Link Prediction Tools for Networked Economic and Financial Systems

Alessandro Spelta (University of Pavia) and Simone Giansante (University of Palermo)

PAVIA
27 September 2022
2-p.m. Common Room at the Department of Economic and Management Via San Felice 5
On-line
https://us02web.zoom.us/j/8087337537

ID riunione: 808 733 7537

The data in different analysis applications such as social networks, economic and financial networks and web analysis consists of relationships, which can be considered as links, between objects. These relationships can be modeled as a graph, where nodes correspond to the data objects (e.g., people) and edges correspond to the links (e.g., a phone call was made between two people). The link structure of the resulting graph can be exploited to detect underlying groups of objects, predict links, rank objects according to importance etc…

In particular, link prediction is the problem of predicting the presence or absence of edges between nodes of a graph. There are two types of link prediction: (i) *structural*, where the input is a partially observed graph, and we wish to predict the status of edges for unobserved pairs of nodes, and (ii) *temporal*, where we have a sequence of fully observed graphs at various time steps as input, and our goal is to predict the graph state at the next time step.

Structural models

Recent developments in the field of financial networks and matrix completion have provided new techniques to forecast the missing bilateral positions from the aggregated data disclosed in the institutions financial reports. The matrix completion task in financial networks aims at finding matrices that satisfy the row and column sum constraints along with other specific topological properties, such as network density, degree distributions, etc.

Two main classes of matrix reconstruction methods have been empirically tested. The first one refers to the *iterative* methods in which transformations/rescaling procedures are applied to initial guesses of the matrix to meet the aggregated assets and liabilities constraints. The second category refers to the *sampling* methods using heuristics, mainly Monte Carlo sampling, to generate an ensemble of reconstructed financial networks that satisfy the constraints, either for each proposed matrix or on average across the ensemble. Within the sampling method, a new statistical mechanics approach to the problem of financial network reconstruction will be presented. The new technique solves the matrix completion problem in financial networks when both aggregated constraints, density and bilateral netting strategy are considered. We provide empirical evidence of the misrepresentation of systemic risk assessments of financial derivatives networks when net positions are discarded, both at the system level in terms of system instability and at the individual level in terms of individual probabilities of default.

Temporal models

Dynamic interactions over time introduce another dimension to the challenge of mining and predicting link structure. Here we consider the task of link prediction in time. Given link data for T time steps, can we predict the relationships at time T+1?

Latent feature models derived from Non-negative Matrix Factorization (NMF), Singular Value Decomposition (SVD) or Katz method has proven to be highly accurate in previous work on link prediction, but this matrix-based methods collapse the data into a single matrix by summing (with and without weights) the matrices corresponding to the time slices.

Here we propose optimal transport-based and tensor-based methods, such as the CANDECOMP/PARAFAC (CP) decomposition and the DEDICOM decomposition, which does not collapse the data but instead retains its natural three-dimensional structure.

Tensor factorizations are higher-order extensions of matrix factorizations that capture the underlying patterns in multi-way data sets and have proved to be successful in diverse disciplines including chemometrics, neuroscience and social network analysis.

The link forecast method based on Optimal Transport instead learns the underlying link dynamics among a set of nodes through Wasserstein barycentric coordinates obtained as the optimal Frechet mean of the adjacency matrices describing the network evolution, in the space of probability measures endowed with the Wasserstein metric. Once the learned generative mechanism is established through the Wasserstein regression model, which determines the relationship between nodes in the evolving network, the Wasserstein barycentric coordinates are exploited to predict the future network configuration.

Link forecast methods are applied to different socio-economic and financial dataset such as correlation networks derived from stock price time series, similarity networks based on cosine distances between mortality data and Foreing Direct Investment data.