Hierarchical models of complex networks

Mikkel N. Schmidt, Morten Mørup, and Tue Herlau
Technical University of Denmark

Introduction

Networks or graphs of relations (edges) between entities (nodes) occur in many areas of science, for example in sociology, representing friendships between people; in economy, representing trade relations; and in biology, representing interactions between proteins. Many of these real-world networks exhibit hierarchical organization, in the sense that the complex structure of edges between vertices can be well described by a latent hierarchical structure underlying the network. Generative nonparametric Bayesian methods can be used to model latent hierarchical structure in complex networks, and make it possible to infer whether or not hierarchical structure is present.

A general generative model

We represent a complex network with N vertices by an $N \times N$ adjacency matrix X. In the case of an undirected binary network X is a symmetric binary matrix where $X_{ij} = 1$ indicates that an edge exists between node i and j. The following general outline of a probabilistic generative process can be used to characterize a complex network with a hierarchical cluster structure. Several existing hierarchical network models [2,3,5] are special cases of this approach.

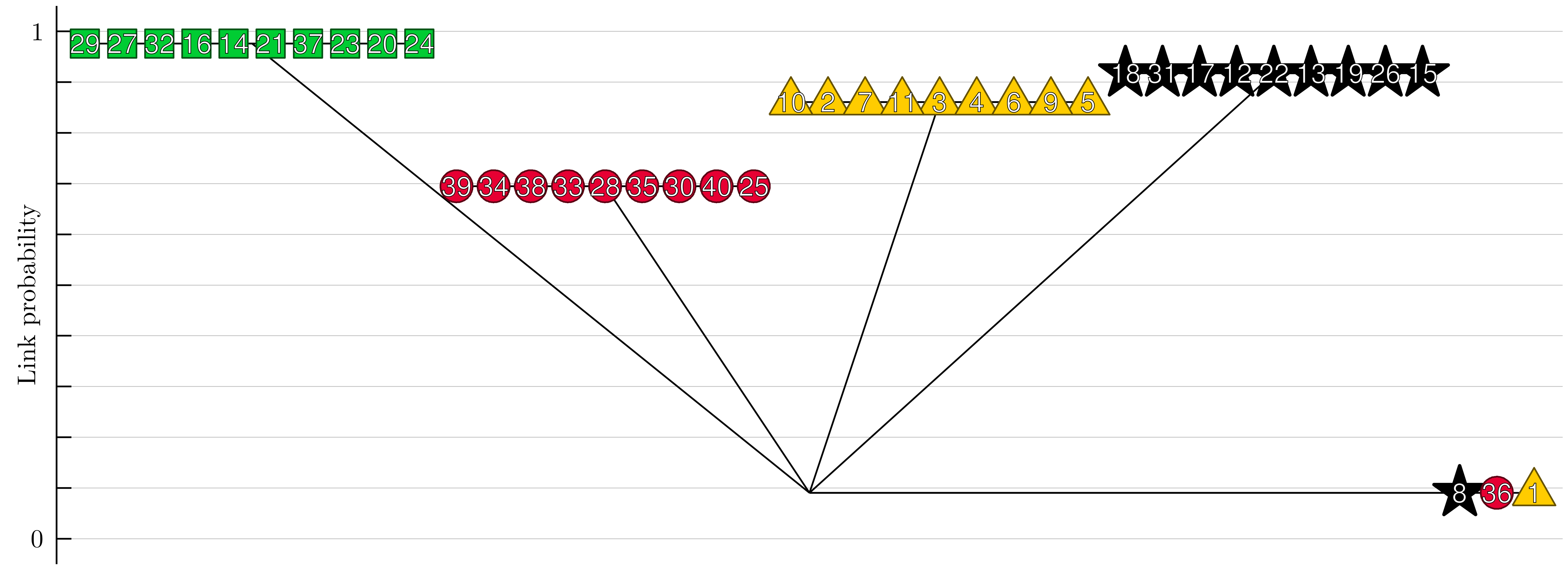
1. Generate a rooted tree where the leaf nodes corresponds to the vertices in the complex network,
$$T \sim p(T|\tau), \tag{1}$$
where τ are parameters of the prior distribution of the tree. Each internal node in the tree corresponds to a cluster of network vertices.
2. For each internal node r in the tree, generate parameters that govern the probabilities of edges between each of its children.
$$R_r \sim p(R_r|T, \rho), \tag{2}$$
where ρ are parameters of the prior distribution of the edge probabilities.
3. For each pair of vertices i and j in the network, generate an edge with probability governed by the parameters located at the common ancestral nodes in the tree
$$X_{ij} \sim p(X_{ij}|T, R). \tag{3}$$

Inference in such models entails finding the posterior distribution of the tree as well as the edge probability parameters given the observed complex network.

Relation to previous work

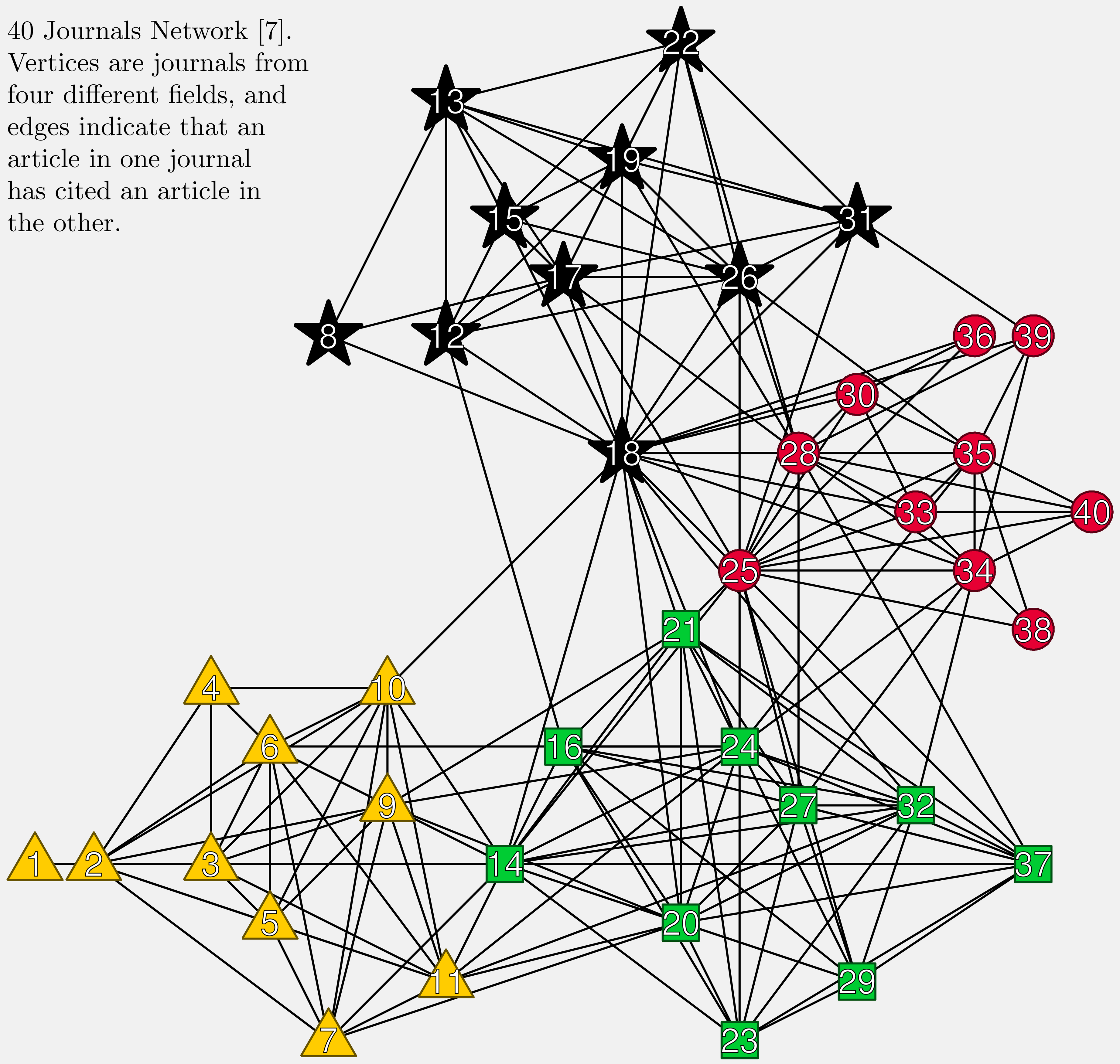
Hierarchical random graph [2]
A flat prior over binary trees (dendrograms) is used. Each node in the tree has a probability parameter, and network vertices connect with probability found at the lowest common ancestral node in the tree.
Learning annotated hierarchies [5]
A flat prior over binary trees (dendrograms) is used. Each node in the tree has a weight parameter. A tree-consistent partition is generated with probability governed by the weights. Each combination of clusters in this partition is assigned a probability parameter, and network edges within and between these clusters are generated with these probabilities.
Tree-like infinite relational model [3]
In its most simple formulation the prior over trees is constructed such that the partitioning of the vertices on the level below the leaf nodes in the tree are generated from a Chinese restaurant process, and the tree below this level is generated from a flat prior over multifurcating trees. In this model, the likelihood is identical to that of Clauset et al. [2].

Result: Hierarchical clustering



Data: Complex network

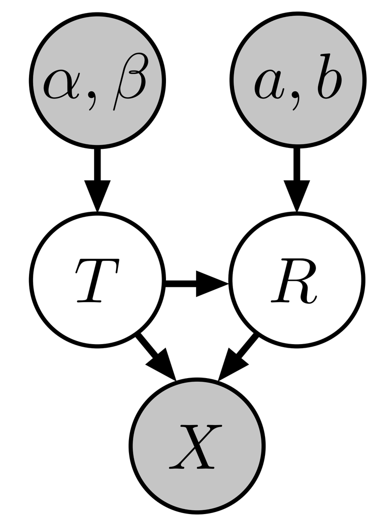
40 Journals Network [7]. Vertices are journals from four different fields, and edges indicate that an article in one journal has cited an article in the other.



A new model within the general framework

Prior over trees
The multifurcating tree is generated from a Gibbs fragmentation process [4].

$$T|\alpha, \beta \sim \text{GFT}(\alpha, \beta)$$



Parameters on each node
Each node in the tree has a single parameter R_r:

$$R_r|T, a, b \sim \text{Beta}(a, b)$$

Network links
Links are generated with probability:

$$p(X_{ij} = 1|T, R) = 1 - \prod_{r \in \text{a}(i) \cap \text{a}(j)} (1 - R_r)$$

References

1. Charles Blundell, Yee Whye Teh, and Katherine Heller. Bayesian Rose Trees. In Uncertainty in Artificial Intelligence, Proceedings of the International Conference on, 2010.
2. Aaron Clauset, Cristopher Moore, and M E J Newman. Hierarchical structure and the prediction of missing links in networks. Nature, 453(7191):98101, 2008.
3. Tue Herlau. Inference in complex networks. Masters thesis, Technical University of Denmark, 2011.
4. Peter McCullagh, Jim Pitman, and Matthias Winkel. Gibbs fragmentation trees. Bernoulli, 14(4):9881002, 2008.
5. Daniel M Roy, Charles Kemp, Vikash K Mansinghka, and Joshua B Tenenbaum. Learning annotated hierarchies from relational data. Advances in Neural Information Processing Systems 19, 19:11851192, 2007.
6. Morten Mørup, Mikkel N. Schmidt, and Lars Kai Hansen. Infinite multiple membership relational modeling for complex networks. In Machine Learning for Signal Processing, IEEE International Workshop on (MLSP), 2011.
7. Martin Rosvall and Carl T Bergstrom. An information-theoretic framework for resolving community structure in complex networks. Proceedings of the National Academy of Sciences of the United States of America, 104(18), 7327-7331. 2007.