Department of Computing, Imperial College London

Introduction to Bioinformatics:
Biological Networks

Spring 2010

Lecturer: Nataša Pržulj
Office: 407A Huxley
E-mail: natasha@imperial.ac.uk

Lectures:

Time	Place
LEC: Thursdays 16:00 - 18:00	Huxley 331
TUT: Thursdays 11:00 – 12:00	Huxley 145

Office Hours: By appointment. The best way to reach me is by e-mail.

Marking Scheme:

   (1) About 17% two coursework (30 marks)
   (2) About 83% final exam (150 marks)

Mark range (%)	Grade	Undergraduates	MSc Postgraduates
[90-100]	A+		
[70-90)	A		PASS
[60-70)	B	PASS	
[50-60)	C		
[40-50)	D		
[30-40)	E	FAIL	FAIL
[0-30)	F		

Recommended Texts and Readings:

a) Bjorn H. Junker and Falk Schreiber, "Analysis of Biological Networks," Wiley, 2008.
b) Bornholdt and Schuster (Editors), "Handbook of Graphs and Networks: From the Genome to the Internet," Wiley, 2003.
c) Douglas B. West, "Introduction to graph theory," 2nd edition, Prentice Hall, 2001.
d) "LEDA: A Platform for Combinatorial and Geometric Computing," by Kurt Mehlhorn, Stefan Näher, Cambridge University Press, 1999.
e) A list of papers selected by Dr. Pržulj (see below).

Course Overview and Goals:

Vast amounts of biological network data have recently been generated due to advances in experimental biology. These data sets are increasingly beings studied to obtain systems-level understanding of biological structures and processes. Various mathematical and computational tools are being used and developed to analyze and model these data aiming to achieve a better description and understanding of biological processes, disease, and contribute to the time and cost effectiveness of biological experimentation.

This course will give an overview of the existing types of biological network data, point to sources of errors and biases in the data, and introduce the current methods, models and literature on graph theoretic modeling and discrete algorithmic analyses applied to these data. The course will also present an overview of the works of several major network biology labs around the world (e.g., U. Alon, M. Vidal, M. Tyers, M. Stumpf, J. Doyle, A.-L. Barabasi etc.).

Topics Outline:

The course will cover the following topics:
a) Types of biological networks: metabolic, signaling, protein-protein interaction, etc.
b) Major databases storing biological network data (e.g. MINT, DIP, HPRD, GRID, MIPS, KEGG).
c) Sources of noise and biases in various types of the biological network data (e.g., biotechnological biases and limitations, effects of sampling).
d) Computational challenges in network analysis: introduction to basic graph theoretic and computational complexity concepts such as subgraph isomorphism and NP-completeness.
e) Properties of large networks: global (e.g., degree distribution, clustering coefficient, average diameter) and local (e.g., network motifs and graphlets).
f) Network models: various types of random graphs (e.g., Erdos-Renyi, small-world, scale-free, hierarchical, geometric) and network growth models (e.g., preferential attachment).
g) Network motifs: techniques for their detection (exhaustive and heuristic network search algorithms) and biological function (e.g., feed-forward loops in transcriptional regulation networks).
h) Interplay of network topology and function (e.g., "lethality" and "centrality," "synthetic lethality" and network "redundancy," graph theoretic pathway models).
i) From models to heuristic algorithms (e.g., exploiting network model properties for "optimal" walks through a network, or detection of small network substructures).
j) Graph alignment heuristics (e.g., PathBLAST, IsoRank, GRAAL).
k) Network evolution (e.g. gene duplication and divergence in biological network growth models).
l) Clustering problems in biological networks (e.g., detection of protein complexes).
m) Software tools and libraries for network analysis (e.g., LEDA, Cytoscape, Pajek).

Some Additional Reading Materials:

Review Papers:

1. O. Ratman, C. Wiuf, and J. W. Pinney, "From evidence to inference: probing the evolution of protein interaction networks," *HFSP Journal*, October 2009.

2. E. Pieroni, S. de la Fuente van Bentem, G. Mancosu, E. Capobianco, H. Hirt, and A. de la Fuente, "Protein networking: insights into global functional organization of proteomes," *Proteomics* 8, 799-816, 2008.
3. E. de Silva and M. P. H. Stumpf, "Complex networks and simple models in biology," *Journal of the Royal Society Interface* 2: 419-430, 2005.
   N. Pržulj, "Graph theory analysis of protein-protein interactions," a chapter in "Knowledge Discovery in Proteomics," CRC Press, 2005.
4. A.-L. Barabasi and Z. N. Oltvai, "Network Biology: Understanding the Cell's Functional Organization", *Nature Reviews Genetics* 5, 2004.
5. R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Molecular Systems Biology* 3:88, 2007.
6. T. Ideker and R. Sharan, "Protein networks in disease," *Genome Research*, 18:644-652, 2008.
7. R. Sharan and T. Ideker, "Modeling cellular machinery through biological network comparison," *Nature Biotechnology* 24:4, 427-433, 2006.
8. S. I. Berger and R. Iyengar, "Network analysis in systems pharmacology," *Bioinformatics* 25:19, 2009.

Research Papers:

9. S. Maslov and K. Sneppen, "Specificity and Stability in Topology of Protein Networks", *Science* 269, 2002.
10. H. Jeong, S. P. Mason, and A.-L. Barabasi, "Lethality and centrality in protein networks", *Nature* 411, 2001.
11. S. Coulomb, M. Bauer, D. Bernard, and M.-C. Marsolier-Kergoat, "Gene essentiality and the topology of protein interaction networks", *Proc Roy Soc B* 272, 1721-1725, 2005.
12. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, "The large-scale organization of metabolic networks", *Nature* 407, 2000.
13. R. Tanaka, "Scale-Rich Metabolic Networks", *Physical Review Letters* 94, 2005.
14. C. von Mering *et al*., "Comparative assessment of large-scale data sets of protein-protein interactions", *Nature* 417, 2002.
15. P. Braun *et al*., "An experimentally derived confidence score for binary protein-protein interactions," *Nat Methods*; 6:91-97, 2009.
16. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks", *Science* 289, 2002.
17. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of *E. coli*", *Nature Genetics* 31, 2002.
18. R. Milo *et al*., "Superfamilies of Evolved and Designed Networks", *Science* 303, 2004.
19. Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone, "Comment on "Network Motifs: Simple Building Blocks of Complex Networks" and "Superfamilies of Evolved and Designed Networks"", *Science* 305, 2004.
20. R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, and U. Alon, "Response to Comment on "Network Motifs: Simple Building Blocks of Complex Networks" and "Superfamilies of Evolved and Designed Networks"", *Science* 305, 2004.
21. S. Wuchty, Z. N. Oltvai, and A.-L. Barabasi, "Evolutionary conservation of motif constituents in the yeast protein interaction network", *Nature Genetics* 35 (2), 2003.
22. N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs", *Bioinformatics* 20 (11), 2004.

23. N. Przulj, D. G. Cornail, and I. Jurisica, "Modeling Interactome: Scale-Free or Geometric?" *Bioinformatics* 20 (18), 2004.
24. N. Przulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics* 23, e177-e183, 2007.
25. D. J. Higham, M. Rasajski, and N. Przulj, "Fitting a geometric graph to a protein-protein interaction network," *Bioinformatics* 24:8, 1093-1099, 2008.
26. O. Kuchaiev, M. Rasajski, D. J. Higham, and **N. Przulj**,"Geometric De-noising of Protein-Protein Interaction Networks," *PLoS Computational Biology*, volume 5, issue 8, e1000454, August 2009.
27. T. Milenkovic and N. Przulj, "Uncovering biological network function via graphlet degree signatures," *Cancer Informatics*, 6:257-273, 2008.
28. N. Przulj, D. G. Cornail, and I. Jurisica, "Efficient estimation of graphlet frequency distributions in protein-protein interaction networks", *Bioinformatics* 22:8, 974-980, 2006.
29. M. Lappe and L. Holm. "Unraveling protein interaction networks with near-optimal efficiency." *Nature Biotechnology*, 22(1):98–103, 2004.
30. A. S. Schwartz et al., "Cost-effective strategies for completing the interactome," *Nature Methods* 6:1, 55-61, 2009.
31. C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks", *Nature* 433, 2005.
32. S. Itzkovitz *at al*., "Coarse-Graining and Self-Dissimilarity of Complex Networks", *Phys Rev E* 71, 016127, 2005.
33. M. Stumpf, C. Wiuf, and R. May, "Subnets of scale-free networks are not scale-free: Sampling properties of networks", *PNAS* 102 (12), 2005.
34. J.-D. Han, D. Dupuy, N. Bertin, M. E. Cusick, and M. Vidal, "Effect of sampling on topology predictions of protein-protein interaction networks", *Nature Biotechnology* 23 (7), 2005.
35. A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, "Modeling of protein interaction networks", *Complexus* 1, 2003.
36. N. Przulj, O. Kuchaiev, A. Stevanovic, and W. Hayes, *Geometric Evolutionary Dynamics of Protein Interaction Networks*, Proceedings of the 2010 Pacific Symposium on Biocomputing (PSB), Big Island, Hawaii, January 4-8, 2010.
37. E. F. Keller, "Revisiting 'scale-free' networks", *BioEssays* 27: 1060-8, 2005.
38. J-F Rual et al., "Towards a proteome-scale map of the human protein-protein interaction network", Nature, 2005.
39. U. Stelzl et al., "A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome", *Cell* 122: 957-68, 2005.
40. S. Suthram, T. Sittler, and T. Ideker, "The Plasmodium protein network diverges from those of other eukaryotes", *Nature* 438 (3): 108-12, 2005.
41. J-D J. Han *et al*., "Evidence for dynamically organized modularity in the yeast protein-protein interaction network," *Nature* 430, 88-93, 2004.
42. Batada, NN, *et al*. "Stratus not altocumulus: a new view of the yeast protein interaction network." *PLoS Biol.* **4**, e317, 2006.
43. Bertin, N, *et al*. "Confirmation of organized modularity in the yeast interactome." *PLoS Biol.* **5**, e153, 2007.

Research Papers on Network Alignment:

44. B. P. Kelley *et al*., "Conserved pathways within bacteria and yeast as revealed by global protein network alignment", *PNAS* 100 (20), 2003.
45. J. Berg and M. Lassig, "Local graph alignment and motif search in biological networks", *PNAS* 101 (41), 2004.

46. Kelley, B.P., Bingbing, Y., Lewitter, F., Sharan, R., Stockwell, B.R., Ideker, T.: PathBLAST: a tool for alignment of protein interaction networks. Nucl. Acids Res. **32**(Web Server issue) W83–W88, 2004.
47. Flannick, J., Novak, A., Balaji, S., Harley, H., Batzglou, S.: Graemlin general and robust alignment of multiple large interaction networks. Genome Res **16**(9), 1169–1181, 2006.
48. Liang, Z., Xu, M., Teng, M., Niu, L.: NetAlign: a web-based tool for comparison of protein interaction networks. Bioinformatics **22**(17), 2175–2177, 2006.
49. Berg, J., Lassig, M.: Cross-species analysis of biological networks by Bayesian alignment. Proceedings of the National Academy of Sciences **103**(29), 10967–10972, 2006.
50. Sharan, R.: Conserved patterns of protein interaction in multiple species. Proc. Natl. Acad. Sci. USA **102,** 1974–1979, 2005.
51. Koyuturk, M., Kim, Y., Topkara, U., Subramaniam, S., Szpankowski, W., Grama, A.: Pairwise alignment of protein interaction networks. Journal of Computational Biology **13**(2), 182–199, 2006.
52. Singh, R., Xu, J., Berger, B.: Pairwise global alignment of protein interaction networks by matching neighborhood topology. In: Research in Computational Molecular Biology. Springer, 16–31, 2007.
53. Flannick, J., Novak, A.F., Do, C.B., Srinivasan, B.S., Batzoglou, S.: Automatic parameter learning for multiple network alignment. In: RECOMB, 214–231, 2008.
54. Zaslavskiy, M., Bach, F., Vert, J.P.: Global alignment of protein-protein interaction networks by graph matching methods. Bioinformatics **25**(12), i259–i267, 2009.
55. Liao, C.S., Lu, K., Baym, M., Singh, R., Berger, B.: Isorankn: spectral methods for global alignment of multiple protein networks. Bioinformatics **25**(12), i253–258, 2009.
56. O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes, and N. Przulj, "Topological network alignment uncovers biological function and phylogeny," *Nature Precedings*, 17 December, 2009.