A new framework for modeling sparse networks that makes sense (and can actually be fit!)

Neil Spencer
Carnegie Mellon University

Abstract

Latent position models are a versatile tool for modeling network data. Applications include clustering entities, network visualization, and controlling for unobserved causal confounding. In these models, the nodes' latent positions are typically treated as independent and identically distributed random variables. However, this assumption implies the average node degree grows linearly with the number of nodes, which is inappropriate when the network is thought to be sparse. In the first part of this talk, I will propose an alternative assumption—that the latent positions are generated according to a Poisson point process—and show that it is compatible with various levels of network sparsity. I will also provide theory establishing that the nodes' latent positions can be consistently estimated, depending on the level of sparsity. In the second part of the talk, I will address the computational challenge of fitting latent position models to large datasets. I will describe a new Markov chain Monte Carlo strategy—based on a combination of split Hamiltonian Monte Carlo and Firefly Monte Carlo—that mixes much faster than the standard Metropolis-within-Gibbs algorithm, especially for large sparse networks. Throughout the talk, I will use an advice-seeking network of elementary school teachers within a school district as a running example.