Inferring Network Structure from Co-Occurrences

Michael G. Rabbat[1], Mário A.T. Figueiredo[2], and Robert D. Nowak[1]

[1] University of Wisconsin, Madison WI 53715, USA,
rabbat@cae.wisc.edu, nowak@ece.wisc.edu
WWW home page: http://www.cae.wisc.edu/~rabbat
http://www.ece.wisc.edu/~nowak
[2] Instituto Sperior Téchnico, 1049-001 Lisboa, Portugal
mtf@lx.it.pt
WWW home page: http://www.lx.it.pt/~mtf

The discovery of network structures is a fundamental problem arising in numerous fields of science and technology, including communication systems, biology, sociology, and neuroscience. Unfortunately, it is often difficult to obtain data that directly reveals network structure, and so one must infer a network from incomplete data. In this talk we describe and analyze a framework for inferring network structure from indirect measurements called "co-occurrences"; a type of observation that identifies which nodes occur in a transmission path but does not indicate their order within the path. Without order information every permutation of the co-occurring nodes leads to a different feasible solution, so the size of the feasible set grows exponentially in the problem dimensions. However, physical principles underlying most networked systems suggest that not all feasible solutions are equally likely. Intuitively, nodes which co-occur more frequently are probably more closely connected. Based on this intuition, we propose a probabilistic observation model: co-occurrences are modelled as samples of a random walk, subjected to a random permutation which accounts for the lack of order information. Treating the permutations as missing data, we derive an exact *expectation-maximization* (EM) algorithm for estimating the random walk parameters. Taking this view results in a significant improvement in computational complexity: compared to scaling exponentially in the network size (for exhaustive search of the feasible set), the EM algorithm scales exponentially in the length of the longest path. Since the exact EM algorithm may still be intractable for networks with very long paths, we also propose a *Monte Carlo EM* (MCEM) algorithm based on importance sampling. Using novel concentration inequalities for self-normalizing sums, we derive conditions which ensure convergence of the algorithm with high probability. Remarkably, the polynomial-complexity MCEM maintains desirably properties of the exact EM algorithm (namely, monotonic convergence to a local maximum). Details of this approach appear in [1].

References

1. Rabbat, M., Figueiredo, M., Nowak, R.: Network inference from co-occurrences. submitted to *IEEE Transactions on Information Theory* (2006)