The World Bank Economic Review, 34(Supplement), 2020, S35–S39
doi: 10.1093/wber/lhz038
Supplementary Article

Representativity and Networked Interference in Data-Rich Field Experiments: A Large-Scale RCT in Rural Mexico

Alejandro Noriega and Alex Pentland

Abstract

Modern availability of rich geospatial datasets and analysis tools can provide insight germane to the design of field experiments. Design of field experiments, and in particular the choice of sampling strategy, requires careful consideration of its consequences on the external representativity and interference (SUTVA violations) of the experimental sample. This paper presents a methodology for a) modeling the geospatial and social interaction factors that drive interference in rural field experiments; and b) eliciting a set of nondominated sample options that approximate the Pareto-optimal tradeoff between interference and external representativity, as functions of sample choice. The study develops and tests the methodology in the context of a large-scale health experiment in rural Mexico, involving more than 3,000 pregnant women and 600 health clinics across 5 states. Relevant for the practitioner, the methodology is computationally tractable and can be implemented leveraging open sourced geo-spatial data and software tools.

JEL classification: I00, C1, C6, C8, C9

Keywords: field experiments, networked interference, SUTVA violations, big data, spatial analysis, cash transfers

1. A Large-Scale Digital Health Experiment in Rural Mexico

Mexico's social assistance program Prospera is one of the largest conditional cash programs in the world, providing health care to nearly 30 million beneficiaries (Oportunidades 2011). The physical remoteness of Prospera's rural beneficiaries drives key challenges in provisioning its services. Moreover, information in Prospera flows through traditional means such as fliers, radio announcements, and door-to-door communication.

National authorities have endeavored to introduce digital means of communication with, and among, Prospera beneficiaries.[1] In this context, this study participated in designing a large-scale randomized control trial (RCT) to assess the effect of such potential interventions on health outcomes. The experiment focuses on maternal and child health; involves more than 600 health clinics and 3,000 pregnant women

Alejandro Noriega (corresponding author) is a visiting scholar at the MIT Human Dynamics Laboratory, and founder of Prosperia Labs; his email address is anc@prosperia.ai. Alex Pentland is director of the MIT Human Dynamics Laboratory, Academic Director of the Harvard-MIT-ODI DataPop Alliance, and Faculty Director of the MIT Connection Science Research Initiative, Cambridge, MA; his email address is pentland@mit.edu.

1 In partnership with a set of academic institutions and NGOs, such as the MIT Media Laboratories.

across 5 states; and tests 3 treatment arms, consisting of top-down, peer-to-peer, and down-to-top feedback communication.

2. External Representativity and Interference in Sample Choice

This paper focuses on extrapolation in the common setting where random sampling is not viable in practice, such as in the Prospera experiment. As pointed out by Muller (2015), sampling at random is commonly "not practically feasible, or researchers have the more ambitious aim of generalizing beyond a single, prespecified population."

Following (Imai, King, and Stuart 2008) decomposition of the population average treatment effect (PATE), nonrandom sampling fails to ensure that that $E[\Delta_{SU}] = 0$, where Δ_{SU} denotes the sampling estimation error due to unobservables. Extrapolation is, however, possible leveraging the set of observed covariates X, particularly in rich covariate settings.[2] In the case of Prospera, rich census and institutional data are available, providing more than 200 covariates at the clinic, village, household, and individual levels; including education, indigenism, newborn weight and measures, birth defect rates, clinics' equipment, and so on.

On the one hand, today's age of data fosters the ability to extrapolate by means of rich covariate sets. On the other, a natural strategy for coping with interference is to choose an experimental sample where units, or clusters, are most apart from each other interference-wise. However, interference-minimizing criteria for sample selection may compromise representativity to the population of interest; and, conversely, common sampling approaches amenable to extrapolation, such as sampling for heterogeneity (Muller 2015) and proportional sampling (Chen, Tse, and Yu 2001), are likely to misalign against interference-minimizing sample selection.

Section 4 introduces a methodology for eliciting the potential tradeoff among these two objectives, and providing the researcher with a set of Pareto nondominated sample options to decide from.

3. Modeling Interference Networks Using Geospatial Data and GIS Tools

Interference Gravity Model

This study proposes and implements a simple class of model where interference between two social clusters—such as villages, schools, or health clinics—is driven positively by the density of experimental subjects relative to population, and negatively by the distance between them. Gravity models of this type have a long-standing history in economics and the social sciences in capturing spatially mediated social interactions.[3] Let dyadic interference between i and j be modeled as

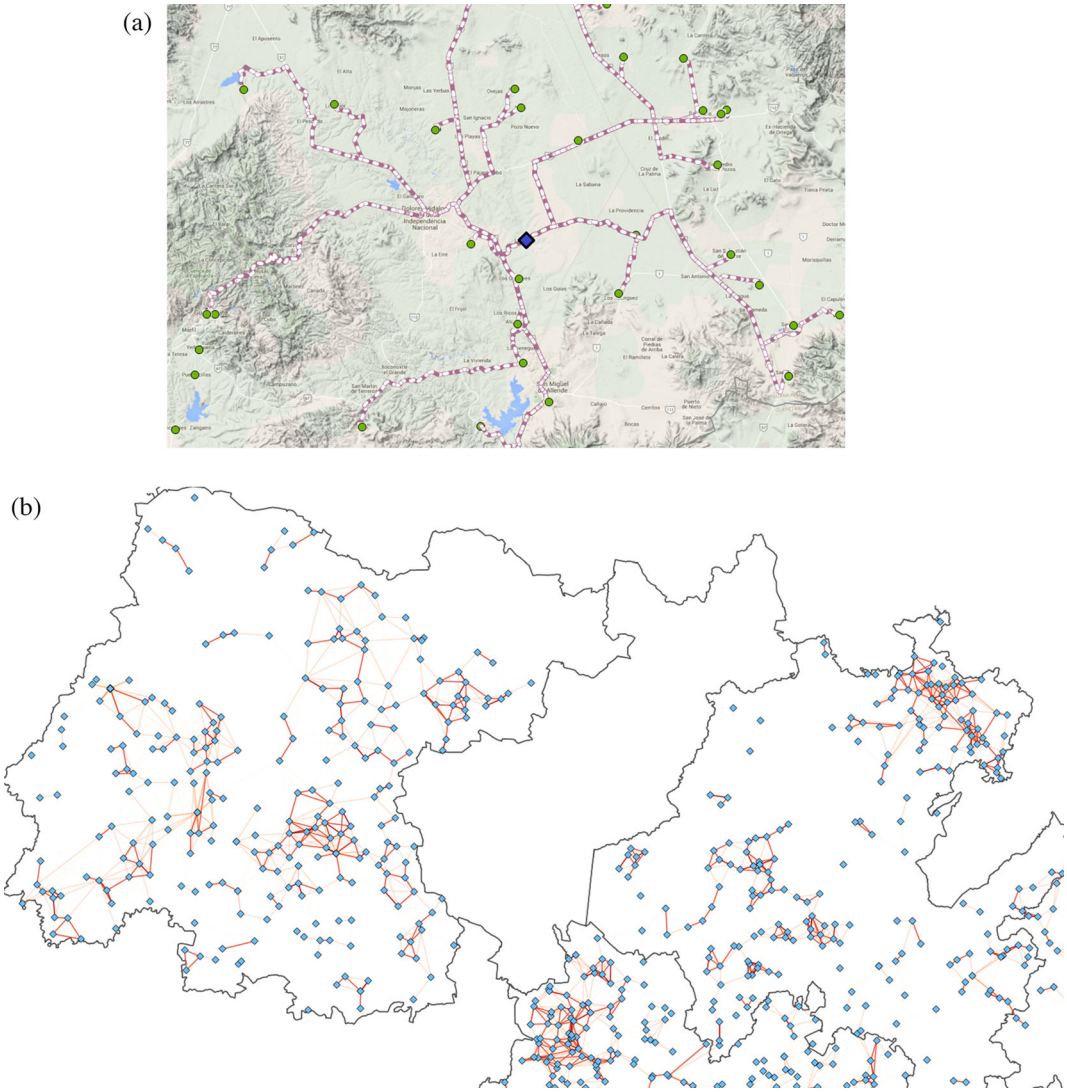
$$I_{ij} = \frac{f(m_i, m_j)}{g(d_{ij})} = \frac{am_i m_j}{d_{ij}^b} \tag{1}$$

where d_{ij} denotes distance between clusters i and j, and m_i denotes i's population density mass, defined as the ratio of the number of experimental subjects in cluster i to its total population. Traditional formulations of gravity models instantiate $f = am_i m_j$ and $g = {d_{ij}}^b$, with $a, b \in R^+$. Functions f and g allow for

2 See Bareinboim and Pearl (2015); Hartman et al. (2015) for formal frameworks and use cases of extrapolation leveraging covariates.

3 From trade (Bergstrand 1985; Deardorff 1998) and migration flows (Ravenstein 1989; Karemera, Oguledo, and Davis 2000), to transportation flows (Erlander and Stewart 1990) and epidemics (Xia, Bjørnstad, and Grenfell 2004).

Figure 1. Panel (a): Paths of Pregnant Women (purple lines) from Their Home Village (green circles) to Their Assigned Health Clinic (blue diamond). Panel (b): InterClinic Interference Network (graded orange lines) for all Prospera Clinics over the States of Guanajuato, Hidalgo and Mexico State.



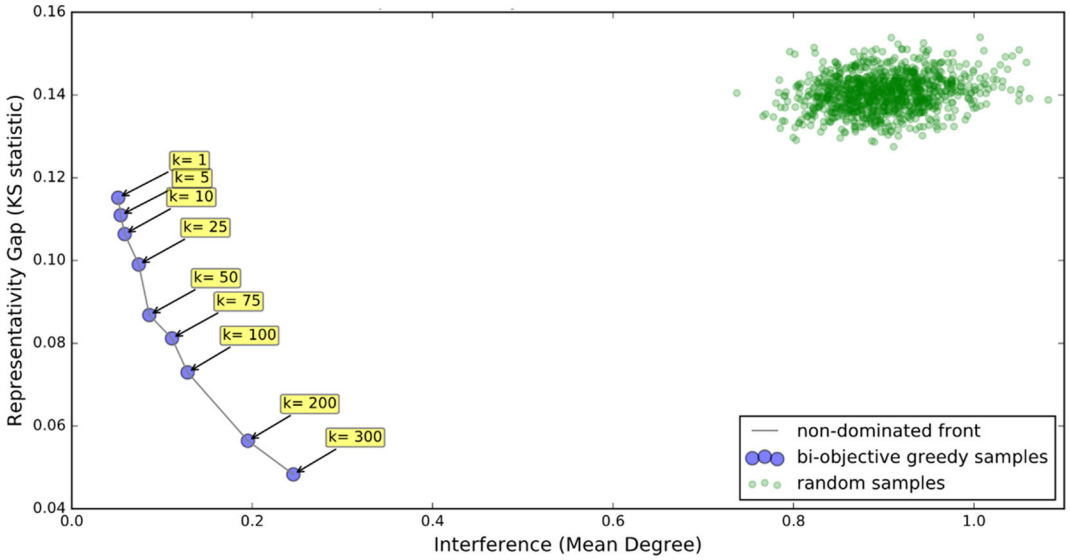
generalizations of the model (Anderson 2010), which can be informed by substantive knowledge or prior studies in the field.[4]

Walk-and-Road Distances

In the context of the Prospera experiment, this study moved from unrealistic straight-line geographic distances to a more meaningful metric of walk-and-road route distance: defined as the shortest routing distance connecting two points, accounting for the possibility of walking route segments outside of the road

4 In the case of the Prospera experiment, threshold functions were used in f and g so that interference was meaningless for distances above and densities below thresholds set with the assistance of field experts. Extensions are also available for gravity analysis where the explicit location of individual units within clusters is deemed relevant.

Figure 2. Pareto Front of the Prospera Experiment, Approximated by a Greedy Bi-Objective Search



network. The analysis was implemented using public geospatial and demographic data, and open-sourced GIS tools (see Noriega and Pentland 2016 for details). Figure 1 shows the resulting interference network.

4. The Interference vs. Representativity Tradeoff

This study is interested in understanding how different choices of experimental sample affect interference and representativity, and ultimately the effect of these on the experiment's power to identify its inferential targets. In the Prospera experiment, this study is interested in selecting a sample of 600 health clinics, out of an eligible pool of 1,700 clinics. However, even for small networks it is unfeasible to conduct power simulations exhaustively over the sample space, which grows exponentially on sample size n_s. Moreover, random samples perform poorly, as shown in fig. 2.

This study implemented a simple bi-objective greedy search heuristic, which searches for sample options that minimize both interference and representativity gap, with the goal of eliciting a set of Pareto nondominated options, that is, sample options that constitute the tradeoff between objectives.[5] Details on the greedy algorithm can be found in Noriega and Pentland (2016).

The greedy heuristic is parameterized on k, which controls the importance given to interference versus representativity in the optimization process. Results associated to a set of different k values provide a set of options that trade off between objectives. Figure 2 shows the set of Pareto nondominated samples elicited for the Prospera experiment, where mean interference values range in the [.05, .25] interval, and representativity gap ranges in the [.05, .12] interval.

This small set of nondominated sample options is amenable for researchers to perform appropriate power calculations and sensitivity analysis, and to choose the sample that best fits research objectives and context of the study.

5 Representativity gap is measured by the Kolmogorov-Smirnoff (KS) distance between the sample and target covariate distributions, analogous to measures of covariate balance in the context of matching (see Diamond and Sekhon 2013).

5. Conclusion

The Prospera experiment implementation commenced in January 2016, and will remain active for 12 months. Near-future work will study observed spillovers in the experiment, and the extent to which the geospatial interference model and its different possible parameterization explain them. As pointed out by Gerber and Green (2012), the study of the existence and nature of spillover effects is of paramount importance, as it provides relevant insights for conducting subsequent research studies, as well as for the design of policy itself.

References

Anderson, J. E. 2010. "The Gravity Model." Working Paper 16576, National Bureau of Economic Research, Cambridge, MA.

Bareinboim, E., and J. Pearl. 2015. "Causal Inference from Big Data: Theoretical Foundations and the Data-Fusion Problem." Tech. Rep., DTIC Document.

Bergstrand, J. H. 1985. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Review of Economics and Statistics* 67 (3): 474–81.

Chen, T. Y., T. Tse, and Y.-T. Yu. 2001. "Proportional Sampling Strategy: A Compendium and Some Insights." *Journal of Systems and Software* 58 (1): 65–81.

Deardorff, A. 1998. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" In *The Regionalization of the World Economy*, edited by J. A. Frankel, 7–32. Chicago: University of Chicago Press.

Diamond, A., and J. S. Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 95 (3): 932–45.

Erlander, S., and N. F. Stewart. 1990. *The Gravity Model in Transportation Analysis: Theory and Extensions*. Topics in Transportation. Vol. 3. Boca Raton, FL: CRC Press.

Gerber, A., and D. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

Hartman, E., R. Grieve, R. Ramsahai, and J. S. Sekhon. 2015. "From SATE to PATT: Combining Experimental with Observational Studies to Estimate Population Treatment Effects." *Journal of the Royal Statistical Society*, 178 (3): 757–778.

Imai, K., G. King, and E. A. Stuart. 2008. "Misunderstandings between Experimentalists and Observationalists about Causal Inference." *Journal of the Royal Statistical Society*, Series A (Statistics in Society) 171 (2): 481–502.

Karemera, D., V. I. Oguledo, and B. Davis. 2000. "A Gravity Model Analysis of International Migration to North America." *Applied Economics* 32 (13): 1745–55.

Muller, S. M. 2015. "Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations." *World Bank Economic Review* 29 (Supp. 1): S217–25.

Noriega, A., and A. Pentland. 2016. "Balancing External Representativity and Networked Interference in Large-Scale Rural Experiments." http://pubdocs.worldbank.org/en/880861466186946430/World-Bank-ABCDE-2016-Noriega-and-Pentland-v6-ABCDE-Representativity-and-interference-in-large-scale-rural-experiments.pdf.

Oportunidades, M. 2011. "Mexico's Targeted and Conditional Transfers: Between Oportunidades and Rights." *Economic & Political Weekly* 46 (21): 49–54.

Ravenstein, E. G. 1889. "The Laws of Migration." *Journal of the Royal Statistical Society* 52 (2): 241–305.

Xia, Y., O. N. Bjørnstad, and B. T. Grenfell. 2004. "Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics." *American Naturalist* 164 (2): 267–81.