Does Topology Matter? The Impact of Network Structure on Graph Cluster Randomization

Kaleigh Clary

Seeking internship

Research interests: statistical relational learning, experimental design, casual estimation

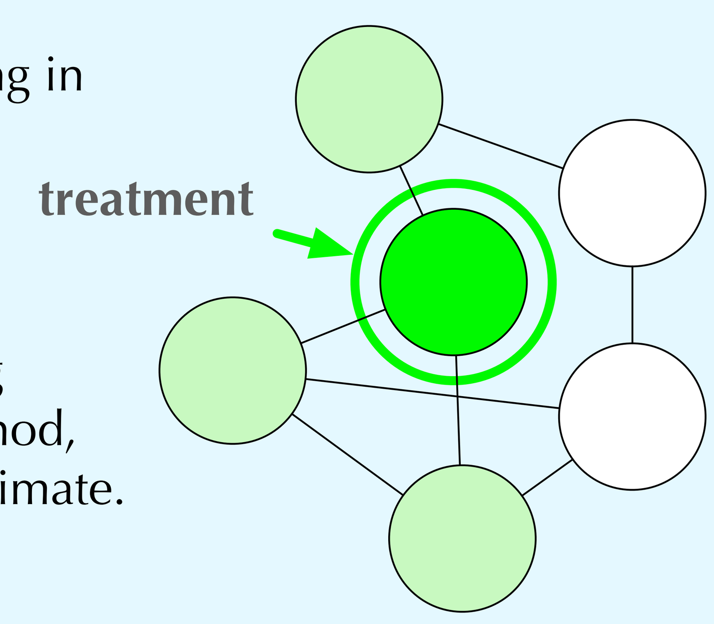
MS/PhD, 2020
BA Computer Science, Mathematics
Hendrix College 2014

1. Motivation

A/B testing is a powerful and widely used method for performing causal inference in real-world settings. These experiments are difficult to reason over in networks, where treatment effects may spill over to individuals assigned to control.

Graph cluster randomization [1] provides a framework for A/B testing in networks which reduces bias from **interference** between units.

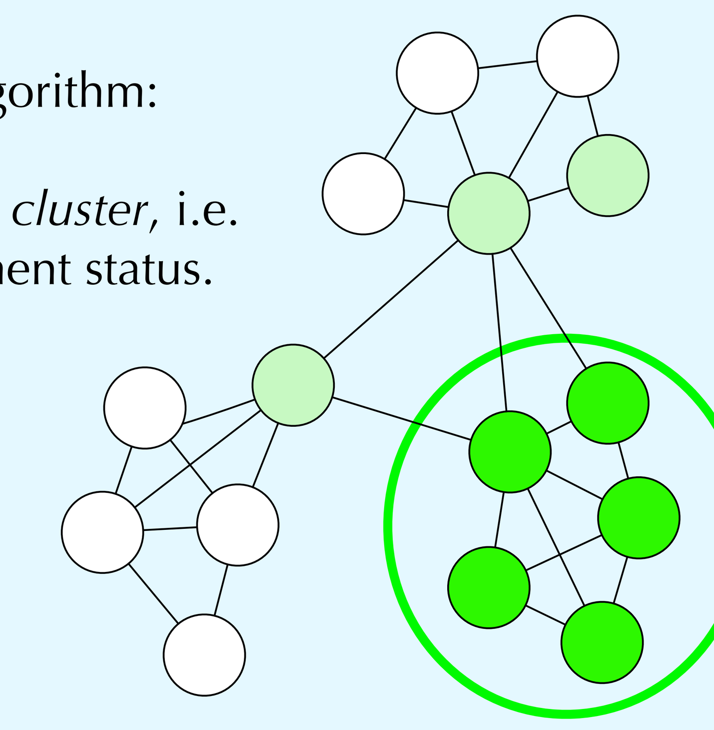
However, little is known about the interaction between the underlying graph topology, the clustering method, and the error of the final causal estimate.



2. Graph Cluster Randomization (GCR)

For a given graph *G* and clustering algorithm:
1. Cluster the graph *G*.
2. Randomly assign treatment to each *cluster*, i.e. all nodes within a cluster share treatment status.
3. Estimate the causal effect.

Assigning treatment to entire clusters provides an approximation to the social behavior of nodes under global treatment or global control and limits spillover effects.



GCR produces an unbiased estimator of average treatment effect (ATE) irrespective of the graph partitioning [1]. However**,** poorly chosen clusters may **increase the variance** of the estimator.

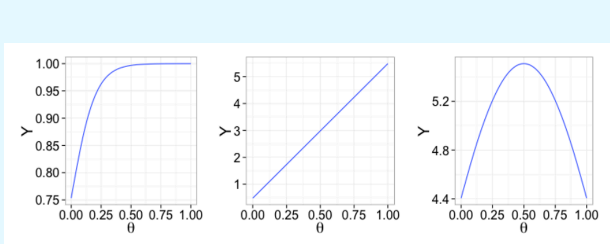
We characterize the relationship between graph topologies and causal effect estimation. We show that **modularity**, a commonly used metric for measuring the quality of community detection algorithms, can be used as a reliable proxy for error in treatment effect arising from the clustering method.

3. Experimental design and topology

Estimation using GCR is influenced by local graph topology, clustering technique, and exposure model. This introduces a large space of variation within the GCR framework.

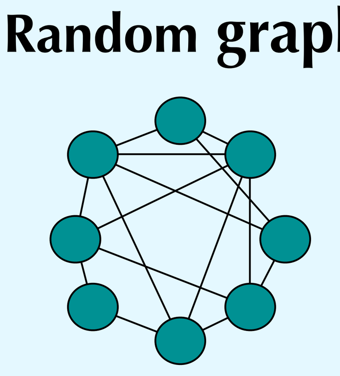
Outcome models:

The form of the response function of an individual according to its number or proportion of treated friends.

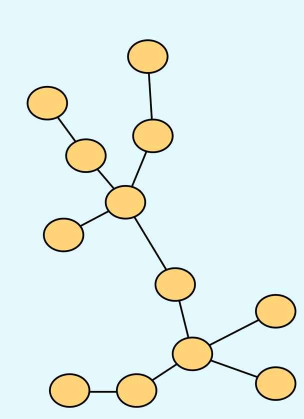


Functional form of three outcome models. *Y* is the treatment response dependent on *Θ*, the proportion of treated neighbors. Reproduced from [2] with permission from the authors.

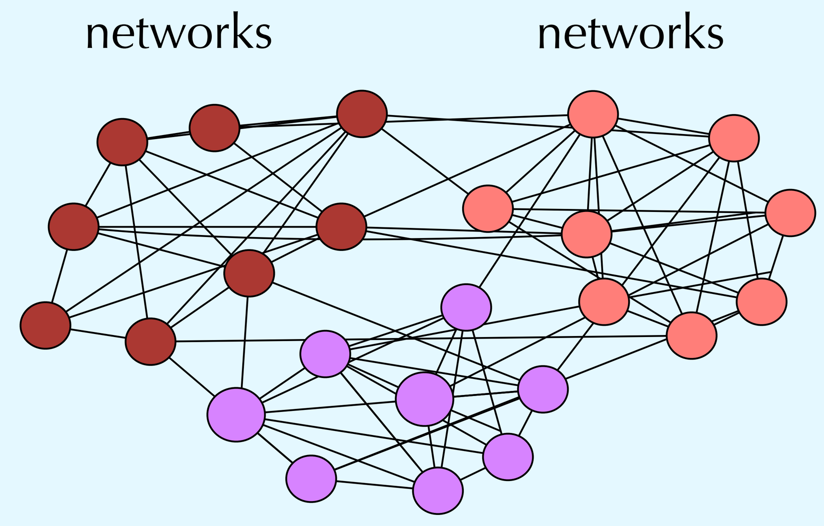
Random graphs:



(a) small-world networks



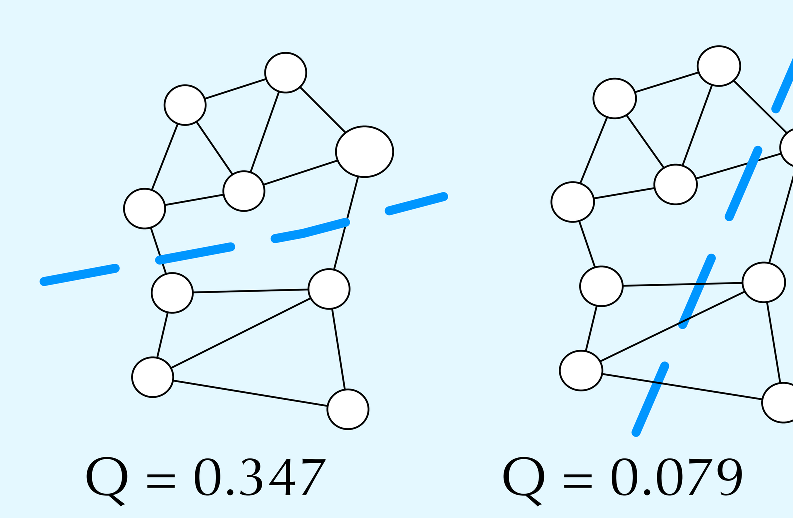
(b) scale-free networks



(c) stochastic block models (SBMs)

4. Modularity and effect estimation

Modularity (Q) is a measure of the division of the network into clusters. It is calculated from the number of edges between nodes in the same cluster and the number of edges between nodes in different clusters.



$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\frac{s_v s_w + 1}{2}$$

Ugander et. al. [1] show the variance of the effect estimator is deeply linked to the **exposure probabilities** of each node.

Exposure probabilities of a node depend directly on the treatment status of its neighbors. These are assigned by cluster, meaning the exposure probabilities are upper bounded by the modularity of the clustering.

5. Modularity as variance bounds

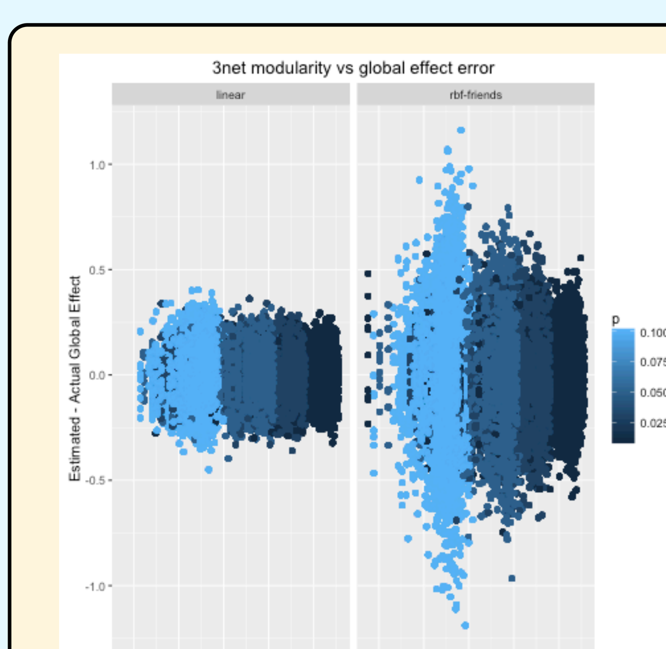
Reducing variance bounds for an unbiased estimator increases confidence in the ATE estimate.

Exposure probabilities also depend on the true outcome model, which is not known in practice. Modularity bounds the variance of the estimator according only to the graph clustering.

Clusterings with high modularity **reduce the variability of the average treatment effect estimate** due to choices in the experimental design.

Experimental setup:
1. Generate random graphs using four different graph generation algorithms, sweeping across parameter settings.
2. Construct graph clusterings using the 3-net algorithm.
3. For each clustering, randomly assign clusters to treatment or control.
4. Estimate the treatment effect as a function of treatment assignment.
5. For each outcome model, calculate the actual treatment effect as a function of treatment assignment, and determine the ATE error of the estimate.

6. Experiments



small-world networks

stochastic block models

For SBMs and small-world networks, graph clusterings with higher modularity have lower variance in average treatment effect error.

The bounds on effect estimate variance depend on the specific graph topology. Both the random graph type and generation parameters influence the variance bounds.

Our results show that **modularity functions as a proxy for the variance in average treatment effect error** induced by the graph topology.

[1] J. Ugander, B. Karrer, L. Backstroke, and J. Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,* pages 329-337. ACM, 2013.
[2] D. Arbor, D. Garant, D. Jensen. Inferring network effects from observational data. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 715-724. ACM, 2016.

Joint work with David Arbour and David Jensen.