Biological Networks: Inference, Analysis, and Applications

By: Soheil Feizi

Massachusetts Institute of Technology

The temporal and spatial expression of genes is coordinated by a hierarchy of transcription factors, whose interactions to each other and to their target genes form directed gene regulatory networks. In addition to individual interactions, the structure of a regulatory network captures a broad systems-level view of regulatory and functional processes, since genes cluster into modules that perform similar functions. Accurate inference of these regulatory networks is important both in the recovery and functional characterization of gene modules, and for comparative genomics of regulatory networks across multiple species. This is especially important because animal genomes, such as fly, worm, and mouse are routinely used as models for human disease.

Here we[1] illustrate the use of spectral, combinatorial, and statistical inference techniques in inference and analysis of gene regulatory networks and modules. First we introduce network maximal correlation (NMC) as an essential measure to capture nonlinear associations between variables in large datasets. NMC infers, possibly nonlinear, transformations of variables with zero means and unit variances by maximizing total nonlinear correlation over the underlying network. For the case of having two variables, NMC is equivalent to the standard Maximal Correlation. We characterize a solution of the NMC optimization using geometric properties of Hilbert spaces for both discrete and jointly Gaussian variables. Under some general conditions, we show that NMC can infer the underlying graphical model for functions of latent jointly Gaussian variables. These functions are unknown, bijective, and can be nonlinear. We apply NMC to different cancer datasets including breast, kidney and liver cancers, and show that NMC infers gene modules that are significantly associated with survival times of individuals while they are not detected using linear association measures

Next, we[2] discuss the inference problem of low-dimensional structures, such as clusters, on large networks. An important class of models describing such structures is the Random Dot Product Graph (RDPG), which assigns low dimensional latent position vectors to nodes and computes edge probabilities using dot products between these vectors. The RDPG provides a more flexible network model compared with the standard Stochastic Block Model (SBM). Here, we introduce the Logistic RDPG, which uses a logistic link function mapping from latent positions to edge probabilities. The logistic RDPG includes most SBMs as well as other low-dimensional structures, such as degree-

[1] This is a joint work with A. Makhdoumi, M. Médard, M. Kellis and K. Duffy.

[2] This is a joint work with L. O'Connor and M. Médard.

corrected models, that are not described by SBMs. Over this model, we derive a method for efficient, asymptotically exact maximum-likelihood inference of latent position vectors. Our method involves computing the top eigenvectors of the mean-centered adjacency matrix and performing a logistic regression step to recover the appropriate eigenvalue scaling. Applied to the network clustering problem on diverse synthetic and real network models, we illustrate that our method is more accurate and more robust than existing spectral and semidefinite network clustering methods.

Finally, we[3] consider the problem of network alignment which is important in comparative analysis across networks. Network alignment aims to find a bijective mapping between nodes of two networks to maximize their overlapping edges while minimizing mismatches. This problem arises in many fields such as computational biology and social sciences, and is often cast as an expensive quadratic assignment problem. Here we introduce EigenAlign, a network alignment algorithm inspired by eigenvector analysis which utilizes a spectral relaxation for the underlying alignment optimization. Unlike existing methods, EigenAlign considers both matched and mismatched interactions in its optimization and therefore, it is effective in aligning networks even with low similarity. We assess the effectiveness of EigenAlign theoretically for certain special classes of networks, and through simulations over various synthetic network models including Erdös-Rényi and power-law graphs. We then apply EigenAlign to compare gene regulatory networks inferred in human, fly and worm, and show that EigenAlign identifies conserved regulatory interactions, centrally-connected genes and regulatory pathways across these species despite the long evolutionary distances spanned.

[3] This is a joint work with G. Quon, M. Médard, M. Kellis, and A.Jadbabaie.