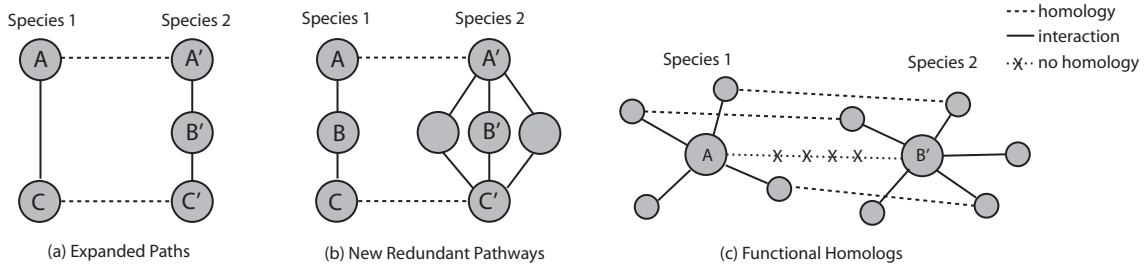


Project Ideas

1. Compare variants of majority rule for function prediction. (E.g. does using inferred labels in subsequent steps help?)
- ~~2. Consider randomized variants of the algorithm of (Newman, PNAS 2006), or apply his algorithm to a protein-protein interaction network. Can you put his algorithm on firm theoretical (CS-style) footing?~~
3. Use the R_{ij} values computed by ISORANK (Singh, Xu, Berger, 2006) for *local* alignment.
- ~~4. Search for the following cross-species motifs between protein-protein interaction networks of two species:~~



What are distinguishing properties of these motifs?

5. Are there highly connected nodes in one species that have low-degree homologs in another species? Are there instances where the gain or loss of a protein domain leads to different topologies?
6. Use METRIC LABELING to improve function annotation.

Metric Labeling. Given a weighted graph $G = (V, E)$ with edge weights $w(\cdot, \cdot)$, a set of labels L , a distance metric $d(\cdot, \cdot)$ between any two labels in L , and an assignment cost $c(u, i)$ between nodes $u \in V$ and labels $i \in L$, find a labeling $f : V \rightarrow L$ to minimize the sum of the assignment costs $\sum_u c(u, f(u))$ and the weighted sum of the distances between labels $\sum_{\{u,v\} \in E} w(u, v) d(f(u), f(v))$.

Metric labeling generalizes multiway cut to take into account distances between *the labels*. This property may be very useful when our labels are biological functions that are related in complicated ways. See (Chekuri et al, SODA 2000) and related papers. See also the 0-extension problem.

7. Explore biases in the methods for generating random graphs of a given degree distribution. How well do several variants perform in practice? (See Milo *et al.* arXiv:cond-mat/0312028v2, 2004.)

8. Explore heuristics / algorithms for generating random graphs that preserve some *other* property or properties: # of hub-hub connections, # of bottlenecks, clustering coefficient, diameter, pairwise distance distribution.
9. ~~What types of graph generation processes that are compatible with known evolutionary processes would yield the kinds of networks we observe today? Consider gene duplication, horizontal gene transfer, gene loss, etc. There is a large amount of literature considering this question, with models such as “preferential attachment.” See for example: Kim et al, *Science* **314**:1938 (2006), Middendorf, Ziv and Wiggins, *PNAS* **102**:3192 (2005), Light, Kraulis, Elofsson, *BMC Genomics* **10**:159 (2005), Pagel, Meade, Scott, *BMC Evol. Biol.* **7** Suppl 1: S16 (2007), Eisenberg and Levanon, *Phys. Rev. Lett.* **91**, 138701 (2003), Wagner, *Proc Biol Sci* **270**:457 (2003), Dhan et al, *Bioinformatics* **18**:1486 (2002).~~
10. How does the performance of function-prediction or module-detection algorithms change as properties of the graph change?
11. ~~Find nucleotide positions with correlated mutations in separate segments in the influenza genome. Identify correlated pairs that don't appear to be correlated because of common ancestry within the phylogenetic tree relating the strains.~~
12. ~~Consider a graph problem in another area of computational biology:~~
 - ~~(a) Alternative gene splicing~~
 - ~~(b) Interpreting Mass Spec data~~
 - ~~(c) Genome rearrangements~~
 - ~~(d) Protein-structure prediction~~
 - ~~(e) Epidemiology (see, e.g., Bansal, Grenfell, Ancestral Meyers, *J. R. Soc. Interface*, **4**, 879–891 (2007)).~~

~~(As previously mentioned, you're welcome to choose any project related to graphs and networks in computational biology — I would recommend, however, sticking with protein-protein interaction graphs or transcriptional networks, unless you have a strong interest otherwise.)~~

~~Project Proposal Presentation~~

- ~~● 10-15 minutes on Oct. 16, 2007.~~
- ~~● Your chance to get feedback / ideas before you get too far along in the project.~~
- ~~● Discuss preliminary work (if any) and related literature~~
- ~~● Defend your reasons for choosing this project.~~
- ~~● Explain why you think this project can be completed in the timeframe available.~~
- ~~● What will you do if the project doesn't work out?~~