On the Effect of Triadic Closure on Network Segregation

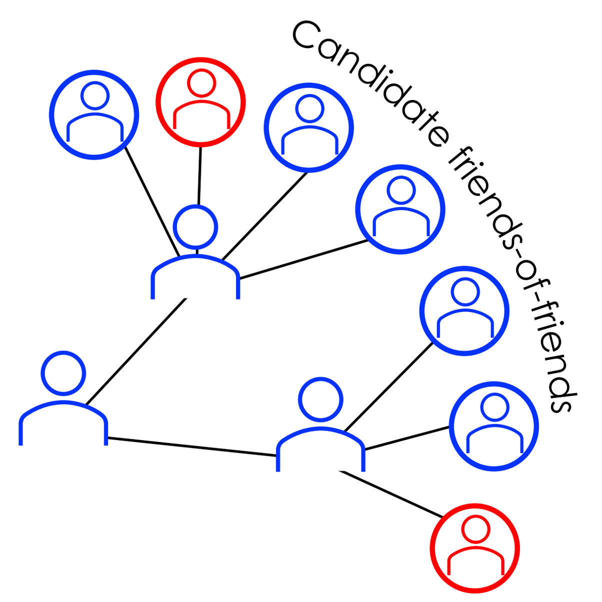
Rediet Abebe[1], Nicole Immorlica[2], Jon Kleinberg[3], Brendan Lucier[2], Ali Shirali[1]

[1]University of California, Berkeley [2]Microsoft Research [3]Cornell University

1) INTRODUCTION

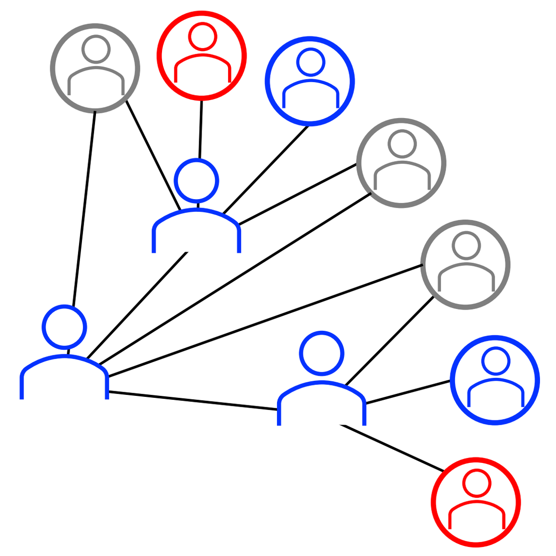
- **Homophily** amplifies segregation in social settings.
- **Key Question:** What existing processes can we leverage to decrease the segregating effects of homophily?
- **Answer:** Challenging a common sociological belief, we show that **triadic closure** is such a phenomenon.
- **Intuition**: Existing literature assumes that triadic closure amplifies segregation since friends-of-friends are likely to be similar. We challenge this conjecture using popular network formation models:

LONG HELD ASSUMPTION



Friends-of-friends are likely to be similar. So, triadic closure amplifies homophily

OUR FINDINGS

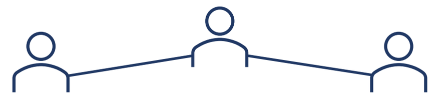


If friends-of-friends are not already connected, it may be because they are dissimilar. So, triadic closure works against homophily.

DEFINITIONS

We consider networks with heterogeneous nodes. **Homophily** is the process by which individuals are more likely to form ties with whom they share similarities.

Triadic closure is a process in which individuals are more likely to form ties with whom they share mutual connections. It operates on a graph-theoretic structure called a **wedge**. Wedges consist of two nodes that have a neighbor in common but themselves are not linked.

Wedge: 

We measure **network integration** using the fraction of edges between dissimilar nodes.

2) STOCHASTIC BLOCK MODEL

Model: Edge (i, j) is formed with probability p if $\text{type}(i) = \text{type}(j)$, q otherwise.

Homophily: $p > q$.

Def. (absolute effect): Measure integration after closing a random wedge.

Def. (relative effect): Measure integration after closing a random wedge *relative* to integration after a baseline intervention e.g., adding a random edge.

3) JACKSON-ROGERS MODEL

Model (informal): At each step, a new node arrives and makes new connections in two phases.

1st phase: it randomly selects N_S and N_D initial friends from similar and dissimilar nodes.

2nd phase: it chooses N_F nodes from the set of accessible nodes through an outbound edge of an initial friend.

This process is also biased: α proportion of these N_F nodes will be selected from the friends of the similar initial friends.

Homophily in 1st phase: $N_S > N_D/(K-1)$ (K is #groups)

Def. (absolute effect): Measure integration in equilibrium by increasing N_F, while N_S and N_D are fixed.

Def. (relative effect): Measure integration in equilibrium by increasing N_F, while $N = N_S + N_D + N_F$ and N_D/N_S are fixed.

4) FIXED-NODE EVOLVING MODEL (ASIKAINEN ET AL.)

Model: With a random initial structure and two node types, at each iteration, a focal node is selected randomly. Then a candidate node is chosen by triadic closure with prob. c or uniformly at random with prob. $1 - c$. A link is formed between focal and candidate nodes with prob. s' if the candidate is selected by triadic closure and s o.w. A random edge connected to the focal node is removed whenever it forms a new edge with a candidate node.

Homophily Originally in Asikainen et al.: $s' = s > 1/2$ → imposes extra homophily to triadic closure.

Homophily in our model: $s > 1/2$ while $s' = 1/2$.

Def. (relative effect): Measure integration by increasing c. This notion implicitly compares the effect of triadic closure with (homophilous) random edge addition.

SUMMARY OF RESULTS

<table>
<tr><th></th><th>Stochastic Block Model (SBM)</th><th>Jackson-Rogers (JR)</th><th>Asikainen et al.</th></tr>
<tr><td>Absolute effect</td><td>[Thm. 2.1]**
Positive* for any SBM iff the network is homophilous.</td><td rowspan="3">[Thm. 3.2]
Positive for any JR model with K types and $1 > \alpha > 1/K$ iff the first phase is homophilous.</td><td>-</td></tr>
<tr><td rowspan="2">Relative effect</td><td>[Thm. 2.2]
(Compared to adding a random edge)
Negative* for any SBM iff the network is homophilous or is sufficiently heterophilous.</td><td rowspan="2">[Thm. 4.1]
Positive for two equiprobable groups iff the network is homophilous.</td></tr>
<tr><td>[Thm. 2.4]**
(Compared to adding a homophilous random edge)
Positive* for any SBM iff the network is neither very homophilous nor very heterophilous.</td></tr>
</table>

* Diverging conclusions when comparing absolute versus relative effects of triadic closure calls for further precision in defining the effect.

** **Beyond network integration:** Similar results hold for an SBM with two types if we consider the relative position of the minority measured by *eigenvector centrality* [Thm. 2.5&2.6].

FURTHER ON JACKSON-ROGERS

Thm. 3.1 (long run behavior of JR Model). For an evolving JR network with K types and $1 > \alpha > 1/K$, the network integration converges to

$$\frac{N_D + (1-\alpha)N_F}{N_S + N_D + \frac{K}{K-1}(1-\alpha)N_F}$$

with the rate of $O(t^{-\frac{N_S+N_D}{N}})$, regardless of the distribution of node types.

Behavior under a series of interventions: We focus on interventions acting solely on the first phase since an authority may have more leverage in the initial phase (e.g., dorm assignment, initial friendship recommendation).

Thm. 3.4 (effect of interventions in JR model, informal). After intervening in a period sufficiently smaller than the age of an evolving JR network by setting N_S while $N_S + N_D$ is kept fixed, (1) the immediate effect of an intervention is independent of other interventions, negatively proportional to the change of N_S, and larger if applied earlier, (2) the long-term effect disappears with the rate of $O(t^{-c})$, $c < 1$. We also provide optimum interventions thm.

EXPERIMENT

For a citation network known to be captured well by the JR model, we first define types by clustering fields of study and then propose a method to estimate model params from observational data. The estimated params along Thm. 3.1 are shown to predict network integration after a few years well.

EXPERIMENT (CONT.)

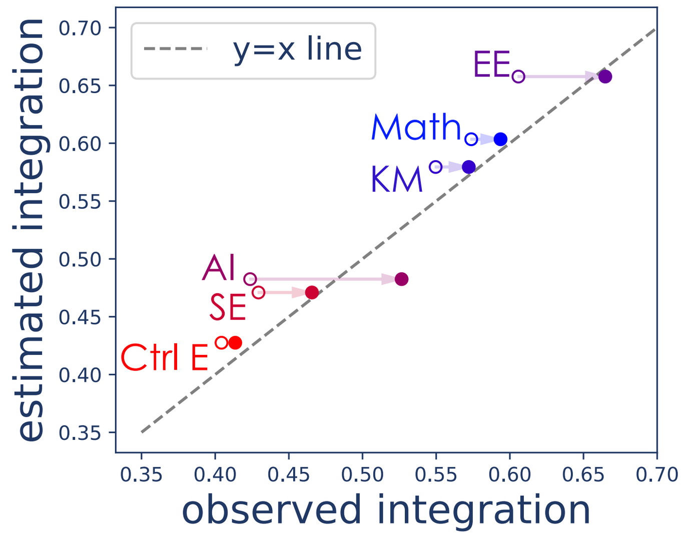


Fig. 1. Estimated vs. observed integration. Empty and filled circles correspond to the beginning and end years of the study. Although clusters have various fields with different frequencies, their behavior in equilibrium is well-predicted from the theory with only a few params.

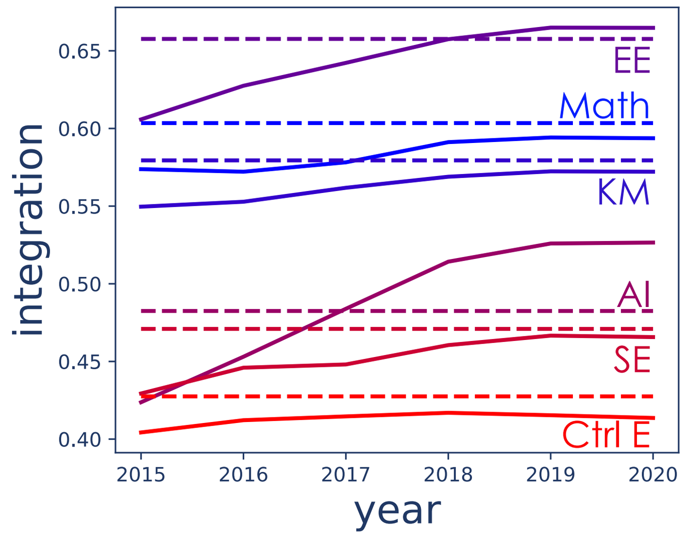


Fig. 2. Convergence behavior of observed integration (solid) to estimated integration (dashed). Except for one field (AI), network integration has converged to the predicted value.

REFERENCES

Jackson, M. O., & Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks?. *American Economic Review*, *97*(3), 890-915.

Asikainen, Aili, et al.. (2020). Cumulative effects of triadic closure and homophily in social networks. *Science Advances*, 6(19), eaax7310.