

Research Proposal

Polina Revina

September 22, 2024

1 Background

As the literature on social network data grows over the years, the causal inference under such data structure remains perplexed. Rubin's causal model, which many causal inference methods rely on, assumes non-interference, meaning that a unit's potential outcome doesn't depend on its neighbours' outcome [1]. However, in many observational studies, this assumption is not plausible if exposure to the outcome can be transmitted through social interaction or physical proximity of units. For instance, students who received tutoring support can interact with other students who weren't assigned to this program thus resulting in the spillover effect [2]. Another example is sharing news on the social network, which can affect other users in the group. Using methods that don't account for this or wrongly imposing this assumption can lead to an entirely wrong conclusion about the effect.

In this research, I aim to (i) generate synthetic network data using autoregression models, (ii) estimate the bias of causal inference methods that don't account for interference, and (iii) apply methods that take into consideration network effect.

2 Research plan

Data Generation Process

To generate a network I'll use the Barabasi-Albert model which is commonly used for modelling real-world systems, for instance, friendship groups or high school networks. It generates networks with a power-law degree distribution, resulting in some nodes having a much larger number of connections than others.

There are several types of network effects to simulate. Relational network effect means that the outcome of the unit depends on its covariates, as well as its neighbours' outcomes and covariates. The positional network effect depicts an effect on a network level, meaning that unit's position in network affects the outcome and the structural means the effect of the network as a whole. To simulate these scenarios we can apply spatial lag model approaches. The value of the outcome of one unit depends on the value of its neighbours, leading to dependencies between neighbours. There are also different ways to simulate treatment, it can be random or depend on unit's covariates. There are several ways to model this, using logistic regression model or some linear threshold.

References

- [1] David Roxbee Cox. "Planning of experiments." In: (1958).
- [2] Laura Forastiere, Edoardo M Airolidi, and Fabrizia Mealli. "Identification and estimation of treatment and interference effects in observational studies on networks". In: *Journal of the American Statistical Association* 116.534 (2021), pp. 901–918.