Adaptive A/B Tests on Networks with Cluster Structures

Yang Liu

George Washington University

louisliu@gwu.edu

Joint Work with Yifan Zhou, Ping Li and Feifang Hu

March 6, 2022

Background

- IT firms such as LinkedIn, Facebook, Tweeter, etc., conduct thousands of A/B tests per day to evaluate the performance of their product (Kohavi et al., 2013) .
- Their goal of the experiment is to detect the **"all versus nothing"** average treatment effect (ATE) that compares the subjects are all treated versus they are all assigned with the control.
- However, the subjects in the experiments are often involved in **social networks** which can complicate the estimation of the ATE.
- **Blessing!** The networks in A/B tests often have certain cluster structures which can be used to mimic the scenario that all of the subjects are assigned with the same treatment.

Limitations

How to do it? (Existing work): graph-cluster randomization + Horvitz-Thompson estimator, i.e. using cluster as the unit for randomization (Ugander et al., 2013; Gui et al., 2015).

- This approach requires bounded cluster sizes.
- However, the clusters detected from many community detection algorithms may not have bounded cluster sizes. Therefore, the usage of graph-cluster randomization can be restricted.

Limitations

This approach also subjects to the following two problems:

- Graph-cluster randomization may generate **severe imbalance** with respect to the **cluster-treated structures**. These structures are the cluster level measures characterizing the network features according to the node's status in its cluster.
- The value of the cluster-treated structures for one cluster may depend on both its own treatment assignment as well as the treatment assignments of its connected clusters. This poses certain **challenges for the design of A/B tests**.
- The **weights** used by the Horvitz-Thompson estimator largely depend on the randomization scheme, and can not reflect the relationship of the **between-cluster network interference**.

Our Contribution

- We show that if the ATE from the **cluster perspective** is considered, then the corresponding Horvitz-Thompson estimator can partially **alleviate the restriction** on the usage of graph-cluster randomization.
- By using an illustrative example, we show that the **imbalance of the cluster-treated structures** and the **inappropriate weights** used by the Horvitz-Thompson estimator can greatly **impair** the evaluation of the ATE. (This problem may not be resolved since the clusters are not independent due to the connection through the network.)

Our Approach

- Design Stage — Rerandomized-Adaptive Randomization
  - Use pairwise-sequential randomization (Qin et al., 2016) to balance the cluster-treated structures whose values does not depend on the assignments of other clusters.
  - Rerandomize the assignments and further select the one that also balance the cluster-treated structures whose values also depend on the assignments of other clusters.
- Estimation Stage— Cluster-Adaptive Estimator
  - propose new weights to adjust the nodes according to the nodes' treated status.

Thank you!

References I

Gui, H., Xu, Y., Bhasin, A., and Han, J. (2015). Network a/b testing: From sampling to estimation. In *Proceedings of the 24th International Conference on World Wide Web*, pages 399–409.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., and Pohlmann, N. (2013). Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1168–1176.

Qin, Y., Li, Y., Ma, W., and Hu, F. (2016). Pairwise sequential randomization and its properties. *arXiv preprint arXiv:1611.02802*.

Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. (2013). 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.