. Treated is an indicator for assignment to the treatment group in 2016. Number of Treated Friends is the number of friends listed on the 2016 baseline survey who were assigned to treatment. The sample includes all study participants in both the treatment and control groups. All regressions include fixed effects for the number of friends in the study. Standard errors clustered using spectral clustering are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.