Comparative Analysis of Molecular Interaction Networks

Ananth Grama
Coordinated Systems Lab and,
Computer Science Department,
Purdue University

http://www.cs.purdue.edu/people/ayg

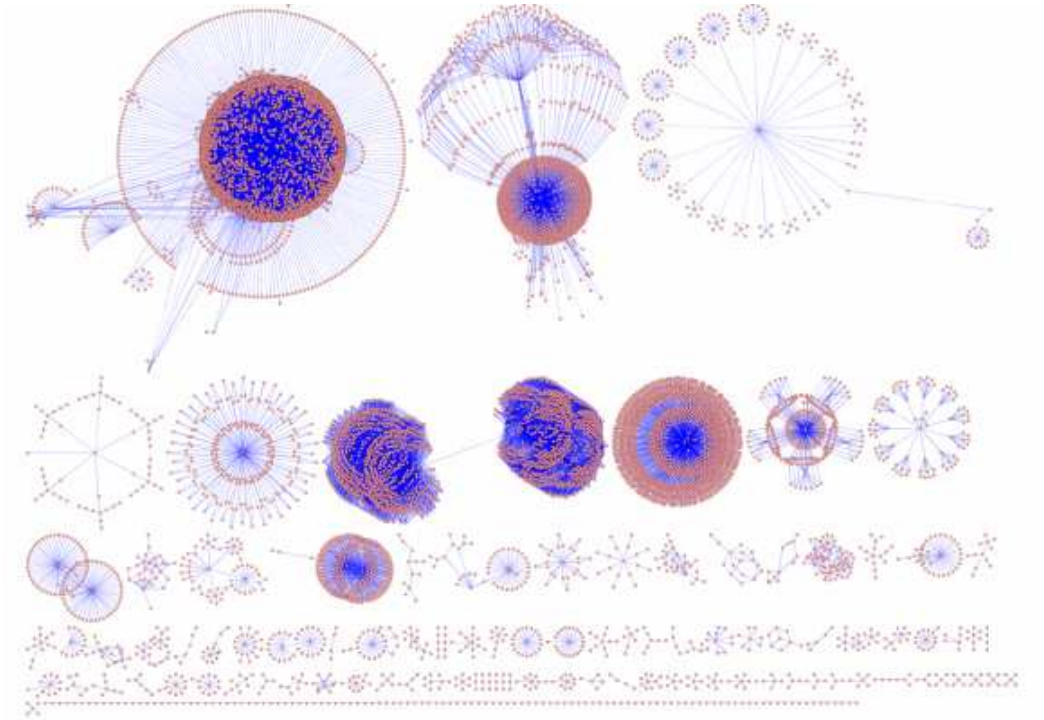
Various parts of this work involve collaborations with Shankar Subramaniam (UCSD) and Mehmet Koyuturk (Case Western)

This work is supported by National Institutes of Health and National Science Foundation.

Outline

- Molecular Interaction Networks
  - Modeling, evolution, problems, practical implications
- Algorithms for Analyzing Molecular Interaction Networks
  - Conservation, Inference, Projection, Up- and Down-scaling
- Networks, Annotations, and Phenotypic Characterization
  - Correlations across genotype, SNPs, transcriptional relationships, etc.

The AfCS Molecule Pages



The complete state network of Molecule Pages, which currently consists of 13883 states and 20556 transitions among these states (`http://www.signaling-gateway.org/molecule/`).

Evolution of Molecular Interactions

- "Evolution thinks modular" (Vespignani, *Nature Gen.*, 2003)
- Cooperative tasks require all participating units
  - Selective pressure on preserving interactions & interacting proteins
  - Interacting proteins follow similar evolutionary trajectories (Pellegrini et al., *PNAS*, 1999)
- Orthologs of interacting proteins are likely to interact (Wagner, *Mol. Bio. Evol.*, 2001)
  - Conservation of interactions may provide clues relating to conservation of function
- Modular conservation and alignment hold the key to critical structural, functional, and evolutionary concepts in systems biology

Conserved Interaction Patterns

- Given a collection of interaction networks (belonging to different species), find sub-networks that are common to an interesting subset of these networks (Koyutürk, Grama, & Szpankowski, *ISMB*, 2004)
  - A sub-network is a group of interactions that are tied to each other (connected)
  - Frequency: The number of networks that contain a sub-network, is a coarse measure of statistical significance
  - Subgraph Isomorphism!

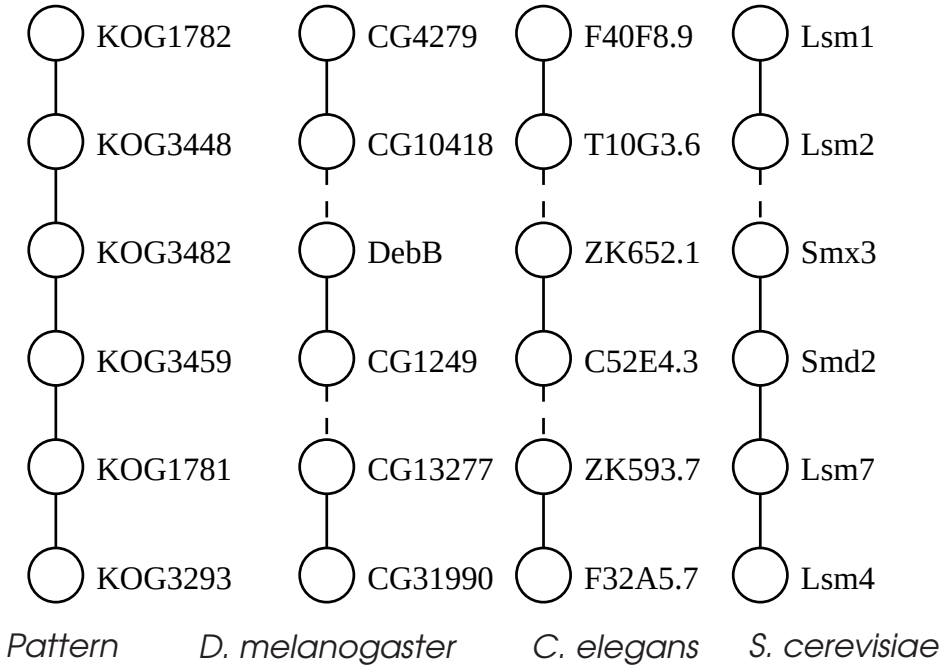
Algorithmic Insight: Ortholog Contraction

- Contract orthologous nodes into a single node
- No subgraph isomorphism
  - Graphs are uniquely identified by their edge sets
- Key observation: Frequent sub-networks are preserved $\Rightarrow$ No information loss
  - Sub-networks that are frequent in general graphs are also frequent in their ortholog-contracted representation
  - Ortholog contraction is a powerful pruning heuristic
- Discovered frequent sub-networks are still biologically interpretable!
  - Interaction between proteins becomes interaction between ortholog groups

Results: Analyzing PPI Networks

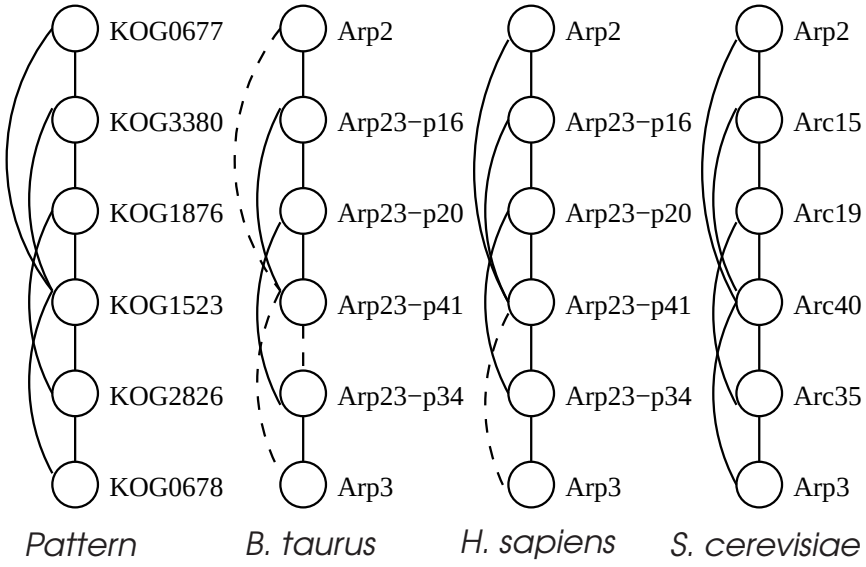
- PPI networks for 9 eukaryotic organisms derived from BIND and DIP
  - *A. thaliania, O. sativa, S. cerevisiae, C. elegans, D. melanogaster, H. sapiens, B. taurus, M. musculus, R. norvegicus*
  - # of proteins ranges from 288 (*Arabidopsis*) to 8577 (*fruit fly*)
  - # of interactions ranges from 340 (*rice*) to 28829 (*fruit fly*)
- Ortholog contraction
  - Group proteins according to existing COG ortholog clusters
  - Merge Homologene groups into COG clusters
  - Cluster remaining proteins via BLASTCLUST
  - Ortholog-contracted *fruit fly* network contains 11088 interactions between 2849 ortholog groups
- MULE is available at `http://www.cs.purdue.edu/pdsl/`

Conserved Protein Interaction Patterns



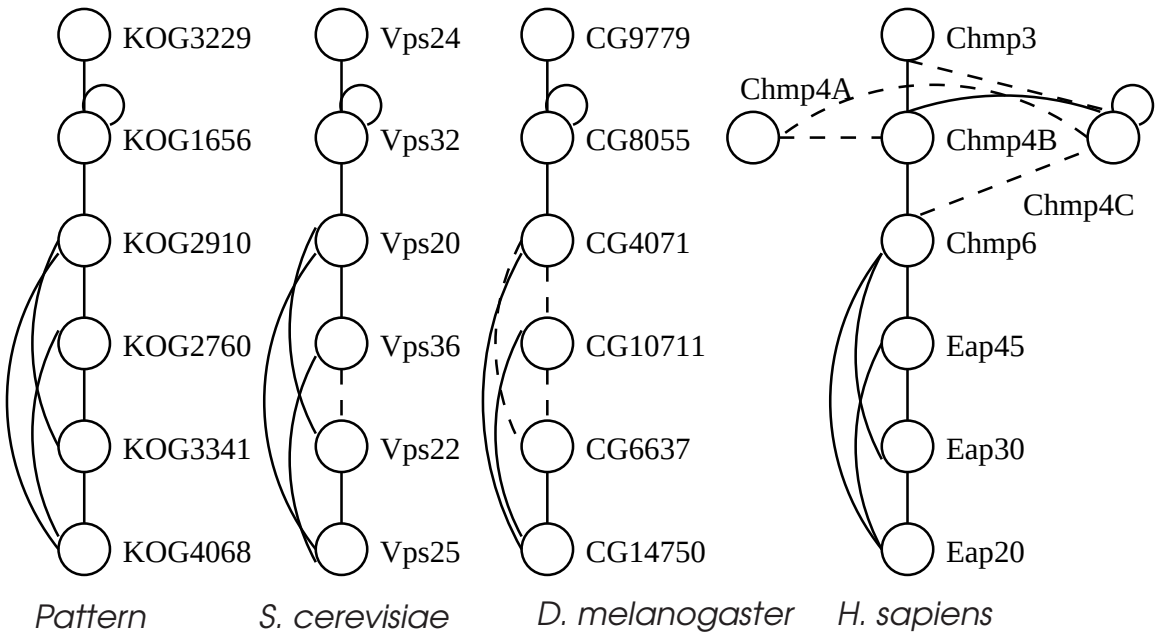
Small nuclear ribonucleoprotein complex ($p < 2e - 43$)

Conserved Protein Interaction Patterns



Actin-related protein Arp2/3 complex ($p < 9e - 11$)

Conserved Protein Interaction Patterns



Endosomal sorting ($p < 1e - 78$)

State of the Art

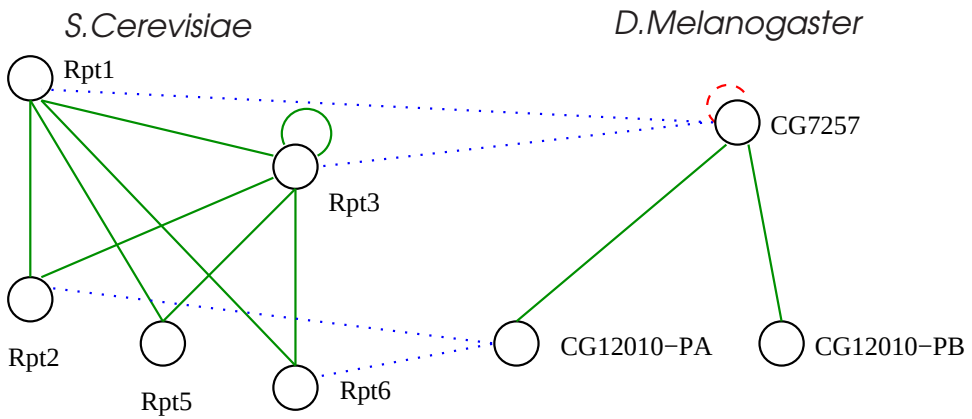
- Data is the bottleneck (in-silico construction of the network?).
- Reliable data on no more than ten organisms. Networks with thousands of nodes and tens of thousands of interactions.
- The exponential complexity of these algorithms makes them ideal candidates for large-scale platforms.
- Current implementations scale to thousands of processors and take days of runtime.

Alignment of Networks

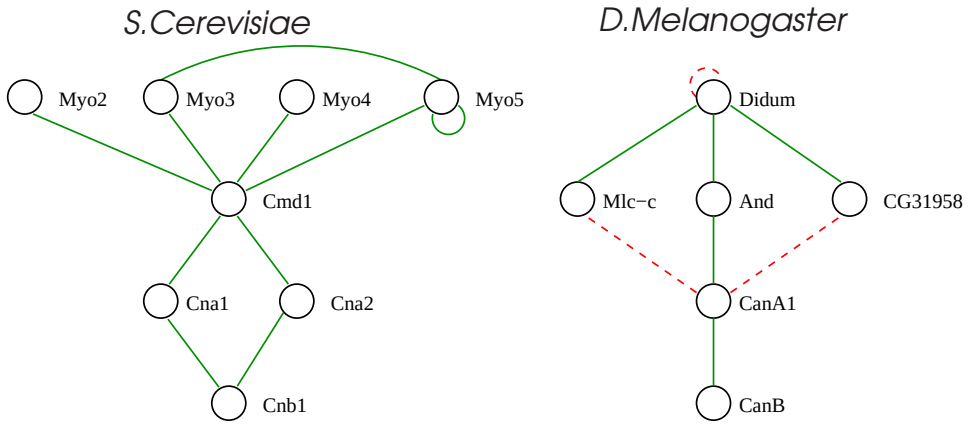
- Given two PPI networks that belong to two different organisms, identify sub-networks that are similar to each other
  - Biological implications
  - Mathematical modeling
- Existing algorithms
  - PathBLAST aligns pathways (linear chains) to simplify the problem while maintaining biological meaning (Kelley et al., *PNAS*, 2004)
  - NetworkBLAST compares conserved complex model with null model to identify significantly conserved subnets (Sharan et al., *J. Comp. Biol.*, 2005)
- Our approach (Koyutürk et al., *RECOMB*, 2005) (Koyutürk et al., *J. Comp. Biol.*, 2006)
  - Guided by models of evolution
  - Scores evolutionary events
  - Identifies sets of proteins that induce high-scoring sub-network pairs

Subnets Conserved in Yeast and Fruit Fly Using our Technique

Proteosome regulatory particle subnet



Calcium-dependent stress-activated signaling pathway



State of the Art

- Local methods are suited to identification of modularity.
- Global methods (IsoRank, Berger et al.) support functional orthology.
- Can we improve the search efficiency and coverage of greedy (local) methods?
- Speed up the eigenvalue computations of spectral methods.
- Current algorithms scale to thousands of processors.
- Data quality is a key concern.

Continuing Developments

- Building comprehensive maps that model crosstalk and weak signals (inference, modularity).
- Constructing accurate flux models (data?)
- Full scale flux analysis for molecular processes.

Continuing Developments

- Scaling molecular networks in space and time.
- Spatial scales must integrate atomistic models, scaling to cellular scale, tissue scale, individual, populations, ecosystems.
- Temporal scales must reconstruct evolution and project trajectories.

Functional and Phenotypic Characterization

- Integrate large-scale SNP data, genomic data into functional characterization of networks.
- Comprehensively characterize causality in disease specific networks.
- Supporting full-scale in-silico design of molecules.

References

- J. Pandey, M. Koyuturk, S. Subramaniam, and A. Grama. Functional coherence in domain interaction networks, Bioinformatics Suppl. on ECCB'08.
- M. Koyuturk, W. Szpankowski, and A. Grama, Assessing significance of connectivity and conservation in protein interaction networks, Journal of Computational Biology, 14(6), 747-764, 2007.
- M. Koyuturk, Y. Kim, S. Subramaniam, W. Szpankowski, and A. Grama, "Detecting conserved interaction patterns in biological networks", *Journal of Computational Biology*, 13(7), 1299-1322, 2006.
- M. Koyuturk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama, Pairwise Alignment of Protein Interaction Networks, *Journal of Computational Biology*, 13(2), 182-199, 2006.
- Y. Kim, M. Koyuturk, U. Topkara, A. Grama, and S. Subramaniam, Inferring Functional Information from Domain Co-evolution, *Bioinformatics*, 22(1), pp. 40-49, 2006.

References

- M. Koyutürk, A. Grama, and W. Szpankowski, An Efficient Algorithm for Detecting Frequent Subgraphs in Biological Networks, *Bioinformatics*, Vol. 20, Suppl. 1, pp i200-i207, 2004.
- M. Koyuturk, A. Grama, and W. Szpankowski, Assessing Significance of Connectivity and Conservation in Protein Interaction Networks, *10th International Conference on Research in Computational Molecular Biology (RECOMB)*, LNBI 3909, pp. 45-59, 2006.
- M. Koyutürk, A. Grama and W. Szpankowski, Pairwise Local Alignment of Protein Interaction Networks Guided by Models of Evolution, RECOMB 2005.

Available from `http://www.cs.purdue.edu/pdsl`