Targeting with Network Structural Information

Bo-An Chen

June 4, 2024

Introduction

A Lead in Example

- Imagine a startup in the fashion industry striving to launch a new line of clothing.
- They want to leverage social media platforms to reach their target audience effectively.
- In their quest to maximize impact, they turn to network targeting strategies.
- What data should they collect, and how could this information be used?

What Information is Helpful?

- On average, how many friends does one person have?
- Does the number of friends differ across groups?
- Does each group occur to have homophily?

How to Survey?

- In general, personal information is easy to extract. However, it's costly to reveal their social interaction.
- Now, we want to conduct a market experiment.
- It can be individual, regarding how many of their friends they chat about fashion.
- It could also be a snowball survey, where we ask their friends to join the survey sequentially.
- How do we analyze an endogenously formed network?

The Basics of Networks

A network can be represented as a graph $G = (V, E)$ (I'll use $\mathbf{y}$) later instead.

- V is the set of vertices, or the agents in networks.
- E is the set of edges or the links between agents in networks.

We also define the following:

- $N(v) \subseteq V$ is the set of neighbors of agent $v \in V$.
- $I(v) \subseteq E$ is the set of edges that include $v \in V$.

ERGM Model

ERGM is a model describing a graph **y** with a probability in a general exponential form:

$$\pi(\mathbf{y}|\theta, \mathbf{z}) = \frac{1}{z(\theta)} \exp\left(\sum_{d=1}^{D} \theta_d s_d(\mathbf{y}) + \sum_{i=1}^{N} z_i f_i(\mathbf{x})\right) = \frac{\exp\left(\theta^{\mathrm{T}} \mathbf{s}(\mathbf{y}) + \mathbf{z}^{\mathrm{T}} \mathbf{f}(\mathbf{x})\right)}{z(\theta)}$$

where the sufficient statistics in our study are:

$$\theta^{\mathrm{T}} \mathbf{s}(\mathbf{y}) = \theta_\nu \nu(\mathbf{y}) + \theta_\rho \rho(\mathbf{y})$$

with

- $\nu(\mathbf{y}) = \sum_{i,j} y_{ij} \ \#$ of edges
- $\rho(\mathbf{y}) = \sum_i \sum_{k>j} \sum_{j \neq i,k} y_{ij} y_{jk} \ \#$ of two-stars

ERGM Model

If we denote

$$\mathbf{q} = (\theta^{\mathrm{T}}\mathbf{s}(\mathbf{y}) + \mathbf{z}^{\mathrm{T}}\mathbf{f}(\mathbf{x}))$$

then the probability for each individual i to form link with any $j \in V$ can be rewrite as

$$\mathbb{P}(i \leftrightarrow j|\mathbf{y}_{-ij}) = \frac{e^{q_{ij}}}{1 + e^{q_{ij}}}.$$

As we have conditioned on a certain graph, we can now establish more on $\mathbf{f}(\mathbf{x})$ for this study:

$$f_{ij}(\mathbf{x}) = 1 - |x_i - x_j|$$

Setup

- There are two groups of potential customers in the market profile, young and senior, respectively.
- Young customers are more active in sharing fashion information.
- The network has a size of N. Suppose we have estimated the parameter of the sampled network.
- We have a budget to advertise $|V^a|$ of them.
- We care about two indexes: (i) How active is the topic? (ii) How many people have received the information?

Simulation

DGP & Estimation

DGP:

- $N = 80$, $|V^a| = 15$
- For sufficient statistics, the coefficients are $\theta_\nu = -5$ (edges), $\theta_\rho = 0.01$ (two-stars).
- The group index $g \in \{s, y\}$ with $\#s = \#y$, independent variable $x_s \sim \text{Uni}\,(0, 1.2)$, $x_y \sim \text{Uni}\,(0.96, 1)$
- We use the Metropolis-Hasting algorithm to generate networks.

Estimation:

- The generated networks are sampled and estimated with R package **ergm**
- For simplicity, we estimate the population network leaving the estimate unbiased without non-representative correction.

True Network

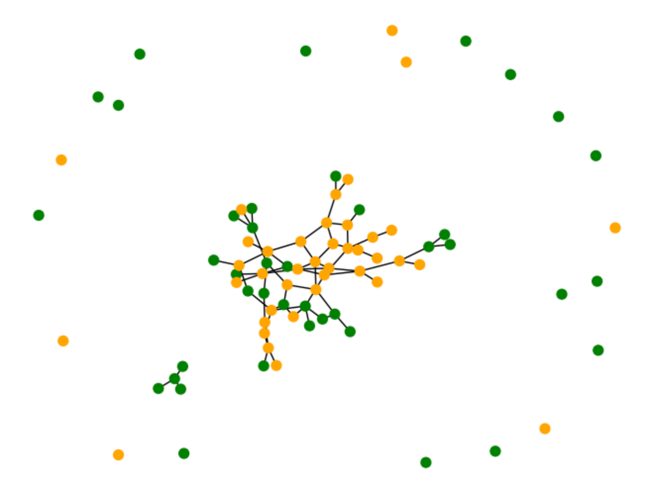

Figure 1: True network with $\theta_\nu = -5$, $\theta_\rho = 0.01$, $z = 1$. Green points refer to senior agents, whereas yellow points refer to young agents.

Is network survey always useful?

- Locally, we can identify the exact behavior of individuals.
- We've received some information about what the network looks like.
- However, beyond local, how could we target those who are not in the samples?

Example Sampling

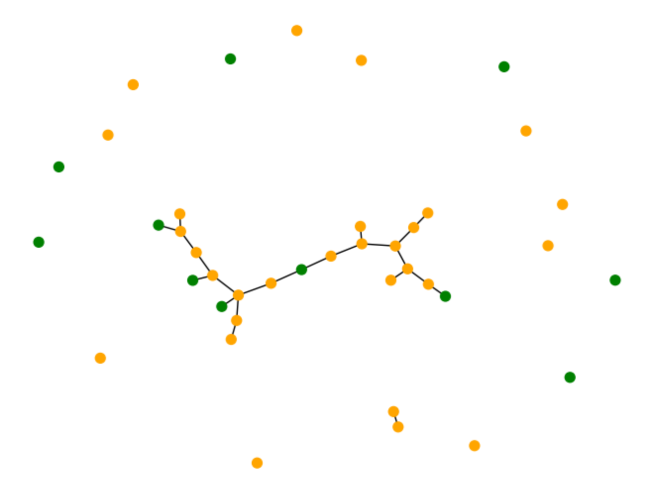

Figure 2: Sampled data from true network. Green points refer to older agents, whereas yellow points refer to young agents.

True Network (Circular)

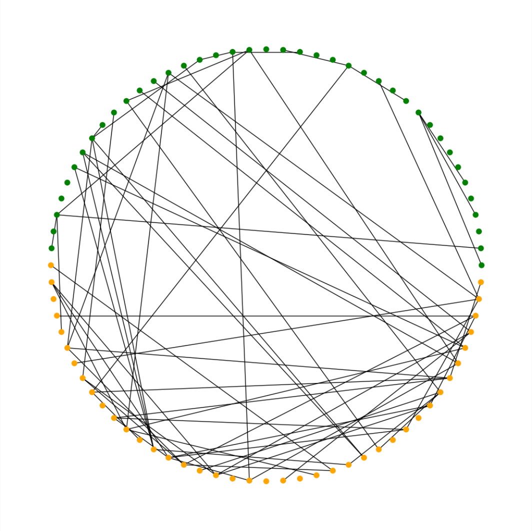

Figure 3: True network in a circular layout. The ratio of average links in the group s and y is about 0.5.

Welfare

Define the targeted nodes as $V^a \in V$

- The empirical welfare for the active level is

$$W_I = |\{\bigcup_v I(v) | v \in V^a\}|$$

- The empirical welfare for the total number who have received the information as

$$W_N = |\{\bigcup_v N(v) | v \in V^a\}|$$

We define the loss as a quadratic loss. We take the square root of loss to better present it.

Assignment Rule

- **Complete Information**: Use mixed integer programming to target the top 15 agents that connect to most nodes with most links.[1]
- **Random**: Uniformly select 15 agents from the population, and target them at the true-network (after empirical realization).
- **Group-level Information**: Calculate the ratio of the average links in two groups, and randomly sample them according to the ratio.
- **Network Information**: Reconstruct the network by the MAP of Metropolis-Hasting posterior sampling, and assign treatment with mixed integer programming.

[1] Targeting nodes are prior to links. After testing, targeting those with the most links is an efficient approximation for saving computation demands.

Targeting with Complete Information

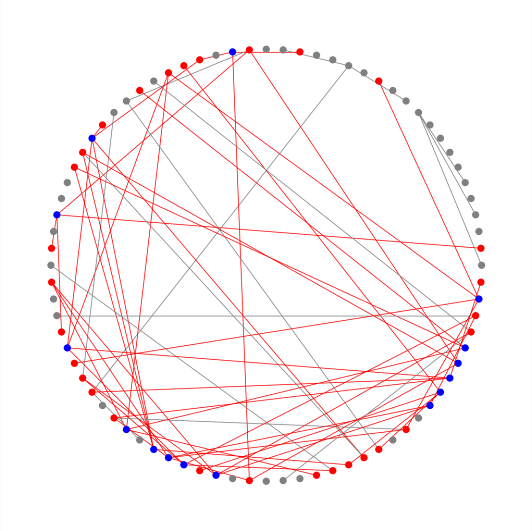

Figure 4: Optimal assignment with complete information. $W_I = 58$, $W_N = 44$.

Targeting with Random Assignment

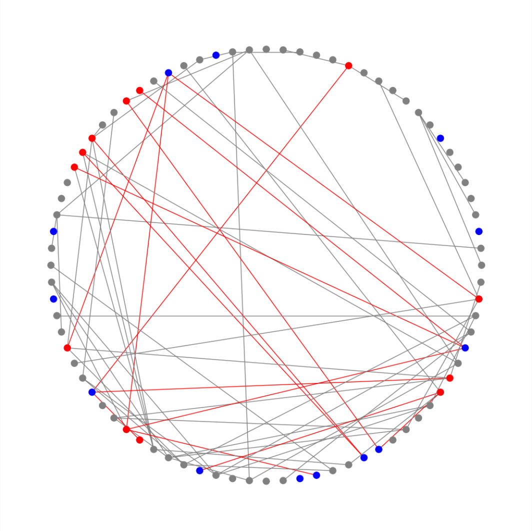

Figure 5: Random assignment. $W_I = 21$, $W_N = 28$.

Targeting with Group Information

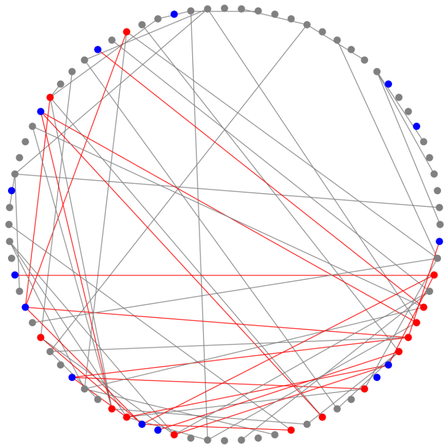

Figure 6: Assignment with group-level information. $W_I = 20$, $W_N = 25$.

Targeting with Network Information

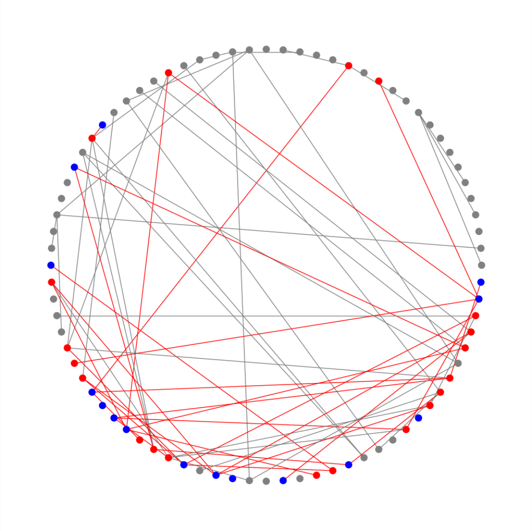

Figure 7: Assignment with network information. $W_I = 55$, $W_N = 44$.

Summary of Empirical Welfare

Table 1: Summary of Results with Different Assignment Rules

Type	W_I	R_I	W_N	R_N
Complete	66.3	–	45.2	–
Random	25.0	35.4	30.2	11.8
Group-level	27.7	32.4	30.8	10.3
Network	57.2	14.2	40.8	6.3

Summary

- Network structural information outperforms all other assignment rules in both empirical welfare estimations.
- Random assignment and structural assignment perform similarly, as the edge distribution is approximately in a random field conditional on groups.
- Our structural estimate contributes significantly in sparse networks.
- It could be improved by better characterizing the social position of individuals, i.e. higher dimensions of homophily, to reduce exchangeability.
- I did not consider the out-sample variance of network estimates, which should be further established in future works.