Web Appendix

Table of Contents

Web Appendix A: Sample Properties by Different Distance Measures

Table A. Sample Network Characteristics Under Different Distance Measures (PowerLaw-Cluster-Base)

	Population	Distance Measures		
		KS distance	KL divergence	Anderson-Darling
Avg. degree	19.996	18.910	17.006	18.260
Avg. clustering coefficient	0.166	0.167	0.171	0.168
Avg. eigen centrality	0.001	0.001	0.001	0.001
Avg. neighbor degree	120.765	120.522	120.612	121.496

Note. This table summarizes sample network characteristics under different distance measures. KS distance stands for the Kolmogorov–Smirnov statistic. KL divergence stands for the Kullback–Leibler (KL) divergence. Anderson-Darling stands for the Anderson–Darling Statistic.

Web Appendix B: Proof of Detailed Balance under Constraint

Suppose in the current iteration, treatment condition $d = T$ is selected and sample S_T is to be updated. First, we prove that the proposal probabilities $q(S'_T|S_T)$ and $q(S_T|S'_T)$ are symmetric. Note that the proposal probabilities describe the chances of constructing S'_T out of S_T and S_T out of S'_T respectively, where $S'_T \neq S_T$. Suppose that we are sampling n ego networks per treatment condition from the population. By construction, the intersection of the two states $S_T \cap S'_T$ contains $n-1$ ego networks because we first remove a randomly drawn ego (and the ego's alters) from the current set of samples S_T and obtain a reduced sample $S_T \cap S'_T$. The probability of arriving at $S_T \cap S'_T$ from the current state S_T is $\frac{1}{n}$. Then, the probability of transitioning from $S_T \cap S'_T$ to the next state S'_T is $\frac{1}{n_{C(S_{D_r})}}$ where $n_{C(S_{D_r})}$ is the number of nodes in the restricted candidate set:

$$C(S_{D_r}) = \{i' \in G \,|\, i' \notin S_{D_r} \,|\, i' \notin A(S_{D_r})\}, \text{ where } S_{D_r} = \{S_T \cup S_C \backslash i\},$$

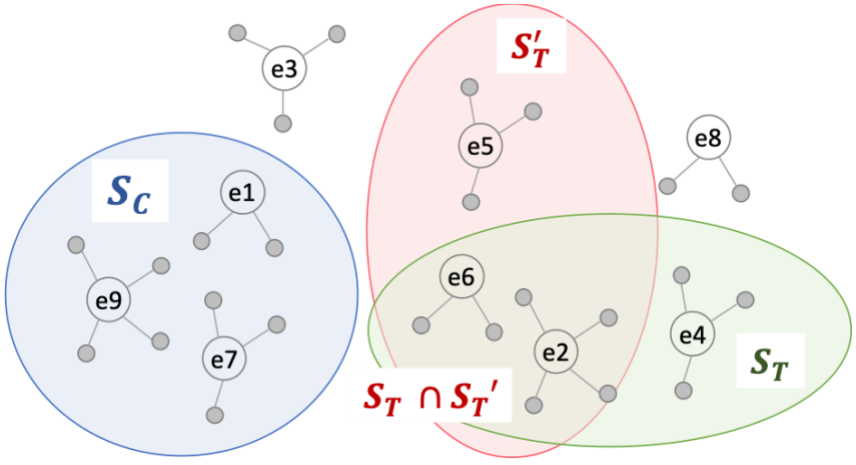
which specifies that any node in this restricted candidate set is at least 2 hops apart from the current remaining sampled egos $S_{D_r} = \{S_T \cup S_C \backslash i\}$. Hence, we can write the proposal probabilities as:

$$P(S_T \to S_T \cap S'_T) = \frac{1}{n}, \text{ and } P(S_T \cap S'_T \to S'_T) = \frac{1}{n_{C(S_{D_r})}}, \text{ so } q(S'_T|S_T) = \frac{1}{n} \cdot \frac{1}{n_{C(S_{D_r})}},$$

Similarly,

$$P(S'_T \to S_T \cap S'_T) = \frac{1}{n}, \text{ and } P(S_T \cap S'_T \to S_T) = \frac{1}{n_{C(S_{D_r})}}, \text{ so } q(S_T|S'_T) = \frac{1}{n} \cdot \frac{1}{n_{C(S_{D_r})}}.$$

Therefore, $q(S'|S) = q(S|S')$. For $S' = S$ this is trivial. The proposal distribution of proposing to move from state S_T to state S'_T remains symmetric. This symmetric proposal distribution, when combined with the same specification on the acceptance probability of S', satisfies the detailed balance condition. We illustrate the idea in the following diagram for $n = 3$.



Web Appendix C: The Steps of the Proposed MH Ego Network Sampling Algorithm

1. Initialize
   (1) Pick $2n$ ego networks from G subject to the contamination constraint, and denote this initial set as $S_{initial}$.
   (2) Randomly pick n ego networks from $S_{initial}$ for treatment condition T and the remaining n ego networks for condition C.
   (3) Set sample states as $\{S_C^m, S_T^m\}$, the restricted candidate set as $C(S_{d_r}^m)$, the set of best samples as $\{S_C^{best}, S_T^{best}\}$ and $m = 0$.

2. Iterate
   (1) Draw a Bernoulli random variable $g \sim \text{Ber}(p)$, with $p = \frac{\Delta\left(f_{1,2}(S_T),\ f_{1,2}(G)\right)}{\sum_d \Delta\left(f_{1,2}(S_d), f_{1,2}(G)\right)}, d \in \{T, C\}$.
   (2) Select sample S_d^m for treatment condition $d = T$ (if $g = 1$) or $d = C$ (if $g = 0$) to be updated.
   (3) Generate a new state S_d' by removing one ego from the current state S_d^m and replacing it with a node from the restricted candidate set $C(S_{d_r}^m)$ according to the proposal distribution $q(S_d'|S_d^m)$.
   (4) Calculate the acceptance probability $a(S_d', S_d^m) = \min(1, \frac{\pi(S_d')}{\pi(S_d^m)})$.
   (5) Accept or reject
      a. Generate a uniform random number $\alpha \in [0,1]$.
      b. If $\alpha \leq a(S_d', S_d^m)$, then accept the new state.
         i. Update the sample state $S_d^{m+1} = S_d'$ and update the restricted candidate set $C(S_{d_r}^{m+1})$.
         ii. If $\Delta\left(f_{1,2}(S_d'),\ f_{1,2}(G)\right) < \Delta\left(f_{1,2}\left(S_d^{best}\right),\ f_{1,2}(G)\right)$, then update the best sample $S_d^{best} = S_d'$.
      c. If $\alpha > a(S_d', S_d^m)$, then reject the new state and set $S_d^{m+1} = S_d^m$.
   (6) Increment: set $m = m + 1$.

The sampling procedure repeats for M iterations and stops when the chain reaches convergence with minimal gain in the distance measure.

Web Appendix D: Theoretical Justifications on Sample Representativeness

Let X denote the optimization variables (i.e., degree and clustering coefficient) used in our sampling method. Let Z denote a variable other than X, with the following possibilities.

We first consider the case when Z is a variable independent of X. For example, Z could be a demographic variable not correlated with degree and clustering coefficient. In this case, since $f(Z|X) = \frac{f(X,Z)}{f(X)} = \frac{f(X)f(Z)}{f(X)} = f(Z)$, the distribution of the variable Z is not affected by the distribution of X in the sample. Therefore, Z will be similar between the sample and the population, between the treatment and control samples, and between our proposed representative samples and the benchmark post-exclusion samples.

We then consider the case when Z is correlated with X. For ease of illustration, we first consider the extreme case when Z is a function of X, i.e., $Z = g(X)$. In this case, there is a deterministic relationship between Z and X. Therefore, given the distribution of X in the sample, the distribution of Z is fixed. Let $F_{SZ}(z)$ and $F_{GZ}(z)$ respectively denote the sample and population cumulative distribution function of Z assessed at z. The distance between the sample and population distributions of Z, assessed through the KS-statistic, becomes

$$\sup_z |F_{SZ}(z) - F_{GZ}(z)| = \sup_{g^{-1}(z)} |F_{SX}(g^{-1}(z)) - F_{GX}(g^{-1}(z))| \leq \sup_x |F_{SX}(x) - F_{GX}(x)|,$$

where $F_{SX}(x)$ and $F_{GX}(x)$ are respectively the sample and population cumulative distribution function of X assessed at x. Since our proposed sampling method minimizes the $\sup_x |F_{SX}(x) - F_{GX}(x)|$ by design, the above inequality indicates that the distributions of Z in our proposed samples are guaranteed to be similar to that in the population, and therefore balanced between the treatment and control conditions.

For the general case when Z is not perfectly correlated with X, it can be written as $Z = g(X) + \varepsilon$, where ε is independent of X. In this case, given $X = x$, the conditional distribution of Z is the distribution of ε shifted by a fixed $g(x)$. The CDF of Z can be written as

$$F_Z(z) = P(Z \leq z) = \int P\big(\varepsilon \leq z - g(x)\big) f_X(x) dx = \int F_\varepsilon(z - g(x)) f_X(x) dx,$$

where $F_\varepsilon(z - g(x))$ is the CDF of ε evaluated at $z - g(x)$. Note that since $0 \leq F_\varepsilon(z - g(x)) \leq 1$, it reduces the large fluctuations in $f_X(x)$. As a result, $F_Z(z)$ is a smoothed version of $F_X(x)$, that is, $F_Z(z) = \tilde{F}_X(x)$. Due to the smoothing mechanism (i.e., with weight between 0 and 1) that reduces the variance of the distribution, the maximum discrepancy between the sample and population distributions of the smoothed data will be less than or equal to the maximum discrepancy between the sample and population distributions of the original data, that is,

$$\sup_z |F_{SZ}(z) - F_{GZ}(z)| = \sup_x |\tilde{F}_{SX}(x) - \tilde{F}_{GX}(x)| \leq \sup_x |F_{SX}(x) - F_{GX}(x)|.$$

Since our proposed sampling method minimizes the $\sup_x |F_{SX}(x) - F_{GX}(x)|$ by design, the above inequality indicates that the distributions of Z in our proposed samples are guaranteed to be similar to that in the population, and therefore balanced between the treatment and control conditions.

Web Appendix E: Comparison of Sample Representativeness

We summarize the sample representativeness in each population network. In each table, Column 2 shows the average value of the network attributes and individual characteristics at the population-level. Columns 3 and 4 present the average values of treatment and control samples generated by representative sampling method, based on 100 simulations. Columns 5 and 6 present the results for post-exclusion samples, based on 100 simulations. For the exclusion approach, the initial sample size per condition is 500, and the size of the representative samples is matched to the size of the post-exclusion samples. Numbers in parentheses show statistical test, including chi-square tests used for categorical variables, z-tests for binary variables, and KS distance for continuous variables.

Table E1. LastFM-Asia

		Representative Samples		**Post-Exclusion Samples**	
	Population	Treatment	Control	Treatment	Control
Avg. degree	7.294	6.661	6.692	3.111	3.114
KS distance		(0.013)	(0.013)	(0.232)	(0.234)
Avg. clustering coefficient	0.219	0.225	0.224	0.189	0.186
KS distance		(0.013)	(0.013)	(0.188)	(0.189)
Avg. eigen centrality	0.0020	0.0017	0.0018	0.0005	0.0005
KS distance		(0.048)	(0.046)	(0.147)	(0.147)
Avg. neighbor degree	23.924	24.632	24.555	23.654	23.743
KS distance		(0.050)	(0.049)	(0.106)	(0.107)
Country (category)	9.275	9.250	9.238	9.357	9.320
% chisq tests reject H0		(0.060)	(0.020)	(0.020)	(0.040)
# artists followed	395.378	397.521	397.874	394.907	397.096
KS distance		(0.052)	(0.053)	(0.053)	(0.053)

Table E2. Twitch-Gamer

		Representative Samples		**Post-Exclusion Samples**	
	Population	Treatment	Control	Treatment	Control
Avg. degree	80.868	61.584	61.734	37.149	37.148
KS distance		(0.019)	(0.018)	(0.144)	(0.144)
Avg. clustering coefficient	0.16	0.168	0.168	0.169	0.168
KS distance		(0.020)	(0.020)	(0.059)	(0.058)
Avg. eigen centrality	0.001	0.001	0.001	0.001	0.001
KS distance		(0.038)	(0.036)	(0.124)	(0.128)
Avg. neighbor degree	2215.552	2298.369	2287.355	2364.138	2335.853
KS distance		(0.048)	(0.046)	(0.060)	(0.053)
View count	188161.766	188151.916	188021.037	225847.357	188244.062

KS distance		(0.045)	(0.045	(0.046)	(0.044)
Lifetime	1541.812	1548.87	1542.931	1539.546	1545.262
KS distance		(0.045)	(0.046	(0.045)	(0.044)
Mature	0.47	0.467	0.471	0.473	0.471
Ztest (% reject null)		5%	3%	3%	5%
Affiliate	0.485	0.486	0.486	0.488	0.482
Ztest (% reject null)		3%	3%	5%	3%

Table E3. Pokec

		Representative Samples		**Post-Exclusion Samples**	
	Population	Treatment	Control	Treatment	Control
Avg. degree	27.317	27.270	27.280	26.120	26.594
KS distance		(0.008)	(0.008)	(0.037)	(0.036)
Avg. clustering coefficient	0.109	0.109	0.109	0.108	0.109
KS distance		(0.008)	(0.008)	(0.039)	(0.036)
Avg. eigen centrality	0.00013	0.00012	0.00012	0.00013	0.00012
KS distance		(0.027)	(0.027)	(0.037)	(0.039)
Avg. neighbor degree	86.297	86.550	86.516	87.547	86.874
KS distance		(0.032)	(0.032)	(0.039)	(0.037)
Age	17.065	17.079	17.087	17.100	17.093
KS distance		(0.025)	(0.025)	(0.034)	(0.035)
Percent profile completion	39.790	39.765	39.897	39.825	39.633
KS distance		(0.033)	(0.033)	(0.037)	(0.035)
Gender (binary)	0.493	0.489	0.492	0.494	0.491
Ztest (% reject null)		7%	7%	5%	4%
Profile public (binary)	0.662	0.661	0.664	0.661	0.664
Ztest (% reject null)		6%	4%	4%	4%

Table E4. PowerLaw-Cluster-Small

		Representative Samples		**Post-Exclusion Samples**	
	Population	Treatment	Control	Treatment	Control
Avg. degree	5.998	5.084	5.119	4.358	4.363
KS distance		(0.050)	(0.048)	(0.112)	(0.115)
Avg. clustering coefficient	0.158	0.163	0.164	0.172	0.171
KS distance		(0.050)	(0.048)	(0.076)	(0.075)
Avg. eigen centrality	0.004	0.004	0.004	0.003	0.003
KS distance		(0.052)	(0.056)	(0.078)	(0.080)

Avg. neighbor degree	23.094	23.094	23.094	23.094	23.094
KS distance		(0.051)	(0.051)	(0.054)	(0.051)
Avg. $Z1$ (continuous)	8.359	8.383	8.383	8.404	8.405
KS distance		(0.052)	(0.050)	(0.061)	(0.060)
Avg. $Z2$ (continuous)	2.821	2.799	2.801	2.784	2.786
KS distance		(0.053)	(0.047)	(0.053)	(0.051)
Avg. $Z3$ (binary)	0.494	0.494	0.496	0.491	0.496
Ztest (% reject null)		6%	5%	7%	4%

Table E5. PowerLaw-Cluster-Base

		Representative Samples		**Post-Exclusion Samples**	
	Population	Treatment	Control	Treatment	Control
Avg. degree	19.996	18.910	18.903	16.858	16.899
KS distance		(0.006)	(0.006)	(0.061)	(0.061)
Avg. clustering coefficient	0.166	0.167	0.167	0.171	0.172
KS distance		(0.008)	(0.008)	(0.057)	(0.056)
Avg. eigen centrality	0.0013	0.0012	0.0013	0.0011	0.0012
KS distance		(0.036)	(0.036)	(0.052)	(0.054)
Avg. neighbor degree	120.765	120.522	121.004	120.973	122.330
KS distance		(0.037)	(0.038)	(0.043)	(0.041)
Avg. $Z1$ (continuous)	9.668	9.670	9.670	9.673	9.672
KS distance		(0.041)	(0.041)	(0.046)	(0.043)
Avg. $Z2$ (continuous)	0.986	0.983	0.984	0.982	0.983
KS distance		(0.041)	(0.041)	(0.043)	(0.042)
Avg. $Z3$ (binary)	0.5	0.497	0.496	0.500	0.502
Ztest (% reject null)		5%	4%	6%	6%

Table E6. PowerLaw-Cluster-Sparse

		Representative Samples		**Post-Exclusion Samples**	
	Population	Treatment	Control	Treatment	Control
Avg. degree	4	3.891	3.890	3.716	3.693
KS distance		(0.004)	(0.004)	(0.028)	(0.029)
Avg. clustering coefficient	0.075	0.075	0.076	0.075	0.076
KS distance		(0.004)	(0.004)	(0.023)	(0.023)
Avg. eigen centrality	0.0005	0.0004	0.0004	0.0004	0.0004
KS distance		(0.039)	(0.037)	(0.039)	(0.040)
Avg. neighbor degree	21.855	22.179	21.752	22.136	21.998

KS distance		(0.040)	(0.037)	(0.040)	(0.038)
Avg. $Z1$ (continuous)	9.352	9.352	9.351	9.355	9.355
KS distance		(0.038)	(0.038)	(0.039)	(0.039)
Avg. $Z2$ (continuous)	1.208	1.206	1.204	1.204	1.204
KS distance		(0.038)	(0.040)	(0.040)	(0.039)
Avg. $Z3$ (binary)	0.502	0.505	0.500	0.503	0.508
Ztest (% reject null)		7%	4%	3%	9%

Web Appendix F: Sample Representativeness and Boundary Conditions When Controlling Both First- and Second-Degree Contamination

We further compare representative sampling method and the exclusion approach on sample representativeness and boundary conditions when both first- and second-degree contamination are controlled.

Similar to Section 4.2 in the main text, we compare the sample representativeness of the two methods under matched sample sizes. For the post-exclusion method, we begin with 500 ego networks per treatment condition and remove all egos that are either (a) directly connected to another sampled ego (first-degree contamination) or (b) share at least one alter with another sampled ego (second-degree contamination).

For the representative sampling method, we match the resulting sample size achieved by post-exclusion and set the distance constraint parameter to $h = 3$, ensuring that sampled egos are at least three hops apart—that is, they are neither directly connected nor share common alters.

Table F1 reports the population averages and the corresponding sample averages, along with the Kolmogorov–Smirnov (KS) distances for four key network attributes: degree, clustering coefficient, eigenvector centrality, and neighbor degree. Across all networks and all variables, representative samples consistently achieve smaller KS distances, indicating closer alignment with the population distributions.

Table F1. Comparison of Sample Representativeness When Both First- and Second-Degree Contamination are Controlled

	LastFM			Twitch			Pokec		
	Pop.	Rp.	Excl.	Pop.	Rp.	Excl.	Pop.	Rp.	Excl.
2nd degree clean sample		195	195		17	17		287	287
Avg. degree	7.290	5.463	1.776	80.868	54.588	2.946	27.317	25.190	10.985
KS distance		0.055	0.438		0.072	0.800		0.014	0.242
Avg. clustering coefficient	0.220	0.248	0.113	0.160	0.184	0.116	0.109	0.113	0.122
KS distance		0.050	0.399		0.078	0.663		0.014	0.150
Avg. eigen centrality	0.002	0.0006	0.0001	0.0011	0.0003	0.00001	0.00013	0.00007	0.00002
KS distance		0.136	0.370		0.381	0.890		0.056	0.235
Avg. neighbor degree	23.924	15.015	9.932	2215.552	724.555	89.590	86.297	69.428	62.148
KS distance		0.237	0.412		0.44	0.842		0.079	0.180

	Small			Base			Sparse		
	Pop.	Rp.	Excl.	Pop.	Rp.	Excl.	Pop.	Rp.	Excl.
2nd degree clean sample		37	37		12	12		338	338
Avg. degree	5.998	5.139	3.574	19.996	20.059	11.892	4.000	3.644	3.145

KS distance		0.035	0.277		0.060	0.393		0.009	0.072
Avg. clustering coefficient	0.158	0.171	0.175	0.166	0.167	0.220	0.075	0.080	0.080
KS distance		0.038	0.167		0.069	0.384		0.009	0.036
Avg. eigen centrality	0.005	0.0024	0.0008	0.001	0.0009	0.0002	0.0005	0.0002	0.0001
KS distance		0.169	0.449		0.176	0.592		0.081	0.128
Avg. neighbor degree	50.10	16.723	8.537	120.765	91.796	30.280	21.855	13.221	10.650
KS distance		0.179	0.496		0.207	0.678		0.104	0.136

Second, we compare the maximum feasible sample sizes for the two sampling methods when both first- and second-degree contamination are controlled. Recall that when controlling only first-degree contamination, representative sampling achieves substantially larger samples than the exclusion approach — 4 to 57 times larger across the six networks (Table 3 in the main text). When both first- and second-degree contamination must be controlled, the advantage of representative sampling becomes even more pronounced. As shown in Table F2, representative sampling achieves 3 to 22 times larger samples than the exclusion approach, with the largest gains in dense networks (Twitch: 200 vs. 9; Base: 700 vs. 32) and smaller gains in sparse networks (Sparse: 2,800 vs. 993). Moreover, for the exclusion approach, the KS distances at the maximum feasible sample sizes exceed 0.1 in all networks, indicating that it cannot satisfy the representativeness threshold under this stricter criterion. In contrast, representative sampling maintains representativeness while achieving significantly larger sample sizes, demonstrating its flexibility and robustness when higher-order contamination must be controlled.

Table F2. Boundary Conditions for Representative Sampling and Exclusion Approach (KS < 0.1)

	Control both First- and Second-degree Contamination	
Population network	Representative Sampling	Exclusion Approach (KS distance)
LastFM	300	26 (0.242)
Twitch	200	9 (0.696)
Pokec	2,700	286 (0.245)
Small	400	38 (0.279)
Base	700	32 (0.497)
Sparse	2,800	993 (0.213)

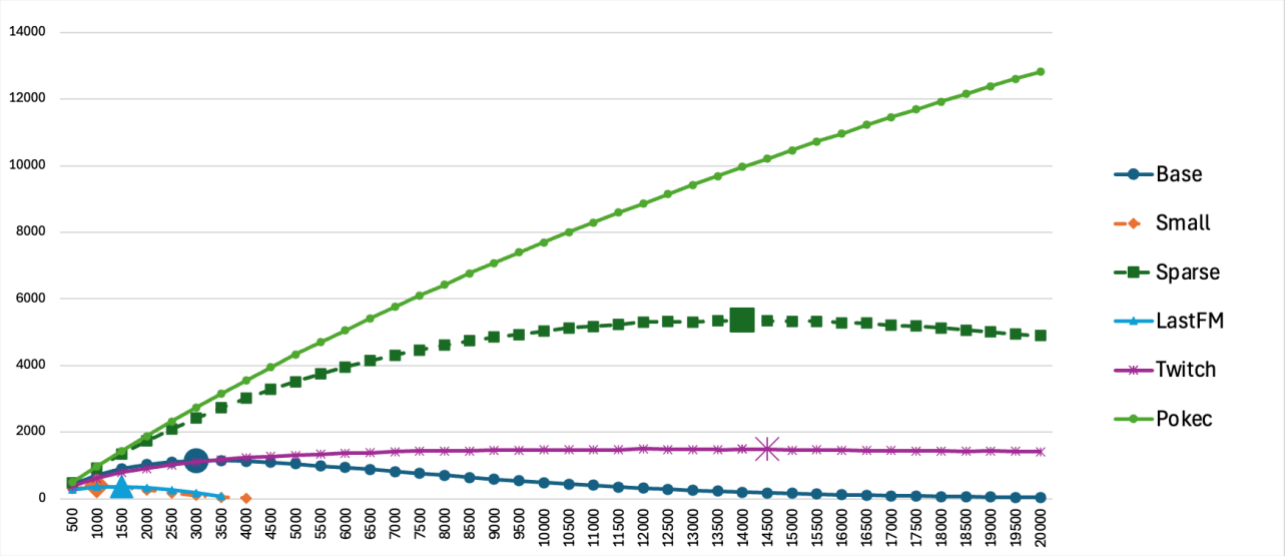
Overall, we find consistent patterns across all analyses. Representative sampling improves both sample representativeness and attainable sample sizes, even when both first- and second-degree contamination are controlled. This highlights the flexibility and robustness of our algorithm. By adjusting the distance constraint, representative sampling can explicitly control the degree of contamination, which is particularly beneficial in small or dense networks.

Web Appendix G: The Number of Qualified Ego Networks in Samples Obtained from the Exclusion Approach

For the exclusion approach, increasing the initial sample size does not necessarily lead to a larger post-exclusion sample. Figure G illustrates this pattern across the six population networks. The x-axis shows the initial sample size per condition and the y-axis shows the number of remaining qualified ego networks that are free of first-degree contamination after exclusion. The enlarged markers indicate the maximum number of qualified ego networks achievable in each network.

For most networks, post-exclusion sample size peaks and then declines as the initial sample grows, because larger initial samples result in higher contamination rates that offset the gain from drawing more egos. This pattern is especially severe for small or dense networks (Base, Small, LastFM, and Twitch) where the maximum is reached early and at low levels. Sparse shows a similar pattern but achieves a larger maximum due to its lower density. Pokec is an exception: as a very large and sparse network with over 1.6 million nodes, its post-exclusion sample continues to grow even at an initial sample of 20,000, suggesting its maximum lies beyond the range we examined.

Figure G. Number of Qualified Ego Networks After Exclusion by Initial Sample Size



Note. Results are shown for first-degree contamination control across the six population networks.

Web Appendix H: IPTW Estimators under Peer Encouragement Designs

A. Experimental Setup and Exposure Mapping

Excluding contaminated nodes leads to nonrepresentative samples, and standard difference-in-means (DIM) estimates may be biased. In such cases, one can apply inverse probability of treatment weighting (IPTW) to reweight nodes by the inverse of the probability of their observed exposure condition (Aronow and Samii 2017).

In a peer encouragement design, treatment is assigned at the ego level, but both ego and alter outcomes may depend on the treatment status of themselves and their neighbors. Let $D_i \in \{0,1\}$ denote whether ego i is treated, and $n(i)$ the set of its alters. Each ego's exposure condition depends on both D_i and $n(i)$:

- d^e_{10} (direct exposure): treated ego, not connected to other egos.
- d^e_{01} (indirect exposure): untreated ego, connected to other egos.
- d^e_{11} (direct+indirect exposure): treated ego, connected to other egos.
- d^e_{00} (no exposure): untreated ego, not connected to other egos.

The contrast between d^e_{10} versus d^e_{00} isolates the direct treatment effect in the absence of interference.

For alter $j(i)$ associated with ego i, exposure depends on the ego's treatment assignment D_i and the alter's direct neighbors $n(j)$:

- d^a_{10} (indirect exposure via focal ego): focal ego treated, not connected to other egos.
- d^a_{01} (indirect exposure via other ego): focal ego untreated, connected to other egos.
- d^a_{11} (indirect exposure via focal+other egos): focal ego treated, connected to other egos.
- d^a_{00} (no exposure): focal ego untreated, not connected to other egos.

The contrast between d^a_{10} versus d^a_{00} captures the indirect (spillover) treatment effect.

For each node i, define the exposure probability vector $\boldsymbol{\pi}_i = (\pi_i(d_1), \ldots, \pi_i(d_k))$ where $\pi_i(d_k)$ is the probability that node i falls under exposure condition d_k. For example, the exposure mapping defined above gives rise to four exposure conditions if the node i is sampled as an ego, four exposure conditions if it is included as an alter, and one condition if it is not being included in the experiment.

Because exact probabilities $\boldsymbol{\pi}_i$ are intractable when networks are large, we approximate $\boldsymbol{\pi}_i$ using Monte Carlo simulation by repeatedly randomizing treatment assignment and counting the proportion of times node i is observed under condition d_k. Following Aronow and Samii (2017), we use 500,000 draws in each network to ensure small bias.

B. Conditions and Limitations of IPTW Estimators

IPTW estimators are unbiased only if the exposure mapping satisfies the consistent potential outcomes assumption (Conditions 1 and 2, Aronow and Samii 2017). Specifically, if two different treatment assignments z and z' lead to the same exposure condition for node i, they must also lead to

the same potential outcome, $y_i(z) = y_i(z')$. In other words, the exposure condition fully characterizes how other nodes' treatments influence node i.

For egos, this assumption holds because under d_{10}^e versus d_{00}^e, ego outcomes depend only on the ego's own treatment. For alters, however, consistency often fails. The same exposure condition (e.g., d_{10}^a) may correspond to being linked to different egos (e.g., a high-influence vs. a low-influence ego). If these differences affect the alter's response, then d_{10}^a does not uniquely determine $y_{j(i)}$, violating the consistency assumption.

When an exposure condition can be mapped to several distinct causal conditions, the exposure mapping is misspecified (Aronow et al. 2021; Aronow and Samii 2017). Under misspecification, IPTW estimators converge not to the true effect but to a weighted average of heterogeneous effects, and produce inaccurate and unstable estimates.

In our response model, each alter's response depends on the specific ego generating the spillover, making IPTW estimators inappropriate for indirect effects estimation. Empirically, we find that IPTW estimates of indirect treatment effect yield large MAE and RMSE and low power and coverage compared with the standard DIM and our representative samples, confirming the limitation of IPTW under misspecification.

C. Horvitz–Thompson and Hájek Estimators

The average direct treatment effect (ADTE) is estimated as the difference in mean potential outcomes between d_{10}^e and d_{00}^e. The Horvitz–Thompson (HT) estimator (Horvitz and Thompson 1952) weights each observation by the inverse of its exposure probability:

$$\widehat{\tau_{HT}}(d_{10}^e, d_{00}^e) = \frac{1}{N}\sum_1^n I(D_i = d_{10}^e)\frac{Y_i}{\pi_i(d_{10}^e)} - \frac{1}{N}\sum_1^n I(D_i = d_{00}^e)\frac{Y_i}{\pi_i(d_{10}^e)}$$

To improve stability, the Hájek estimator normalizes the weights within each exposure condition:

$$\widehat{\tau_H}(d_{10}^e, d_{00}^e) = \frac{\sum_1^n I(D_i=d_{10}^e)\frac{Y_i}{\pi_i(d_{10}^e)}}{\sum_1^n I(D_i=d_{10}^e)\frac{1}{\pi_i(d_{10}^e)}} - \frac{\sum_1^n I(D_i=d_{00}^e)\frac{Y_i}{\pi_i(d_{10}^e)}}{\sum_1^n I(D_i=d_{00}^e)\frac{1}{\pi_i(d_{10}^e)}}$$

The average indirect treatment effect (AITE) is estimated as the difference in potential outcomes between d_{10}^a and d_{00}^a averaged at ego-network level:

$$\hat{\gamma}(d_{10}^a, d_{00}^a) = \frac{1}{n_T}\sum_{i\in S_T}\bar{Y}_{j(i)}(d_{10}^a) - \frac{1}{n_C}\sum_{i\in S_C}\bar{Y}_{j(i)}(d_{00}^a),$$

where $\bar{Y}_{j(i)}(d)$ is the average outcome of alters connected to ego i under exposure condition d.

Within each ego network, the HT estimator for $\bar{Y}_{j(i)}(d)$ is:

$$\bar{Y}_{j(i)}^{HT}(d) = \frac{1}{k_i}\sum_{j\in n(i)} I(D_i = d)\frac{Y_j}{\pi_j(d)},$$

where k_i is the degree of ego i.

The Hájek estimator is:

$$\bar{Y}^{H}_{j(i)}(d) = \frac{\sum_{j \in n(i)} I(D_i = d) \frac{Y_j}{\pi_j(d)}}{\sum_{j \in n(i)} I(D_i = d) \frac{1}{\pi_j(d)}}.$$

D. Bias–Variance Trade-off

Both HT and Hájek estimators reweight observations to reconstruct representative exposure means. This weighting reduces bias but can increase variance when some $\pi_i(d)$ are small or highly unequal. In network settings, π_i often varies with network structure (e.g., degree or clustering), leading to extreme weights.

The Hájek estimator introduces normalization to stabilize the weights but does not eliminate the variability. While IPTW estimators are unbiased in expectation, their variance is typically larger than for unweighted estimators such as DIM, resulting in larger MAE and RMSE. This bias–variance trade-off has been largely documented in prior literature (Aronow et al. 2021; Aronow and Samii 2017; Eckles, Karrer, and Ugander 2017).

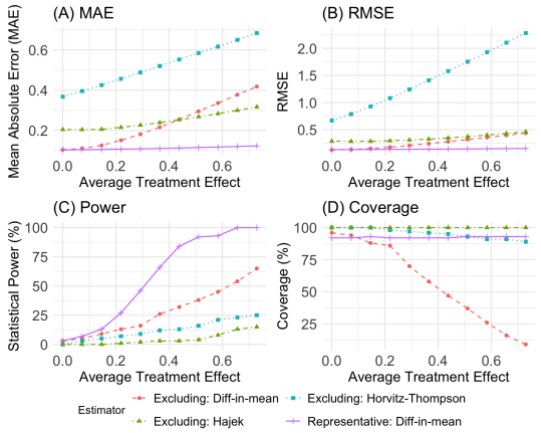
Web Appendix I: ADTE and AITE Estimation

For each population network, we incorporate available individual characteristics along with individual degree K_i as covariates $\mathbf{x}_i$ in the baseline outcome equation of the response model (Eq. 7 in the main text). The covariates used for each network are as follows:

- LastFM: Number of artists followed
- Twitch: View count, Lifetime, Mature, Affiliation
- Pokec: Age, Profile Completion, Gender
- Simulated Networks: Z1, Z2, Z3

Next, we summarize the ADTE (left) and AITE (right) estimates for each network. The patterns are consistent across all six networks. Representative samples consistently achieve lower MAE and RMSE and higher power and coverage than post-exclusion samples under all three estimators. For AITE, IPTW estimators perform especially poorly across all networks, consistent with the consistency condition violation discussed in Section 5.2 in the main text and Web Appendix H.

Figure I1. LastFM-Asia



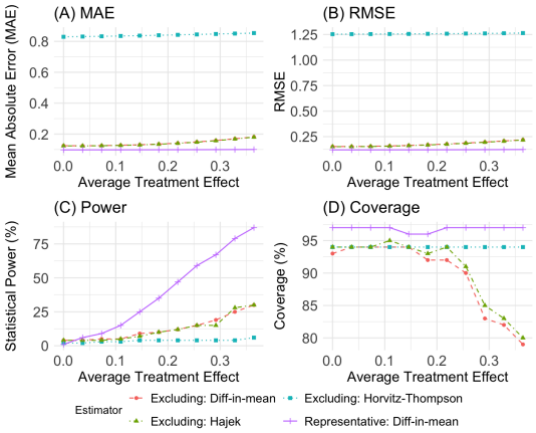
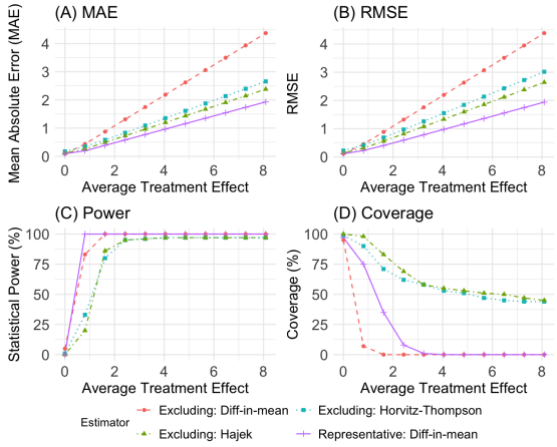


Figure I2. Twitch-Gamer



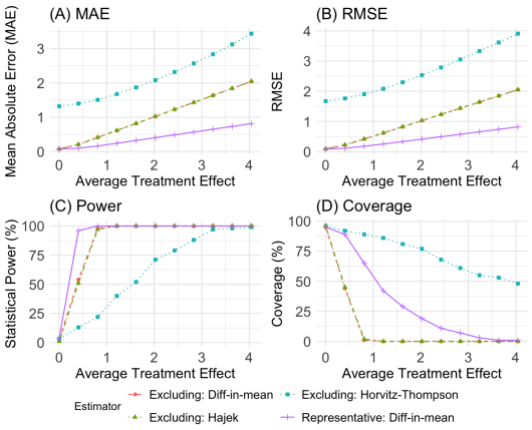


Figure I3. Pokec

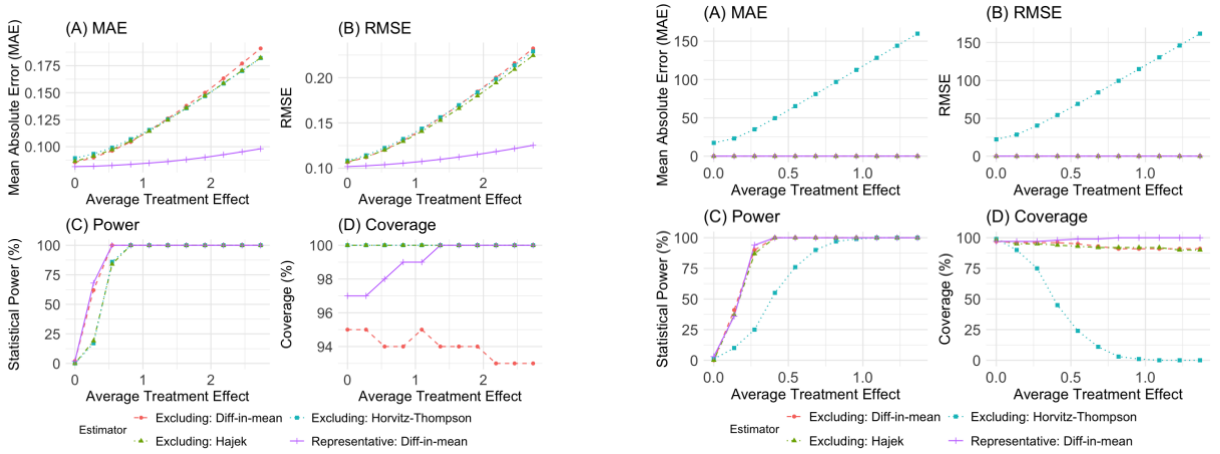


Figure I4. PowerLaw-Cluster-Small

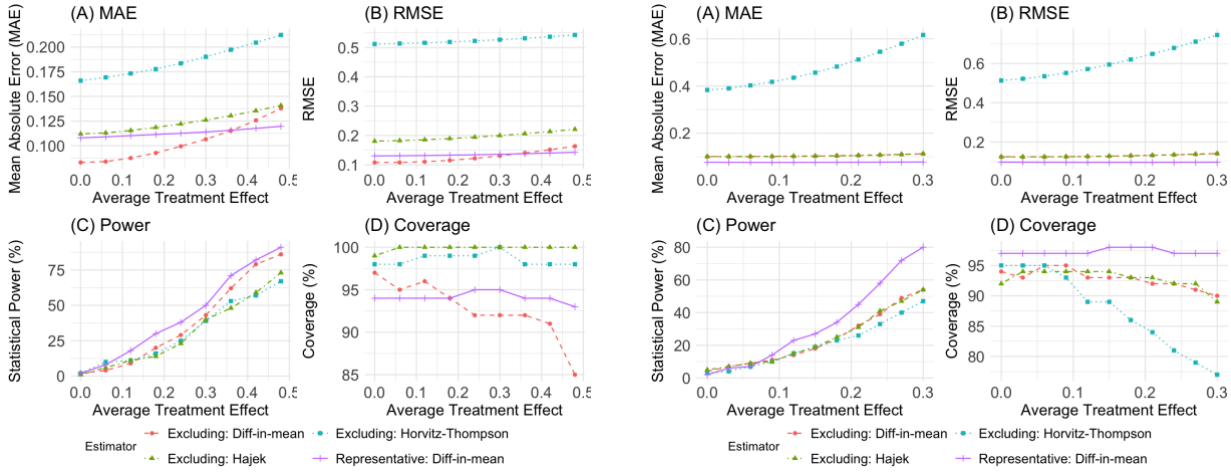


Figure I5. PowerLaw-Cluster-Base

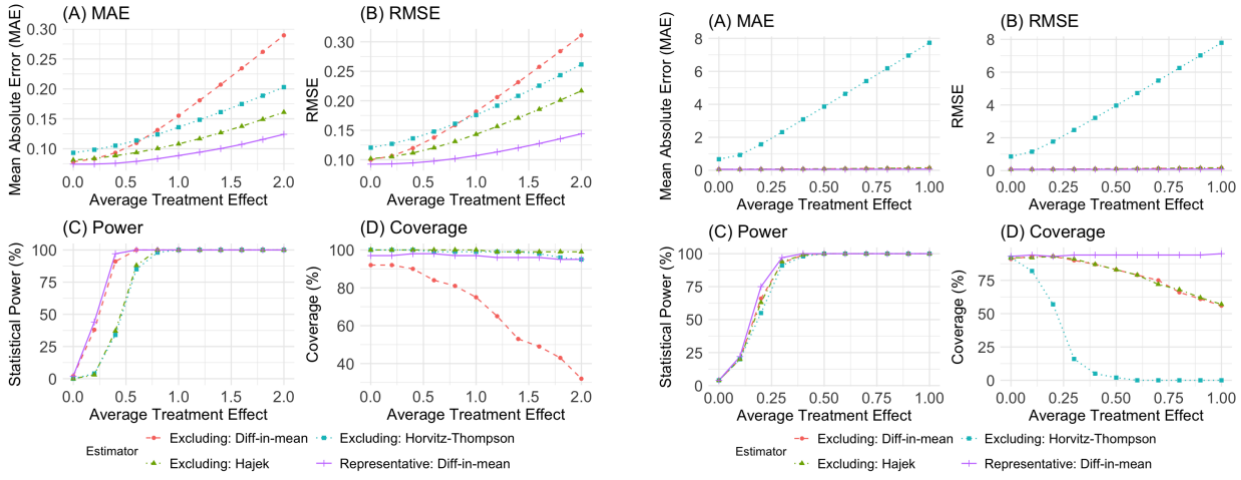
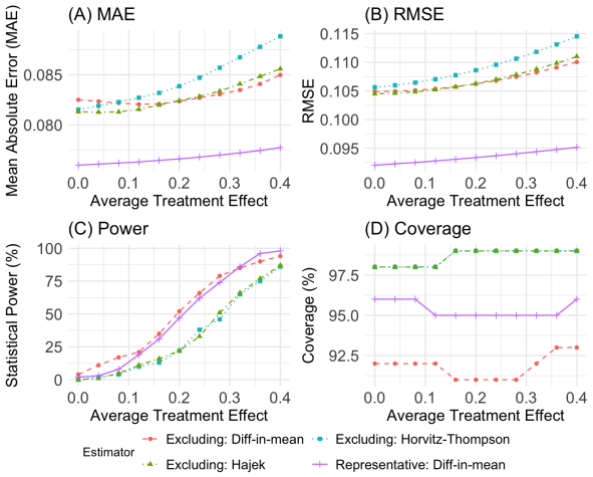
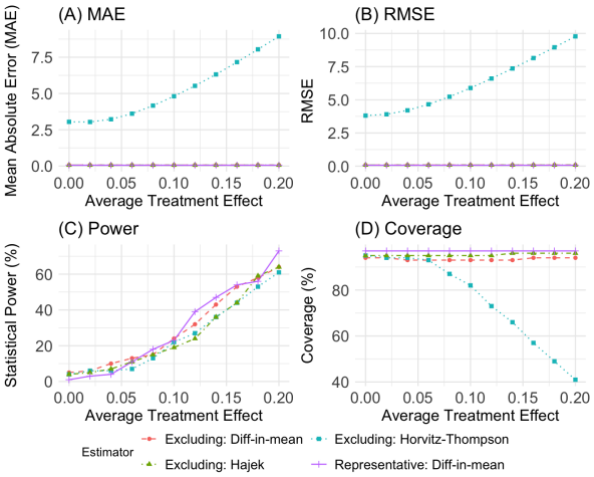


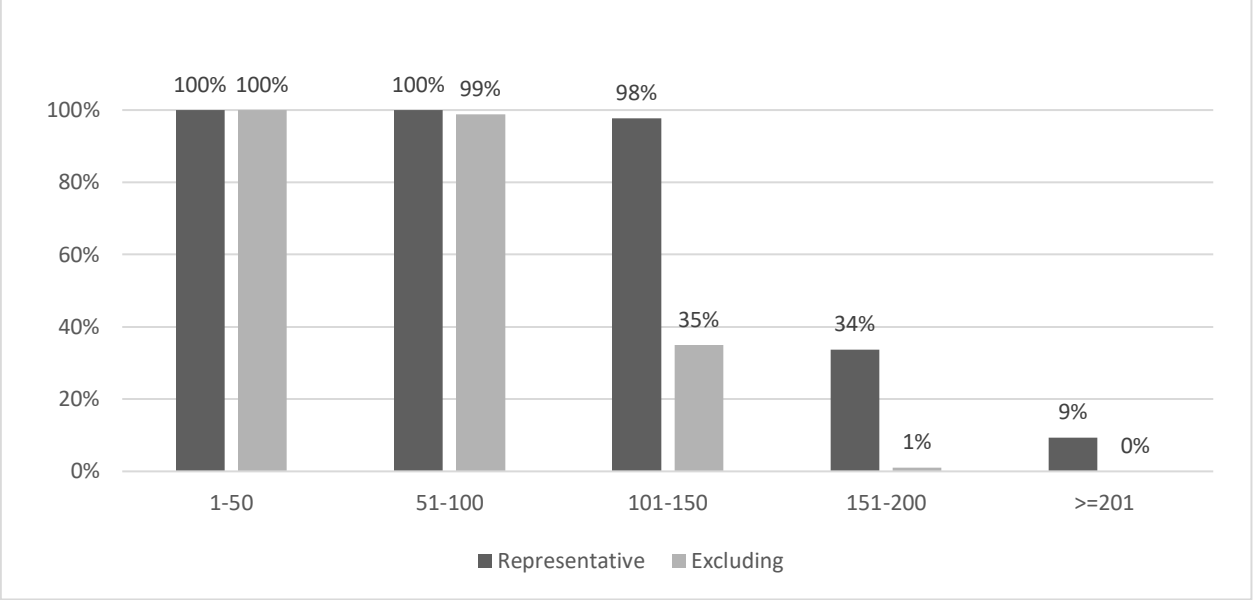
Figure I6. PowerLaw-Cluster-Sparse


(A) MAE
(B) RMSE
(C) Power
(D) Coverage
Mean Absolute Error (MAE)
RMSE
Statistical Power (%)
Coverage (%)
Average Treatment Effect
Estimator
Excluding: Diff-in-mean
Excluding: Horvitz-Thompson
Excluding: Hajek
Representative: Diff-in-mean


(A) MAE
(B) RMSE
(C) Power
(D) Coverage
Mean Absolute Error (MAE)
RMSE
Statistical Power (%)
Coverage (%)
Average Treatment Effect
Estimator
Excluding: Diff-in-mean
Excluding: Horvitz-Thompson
Excluding: Hajek
Representative: Diff-in-mean

Web Appendix J: Probability of Obtaining Clean Samples (Free of First-Degree Contamination) in Each Degree Subgroup

To confirm this potential explanation, we conducted 500 simulations to calculate the proportion of qualified ego networks in each degree subgroup in representative samples and post-exclusion samples. The results support our hypothesis: the exclusion approach struggles to keep adequately qualified ego networks for large-degree groups, with proportions of 35% for the 101–150 degree group, 1% for the 151–200 degree group, and 0% for groups with degree of 200 or more. This scarcity makes it challenging to estimate the CADTE for these subgroups using post-exclusion samples. In contrast, for representative samples, these proportions are significantly higher at 98%, 34%, and 9% for these same high-degree subgroups, respectively, providing sufficient data for estimating the CADTE.



Note. The proportions represent how often the sampling method obtains qualified ego networks in each degree subgroup. The proportions are averaged over 500 simulations.

Reference

Aronow, Peter M., Dean Eckles, Cyrus Samii, and Zonszein (2021), "Spillover Effects in Experimental Data," in *Cambridge Handbook of Advances in Experimental Political Science*, 289:319.