# Estimation, Experimental Design, and Optimal Policy when Facing Partial Network Data from a Single Network

*Keywords: network inference; experimental design; policy optimization; causal inference; measurement error*

## Extended Abstract

Networks pervade social, economic, and epidemiological systems, yet real-world studies typically yield only one large network observation, compromised by sampling gaps, survey error, or privacy-preserving noise. Under such partial and noisy measurement, classical network inference and experimental design methods fail, since they assume multiple independent graphs or fully observed ties. This paper delivers two distinct but complementary contributions—one focused on *inference* under a single imperfect network, the other on *experimental design and policy optimization*, providing a unified toolkit for rigorous causal analysis and intervention planning in realistic settings. Our method is based on representing common processes on graphs as *structural causal models*. A structural causal model is a set of equations and exogenous noise variables that together define how treatments, network structure, and latent disturbances generate observed outcomes, making all counterfactuals well-defined.

**1. Inference with a Single Noisy Network.** We begin by showing how to recover consistent parameter estimates when only a mismeasured network $G^*$ is available. First, an *iterated-expectations* argument yields unbiased moment equations

$$\mathbb{E}\big[m(Y_i,\,S_i,\,V_i;\beta)\big] = 0,$$

where exposures $V_i = f_V(a;\varphi_i(G^*))$ and pre-treatment confounders $S_i = f_S(X;\vartheta_i(G^*))$ are built from the observed graph. Under mild "affinity-set" dependence conditions, we prove a Central Limit Theorem for these moments, despite network-driven correlations. Second, to guard against misspecification of either the outcome model or the network-measurement model, we propose a *doubly robust* estimator that merges outcome regression $\hat{\mu}(S_i,V_i)$ with a graph-reconstruction correction $\hat{G}$. We show consistency if *either* component is correctly specified. Third, leveraging a graphon perspective via stochastic block models (SBMs), we denoise $G^*$ by partitioning nodes into $K = O(\sqrt{n/\log n})$ communities, achieving vanishing $L_2$ error and markedly improving finite-sample bias and variance in both simulation and application.

**2. Experimental Design and Policy Optimization.** Building on our inference machinery, we tackle the challenge of designing experiments and allocating treatments when network ties are noisy. Directly optimizing over all $2^n$ assignments is intractable for moderate $n$. We introduce a *block-saturation* technique: after denoising via SBMs, treatments are assigned at the community level, reducing the decision space from size $2^n$ to $2^K$, with $K \ll n$. We then express the design problem—whether minimizing estimator variance or maximizing expected outcomes under budget constraints—as a convex program over block-level saturation variables. Finally, a Bayesian optimization routine rapidly identifies near-optimal allocations in seconds, even for $n \sim 10^3$, enabling scalable and cost-effective network intervention planning.

**3. Validation and Empirical Case Studies.** In synthetic networks spanning sparse to dense regimes and varying measurement error levels, our inference procedures achieve nominal 95% confidence coverage and reduce bias by up to 50% relative to naive plug-in estimators. In a pit-planting trial in Malawi [1], our design raises knowledge-adoption rates by 22% over standard A/B allocation with identical sample sizes. For example, Figure 1 shows that targeting by estimated block structure outperforms naive degree-based seeding, with the greatest gains in sparser villages. Further refinement by choosing the highest-degree within each block delivers adoption rates very close to the oracle optimum. In an information-diffusion experiment in India [2], our seeding strategy increases call-in responses by 24% relative to random seeding under the same budget.

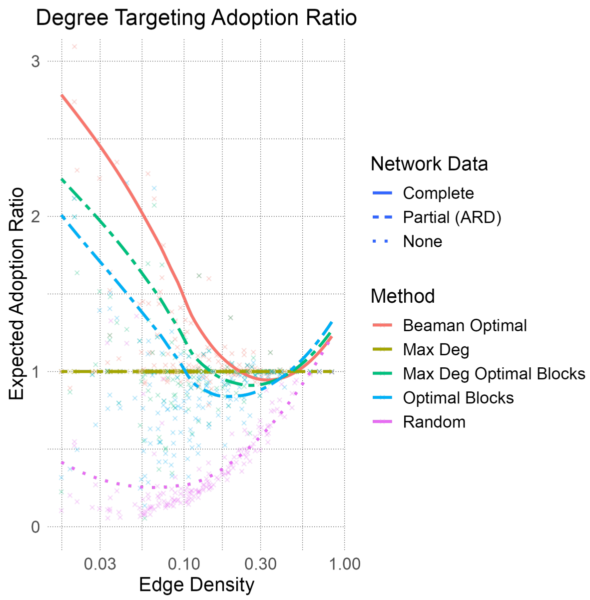


Figure 1: Adoption ratio under complex contagion as a function of village network density, comparing seeding by highest-degree nodes, uniform random, and model-based block targeting (with and without degree refinement).

By explicitly modeling network measurement imperfections and uniting robust estimation with scalable design algorithms, we extend rigorous causal network analysis to settings once considered intractable. This framework empowers researchers and policymakers to draw valid inferences and plan effective interventions from a single noisy network. Future work will explore dynamic networks, endogenize tie formation, and apply these methods to digital platforms and epidemic control.

# References

[1] Lori Beaman et al. “Can Information Campaigns Together with Incentives Promote Technology Adoption? Evidence from Malawi”. In: *American Journal of Agricultural Economics* 103.3 (2021), pp. 789–812. DOI: 10.1111/ajae.12104.

[2] Abhijit V. Banerjee et al. “Gossip: Identifying Information Trade in Networks”. In: *Journal of Political Economy* 127.4 (2019), pp. 1475–1515. DOI: 10.1086/701908.