Community Network Auto-Regression for High-Dimensional Time Series

Abstract

Modeling responses on the nodes of a large-scale network is an important task that arises commonly in practice. This paper proposes a community network vector autoregressive (CNAR) model, which utilizes the network structure to characterize the dependence and intra-community homogeneity of the high dimensional time series. The CNAR model is fundamentally different from the network vector autoregressive (Zhu et al., 2017, NAR) model in that the CNAR is a population-level generative model while the NAR only utilizes a realized sample adjacency matrix. The CNAR model greatly increases the flexibility and generality of the NAR model by allowing heterogeneous network effects across different network communities. In addition, the non-community-related latent factors are included to account for unknown cross-sectional dependence. The number of network communities can diverge as the network expands, which leads to estimating a diverging number of model parameters. We obtain a set of stationary conditions and develop an efficient two-step weighted least-squares estimator. The consistency and asymptotic normality properties of the estimators are established. Theoretical results show that the two-step estimator improves the one-step estimator by an order of magnitude when the error admits a factor structure. The advantages of the CNAR model are further illustrated on a variety of synthetic and real datasets.

KEY WORDS: Network autoregression; Community structure; Common latent factors; High dimensional time series; VAR model.