Special Topics in Computational Network Biology Fall 2019 Syllabus

Sushmita Roy, Anthony Gitter

1 Course description

Overview. A network representation can be a powerful representation for many biological and biomedical problems. Network biology is an emerging area that encompasses theory and applications of networks to study complex systems such as living organisms. This course surveys the current literature on computational, graph-theoretic approaches that use network algorithms for biological modeling, analysis, interpretation, and discovery.

The material covered in this class will come from published literature, review articles, and selected book chapters. Students will participate in discussions of papers and gain hands-on experience in network biology by implementing class projects. This class should be of interest to students from multiple disciplines including computer science, engineering, math, statistics, microbiology, biochemistry, and genetics.

Course objective. The goal of this course is to provide students an introduction to different computational problems that arise in the biological networks, key algorithms to solve these problems, and in-depth case studies showing practical applications of these concepts. The course will provide the necessary relevant background in machine learning, graph theory, and molecular biology needed to grasp the concepts introduced in the class.

Pre-requisites. Some programming experience. Students can email the instructor about questions and concerns about their background.

2 Course topics

2.1 Foundations

- Introductory graph theory
- Introductory probability theory
- Introduction to molecular networks

2.2 Readings

- Life and Its Molecules: A Brief Introduction. AI Magazine 25(1):9-22, 2004.
- "Network Biology" section in Topology of molecular interaction networks. `https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-7-90#Sec2`

3 Representation and learning of molecular networks from data

3.1 Foundations

- Probabilistic graphical model representation of molecular networks
- Bayesian networks : Score-based graph structure learning of Bayesian networks, Module networks
- Dependency networks
- Graphical Gaussian models
- Dynamic Bayesian network
- Regression trees and random forests
- Prior-based approaches for integrating diverse data types
- Regularized regression
- Markov chain Monte Carlo
- Information theoretic methods

3.2 Readings

- Inferring cellular networks – a review. `http://dx.doi.org/10.1186/1471-2105-8-s6-s5`
- **Bayesian networks**
  - Using Bayesian networks to analyze expression data. `http://dx.doi.org/10.1089/106652700750050961`
  - Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge. `http://dx.doi.org/10.2202/1544-6115.1282`
  - Bayesian Inference of Signaling Network Topology in a Cancer Cell Line. `http://bioinformatics.oxfordjournals.org/content/28/21/2804.full.pdf`
  - E. Segal, D. Pe'er, A. Regev, D. Koller, and N. Friedman, "Learning module networks," Journal of Machine Learning Research, vol. 6, pp. 557-588, Apr. 2005. [Online]. Available: `http://www.jmlr.org/papers/volume6/segal05a/segal05a.pdf`
- **Dependency networks/Undirected models**
  - Inferring regulatory networks from expression data using Tree-Based methods. `http://dx.doi.org/10.1371/journal.pone.0012776`
  - Sparse inverse covariance estimation with the graphical lasso. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3019769/`
  - Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. `http://dx.doi.org/10.1093/bioinformatics/btt099`
  - (Optional) Learning Graphical Model Structure using L1-Regularization Paths. `http://www.cs.ubc.ca/~murphyk/Papers/aaai07.pdf`
- **Single cell network inference**
  - Aibar, Sara, Carmen Bravo Gonzalez-Blas, Thomas Moerman, Van Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, et al. 2017. "SCENIC: Single-Cell Regulatory Network Inference and Clustering." Nature Methods 14 (11): 1083-86. `https://doi.org/10.1038/nmeth.4463.`

  - Chan, Thalia E., Michael P. H. Stumpf, and Ann C. Babtie. 2017. "Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures." Cell Systems 5 (3): 251-267.e3. `https://doi.org/10.1016/j.cels.2017.08.014.`
  - (Optional) Fiers, Mark W. E. J., Liesbeth Minnoye, Sara Aibar, Carmen Bravo Gonzalez-Blas, Zeynep Kalender Atak, and Stein Aerts. "Mapping Gene Regulatory Networks from Single-Cell Omics Data." Briefings in Functional Genomics 17, no. 4 (July 1, 2018): 246-54. `https://doi.org/10.1093/bfgp/elx046.`

4 Dynamics and context-specificity of networks

4.1 Foundations

- Models to represent dynamics in networks
- Dynamic Bayesian networks
- Hidden Markov Models
- Multi-task learning
- Graphical Gaussian models
- Non-stationary dynamical models

4.2 Readings

- Discovering regulatory and signalling circuits in molecular interaction networks. `http://dx.doi.org/10.1093/bioinformatics/18.suppl_1.s233`
- Reconstructing dynamic regulatory maps. `http://msb.embopress.org/content/3/1/74`
- Sharing and Specificity of Co-expression Networks across 35 Human Tissues. `http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004220`
- Learning Non-Stationary Dynamic Bayesian Networks. `http://www.jmlr.org/papers/v11/robinson10a.html`
- (Optional) KELLER: estimating time-varying interactions between genes. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2687946/`
- (Optional) Recovering time-varying networks of dependencies in social and biological studies. `http://www.pnas.org/content/106/29/11878`

5 Topological properties of graphs

5.1 Foundations

- Introductory linear algebra
- Clustering
- Modularity measures on graphs
- Spectral clustering
- Properties of graphs: degree distribution, modularity measures on graphs, network motifs.

5.2 Readings

- Assessment of network module identification across complex diseases. `https://doi.org/10.1038/s41592-019-0509-5`.
- A Double Spectral Approach to DREAM 11 Subchallenge 3. `https://www.synapse.org/#!Synapse:syn7349492/wiki/407359`
- (Optional) Evaluating Overfit and Underfit in Models of Network Community Structure. `http://arxiv.org/pdf/1802.10582`

6 Graph comparison and alignment

6.1 Foundations

- Types of graph alignment problems
- Matrix completion and factorization

6.2 Readings

- **Species network alignment**
  - Global alignment of multiple protein interaction networks with application to functional orthology detection. `http://dx.doi.org/10.1073/pnas.0806627105`
  - Fuse: multiple network alignment via data fusion. `http://dx.doi.org/10.1093/bioinformatics/btv731`
  - (Optional) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. `http://dx.doi.org/10.1073/pnas.1534710100`
  - (Optional) LocalAli: an evolutionary-based local alignment approach to identify functionally conserved modules in multiple networks. `http://dx.doi.org/10.1093/bioinformatics/btu652`
- **Factorization and graph-based data alignment/integration**
  - Hie, Brian, Bryan Bryson, and Bonnie Berger. 2019. “Efficient Integration of Heterogeneous Single-Cell Transcriptomes Using Scanorama.” Nature Biotechnology 37 (6): 685-91. `https://doi.org/10.1038/s41587-019-0113-3.`
  - Barkas, Nikolas, Viktor Petukhov, Daria Nikolaeva, Yaroslav Lozinsky, Samuel Demharter, Konstantin Khodosevich, and Peter V. Kharchenko. 2019. “Joint Analysis of Heterogeneous Single-Cell RNA-Seq Dataset Collections.” Nature Methods 16 (8): 695-98. https://doi.org/10.1038/s41592-019-0466-z.
  - Barkas, Nikolas, Viktor Petukhov, Daria Nikolaeva, Yaroslav Lozinsky, Samuel Demharter, Konstantin Khodosevich, and Peter V. Kharchenko. 2018. “Wiring Together Large Single-Cell RNA-Seq Sample Collections.” BioRxiv, November, 460246. https://doi.org/10.1101/460246.
  - (Optional) Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. “Comprehensive Integration of Single-Cell Data.” Cell 177 (7): 1888-1902.e21. `https://doi.org/10.1016/j.cell.2019.05.031.`

7 Network-based data integration and interpretation

7.1 Foundations

- Random walk on the graph
- Diffusion on graphs
- Graph kernels
- Supervised classification

7.2 Readings

- Walking the interactome for prioritization of candidate disease genes. `http://dx.doi.org/10.1016/j.ajhg.2008.02.013`
- Algorithms for detecting significantly mutated pathways in cancer. `http://dx.doi.org/10.1089/cmb.2010.0265`
- MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. `https://www.biorxiv.org/content/early/2017/02/25/111591.full.pdf`
- Similarity network fusion for aggregating data types on a genomic scale. `https://www.nature.com/articles/nmeth.2810`
- (Optional) Diffusion maps for high-dimensional single-cell analysis of differentiation data `https://academic.oup.com/bioinformatics/article/31/18/2989/241305`
- (Optional) GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs `https://academic.oup.com/bioinformatics/article/34/16/2701/4938489`
- (Optional) Compact Integration of Multi-Network Topology for Functional Analysis of Genes. `https://www.cell.com/cell-systems/abstract/S2405-4712(16)30360-X`

8 Deep learning in network biology

8.1 Foundations

- Types of graph neural networks
- Node and graph embedding
- Graph generation

8.2 Readings

- (Optional) A Comprehensive Survey on Graph Neural Networks. `https://arxiv.org/abs/1901.00596`
- (Optional) node2vec: Scalable Feature Learning for Networks. `https://doi.org/10.1145/2939672.2939754`
- (Optional) Protein Interface Prediction using Graph Convolutional Networks. `https://papers.nips.cc/paper/7231-protein-interface-prediction-using-graph-convolutional-networks`

- (Optional) MolGAN: An implicit generative model for small molecular graphs. `https://arxiv.org/abs/1805.11973`
- (Optional) Labeled Graph Generative Adversarial Networks. `https://arxiv.org/abs/1906.03220`