Inferring the Effects of Observed Covariates in Latent Space Models for Networks

Zachary Jones, Matthew Denny, Bruce Desmarais, Hanna Wallach

June 22, 2017

Abstract

Due to the complex dependencies found in network data, scholars draw upon an increasingly sophisticated toolkit for statistical inference. The latent space model (LSM) models the edges in a network via a combination of the generalized linear model (GLM) and a latent spatial embedding of the network's nodes. In many applications, researchers have assumed that this embedding can control for unmeasured confounding network structure. However, there has been little research that considers whether this is indeed the case. Via a simulation study, we investigate the LSM's ability to control for unmeasured confounding network structure. We find that under even moderate confounding, the LSM does not exhibit lower estimation error or inference error than the GLM; it does, however, exhibit substantially lower prediction error. We therefore conclude that the LSM is most appropriate for predictive or exploratory analyses.

1 Introduction

Inferential analyses of network data have grown increasingly sophisticated in recent years. Scholars are well-versed in the risks associated with ignoring unmeasured confounding network structure. If not accounted for, dependencies such as reciprocity, transitivity, and homophily can lead to biased estimates of covariate effects and to hypothesis testing errors, in much the same way that omitted-variable bias can affect conventional regression analyses (Ward, Siverson, and Cao, 2007; Kinne, 2014; Cranmer and Desmarais, 2017; Hays, Kachi, and Franzese, 2010). Researchers have therefore proposed a number of statistical models to account for unmeasured confounding network structure, including the exponential random graph model (ERGM; e.g., Lazer, Rubineau, Chetkovich, Katz, and Michael, 2010; Cranmer and Desmarais, 2011; Desmarais and Cranmer, 2012), the stochastic actor-oriented model (SAOM; e.g., Berardo and Scholz, 2010; Kinne, 2014), and the latent space model (LSM; e.g., Ward et al., 2007; Ward and Hoff, 2007; Kirkland, 2012).

The ERGM and the SAOM both account for dependencies using an approach similar to that used to control for unmeasured confounding variables in regression analyses. The researcher specifies a set of dependencies that (s)he hypothesizes to be important to the network. These dependencies are then explicitly included in a model that simultaneously represents the observed covariates of interest. In contrast, the LSM takes a different approach and uses a latent spatial embedding to control for unmeasured confounding network structure. The LSM therefore has an advantage over the ERGM and the SAOM in that the researcher does not need to hypothesize a set of dependencies that may be important. However, this advantage is contingent upon the LSM's ability to identify confounding network structure that would otherwise be attributed to the observed covariates.

Despite the growing popularity of these models, there have been few studies that investigate their performance at accounting for unmeasured confounding network structure. In this article, we focus on the LSM and study its ability to reduce estimation error and inference error using a latent spatial embedding of network structure.

TODO: Something about type I error.

2 The Latent Space Model

What the LSM is and why it was developed.

What the LSM is used for (currently in 1, 1.1, and 2).

Concerns re. its use, type I and II error, type II addressed.

3

Acknowledgements

This work was supported in part by National Science Foundation (NSF) grants DGE-1144860, SES-1558661, SES-1619644, and CISE-1320219. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect those of the sponsor.

References

Berardo, R. and J. T. Scholz (2010). Self-organizing policy networks: Risk, partner selection, and cooperation in estuaries. *American Journal of Political Science 54*(3), 632–649.

Cranmer, S. J. and B. A. Desmarais (2011). Inferential network analysis with exponential random graph models. *Political Analysis 19*(1), 66–86.

Cranmer, S. J. and B. A. Desmarais (2017). A critique of dyadic design. Accepted to *International Studies Quarterly*.

Desmarais, B. A. and S. J. Cranmer (2012). Micro-level interpretation of expoential random graph models with application to estuary networks. *Political Studies Journal 40*(3), 402–434.

Hays, J. C., A. Kachi, and R. J. Franzese (2010). A spatial model incorporating dynamic, endogenous network interdependence: A political science application. *Statistical Methodology 7*(3), 406–428.

Kinne, B. J. (2014). Dependent diplomacy: Signaling, strategy, and prestige in the diplomatic network. *International Studies Quarterly 58*(2), 247–259.

Kirkland, J. H. (2012). Multimember districts' effect on collaboration between US state legislators. *Leglislative Studies Quarterly 37*(3), 329–353.

Lazer, D., B. Rubineau, C. Chetkovich, N. Katz, and N. Michael (2010). The coevolution of networks and political attitudes. *Political Communication 27*(3), 248–274.

Ward, M. D. and P. D. Hoff (2007). Persistent patterns of international commerce. *Journal of Peace Research 44*(2), 157–175.

Ward, M. D., R. M. Siverson, and X. Cao (2007). Disputes, democracies, and dependencies: A reexamination of the Kantian peace. *American Journal of Political Science 51*(3), 583–601.