Leveraging Contact Network Information in Studies of Contagion Processes.

Mélanie Prague [1]& Patrick Staples [2] & Victor DeGruttola [2] & Jukka-Pekka Onnela [2]

[1] *Inria, SISTM Team, Inserm U1219, Bordeaux Public Health, France.*
[2] *Harvard TH Chan School of Public Health, Boston, USA*

Résumé. Evaluer l'effet d'une exposition ou intervention pour des processus contagieux est une question centrale dans le développement de nouvelles méthodes de contrle d'épidémie ou de propagation d'informations. Nous nous intéressons à des processus de contagion sur des réseaux, le phénomène de transmission ne peux avoir lieu que lorsque deux individus 'contaminé' et non 'contaminé' sont connectés par un lien. Nous nous intéressons à l'usage de l'information issu du réseaux sous-jacent de connection d'individus comme variable d'ajustement, soit en tant que facteur de confusion soit en tant que facteur permettant une meilleure efficacité de l'estimation. Nous utilisons un estimateur augmenté doublement robuste basé sur des équations d'estimation généralisées (GEE) pour évaluer dans quelle mesure la correction du bias et le gain d'efficacité dépend de la structure du réseaux et des caractéristiques du processus de contagion. Nous appliquons notre approche pour estimer l'effet de plusieurs variables d'exposition sur la diffusion d'un programme de microfinance dans des villages de Karnataka en Inde.
Mots-clés. GEE, Processus Contagieux, Réseaux, Statistiques semi-paramétriques

Abstract. Evaluate the effect of an exposition or intervention against a contagion process is a central question in the development of new methods for epidemic control or information propagation. We are interested in contagion processes operating on a contact network, transmission can only occur through ties that connect contaminated and non-contaminated individuals. We investigate the use of contact network features as both confounders and efficiency covariates in exposure effect estimation. Using doubly-robust augmented generalised estimating equations (GEE), we estimate how correction of bias and gains in efficiency depend on the network structure and characteristics of the contagion agent. We apply this approach to estimate the effects of various exposures on the spread of a microfinance program in a collection of villages in Karnataka, India.
Keywords. Contagion process, GEE, Network, Semi-parametric statistics

1 Statistical methodology

1.1 Networks

Most of the time, clustered trials are used to evaluate exposure or intervention effect for contagion outcomes. A cluster can be seen as a network which is itself a collection of networks that can or cannot be connected. Each individual is a node in these networks and each tie is a link between two individuals. We identify a cluster as an independent collection of network, such that outcomes within a network are correlated whereas those across networks (across clusters) are independent. See Figure 1.

Complete network and contagion processes are not generally observable; but even when they are not, it may nevertheless be possible to characterise certain features of processes and/or networks. In this analysis, we considered individual level network features (degree, mean neighbour degree, size of individual's component), cluster-level network features (assortativity, size of largest component, mean component size, number of component) and contagion features (total neighbour infections at baseline, total number of infections at baseline in the cluster, length of the path to the closest infected connected individual).

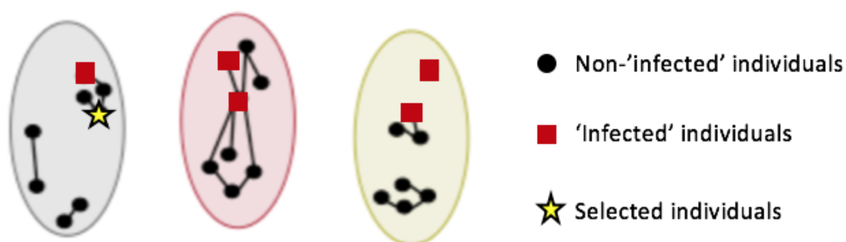

Figure 1: Schematic of three clusters at baseline composed of collection of networks. Between clusters, contagion cannot happen. Within clusters, contagion outcomes are correlated. The first cluster on the left on this figure is composed of 8 individuals and is a collection of three clusters. For the selected individual in starred yellow, the degree is 2, mean neighbour degree is 3/2, size of individual's component 4. For this cluster, assortativity is -0.57, size of largest component is 4, mean component size is 8/3, number of component is 3. Finally, regarding contagion process, total neighbour infections at baseline is 0, total number of infections at baseline in the cluster is 1 and length of the path to the closest infected connected individual is 2.

1.2 Estimation of the effect of an intervention on an outcome

We are interested in computing the marginal effect of exposure or intervention A defined as $\beta = E(Y|A = 1) - E(Y|A = 1)$, Y being the contagion outcome. The generalised estimating equations (GEE) approach provides a general approach for analysing correlated outcomes that: i) is more robust to variance structure misspecification, ii) relies less on parametric assumptions than the standard likelihood methods, and iii) provides population level conclusions on the effect of an exposure on an outcome. The estimator is described in Prague et al. 2016 and further extended in Staples et al. 2019.

The clustered trial may or may not be randomised. in this work, we consider a situation where the data are observational and the exposition is not randomised. In order to estimate the marginal effect of the intervention, we adjust for network feature at two levels: 1/ in the propensity score for probability of exposition status to correct for possible bias, and 2/ in the augmentation model to account for network features that can help in increasing the efficacy of the estimator.

2 Preliminary results

We used simulations to evaluate the validity and properties of our approach. The network generation model in our simulation study is the degree-corrected stochastic block model with degree correlation (Karrer et al. 2011). Each simulation setting was replicated R times. We show that in most settings, adjusting for contact network features and baseline contagion reduced bias and yielded a considerable reduction in RMSE[1] across a range of simulation settings. This result also hold for the setting of simulation similar to the one exhibited in the Karnataka microfinance program study (Banerjee et al. 2013), *i.e.* high average degree of individuals in population, powerlaw degree distribution, assortative community with block structure, contagion process with degree infectivity and low baseline prevalence of contamination.

In this study, the microfinance institution began by inviting "leaders" (*e.g.*, teachers and shopkeepers) to informational meetings and asked them to spread information about the loans to villagers. Each leader could or not take the microfinance at his own discretion. The outcome of interest is a binary variable defined at the individual-level which is equal to one if the individual decided to take the microfinance loan. We are interested in the effect of proportion of leaders on the diffusion of microfinances. As it, the exposure is defined as cluster-level and binary. It is set as high if the fraction of households containing leaders in a cluster is in the top quartile compared to all other clusters. We calculate and

[1] defined as $\sqrt{\frac{1}{R}\left[\widehat{\text{Bias}}(\hat{\beta}_r)^2 + \widehat{\text{Var}}(\hat{\beta}_r)\right]}$

compare the average difference in microfinance uptake between village exposure groups. A crude estimate of this quantity is obtained as the difference between exposed and unexposed villages in average household uptake of microfinance. The standard GEE accounts for the correlation of outcomes within each village. The network-adjusted GEE accounts for the fact that we are interested in a contagion process spreading on a network. Results are presented in Table 1.

Method	Estimate	S.E.	p-value
Crude Estimate	-0.002	0.039	0.478
Unadjusted GEE Estimate	-0.069	0.058	0.120
Network-adjusted GEE Estimate	-0.138	0.031	0.001

Table 1: Estimates of the effect of high fraction of village leaders in a cluster on microfinance uptake for different adjustment strategies. β is the average mean difference in uptake comparing exposed to unexposed clusters.

We managed, using network-adjusted GEE, to exhibit a significant difference between clusters with different number of leaders where other methods fail. We show that a high fraction of leaders is likely to decrease the microfinance uptake, which can be further understood by the fact that each leader making his own decision, the overall influence of each leader is diluted in the population when there is a high number of leaders.

Bibliographie

M. Prague, R. Wang, A. Stephens, E. Tchetgen Tchetgen, and V. DeGruttola. Accounting for interactions and complex inter-subject dependency in estimating treatment effect in cluster-randomized trials with missing outcomes. Biometrics, 2016 72(4): 1066-1077.

B. Karrer and EJ. Newman Stochastic blockmodels and community structure in networks. Phys. Rev. E, 83:016107, Jan 2011.

P. Staples, M. Prague, V. deGruttola and JP. Onnela Leveraging Contact Network Information in Clustered Observational Studies of Contagion Processes. Arxiv, 2019.

A. Banerjee, AG. Chandrasekhar, E. Duflo, and MO. Jackson. The diffusion of microfinance. Science, 341(6144):1236498, 2013.