Estimating Mutation Distance between Biological Networks

Martin McGrane (A/Prof Michael Charleston)
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

Biological Networks

•Organisms are defined by the biological networks that make them up, not just DNA sequence. [1]
•Proteins are key players in the biochemistry of an organism

Protein interaction networks

•Protein Interaction Networks are the networks formed by tracing which protein interacts with which protein
•The **network of interactions** defines a species, not the DNA sequence alone.

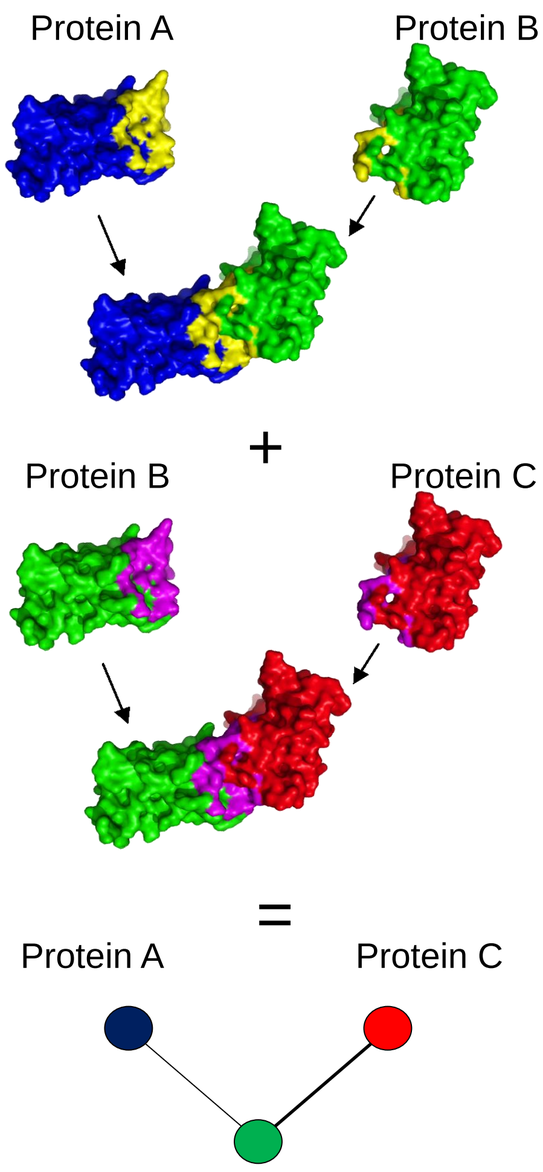
Biological Distance

•Genetic mutations are well known, but are only a small part of a complex system.
•There are many more possible changes that affect a gene's output – changes to siRNA, miRNA, histone acetylation, glycosolation, and more. [2]
•Measuring the distance between the **networks** can tell us more than simply measuring DNA similarity. [3]

Aims

Find the shortest path of mutations between biological networks, quickly and accurately

Protein Interactions as a Graph



Protein Interaction Network

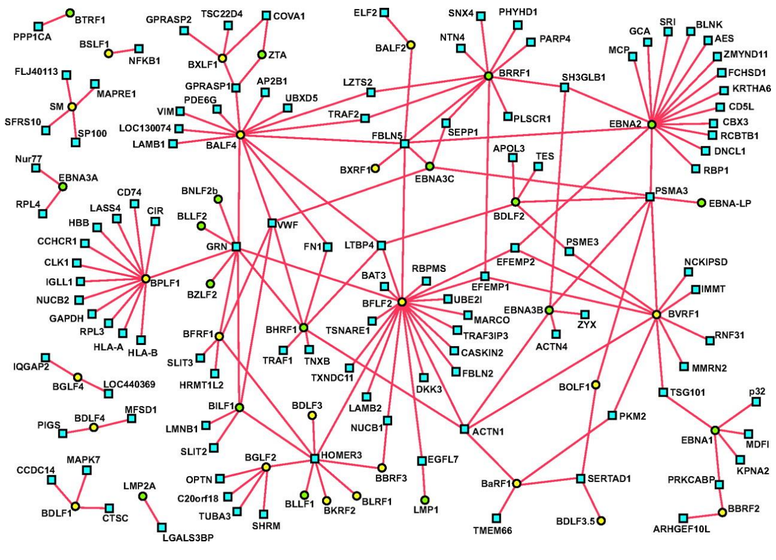


Figure 2: Epstein Barr – human interactome [4]

Methods

Graph building

•Filter protein databases to find reliable interactions

Path building

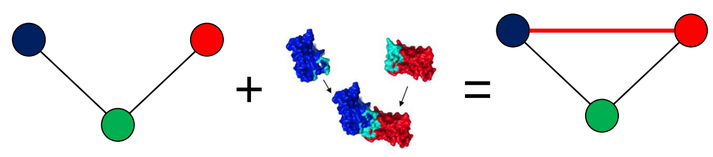
•Step by step build up a path of possible mutations
•Discard bad mutations, keep good mutations

Parallelise

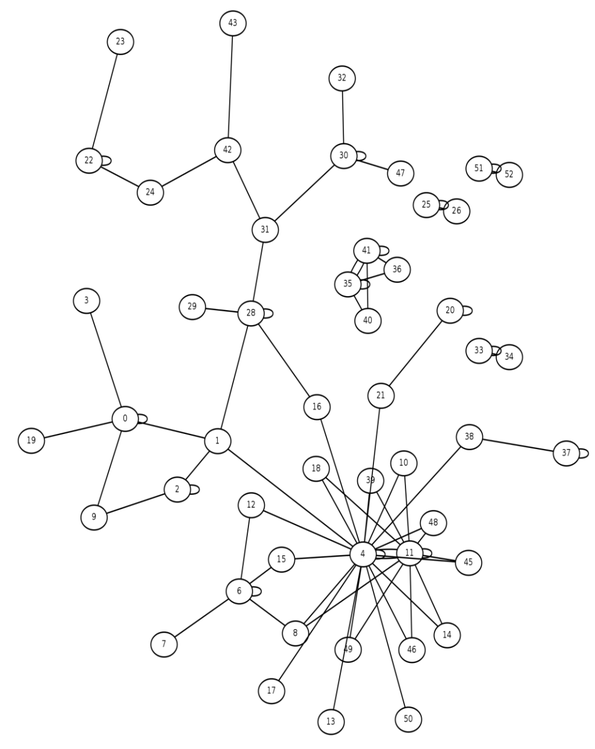
•Build hundreds of possible paths at once
•Keep the best of each path tried

Genetic Algorithm

•Merge the best paths found to produce a new set of base paths to start from
•The most genetically fit mutations survive and help to find the path we need



Mutate Protein A to interact with Protein C



+

1.Duplicate node 7 as 50 linked to (4, 5, 10)
2.Duplicate node 44 as 51 linked to (28)
3.New edge from 1 to 15.
4.New node 52 linked to 23.
5.Delete edge from 31 to 32 and node 32
6.Delete edge from 1 to 4
7.Duplicate node 8 as 53 linked to (0, 2)
8.New node 54 linked to 24
9.New edge from 18 to 54
10.New edge from 41 to 43

=

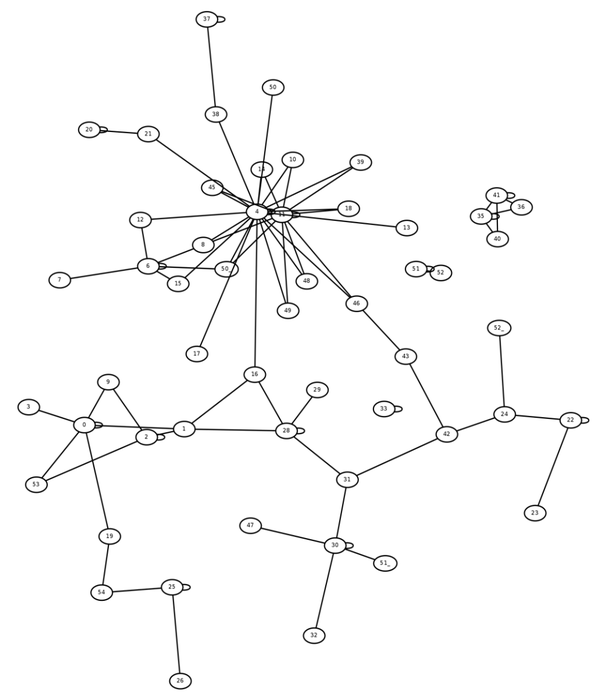


Figure 3: Finding the shortest mutation distance between two networks.

Results

- We find the shortest paths of mutations between two networks
- We rapidly converge on a path
- We find paths equivalent to known shortest path, or shorter

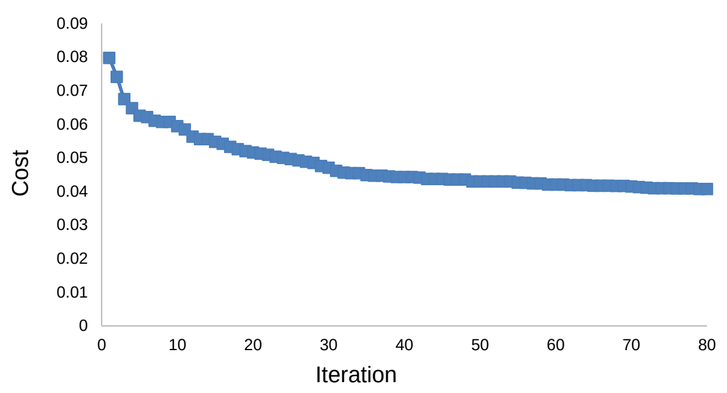


Figure 4: Similarity of mutated original and destination graph

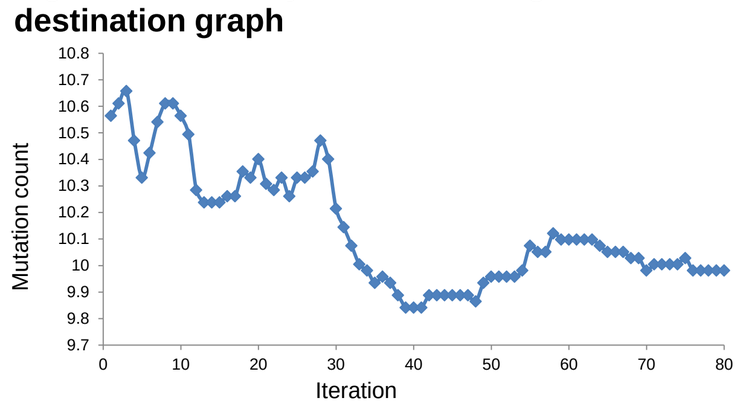


Figure 5: Distance between mutated original and destination graph

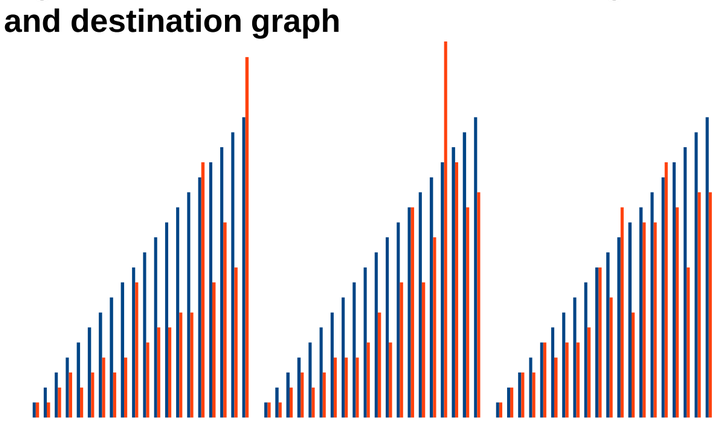


Figure 6: Finding the shortest path in 3 simulated networks of 100 vertices with 1 – 20 mutations added

Conclusions

•This is the **first model of distance between biological networks that considers the effects of mutation on the biological network**, the organism as a whole, rather than simple sequence.
•We can estimate the shortest paths between graphs faster than exhaustive searches
•**We can easily scale** to efficiently use more hardware if needed
•**We can easily expand the framework to handle all existing models of mutations**, and more.

Further Work

•Different existing models of mutation can be directly compared
•The framework can now be applied to more complex models of mutation
•Investigate the relevance of mutation models of network growth to social networks

References

1.Arabidopsis Interactome Mapping Consortium, 2011. Evidence for network evolution in an Arabidopsis interactome map. *Science (New York, N.Y.)*, 333(6042), pp.601–7.
2.Huang, S. (2004). Back to the biology in systems biology: What can we learn from biomolecular networks? Briefings in Functional Genomics and Proteomics, 2(4), 279–297. doi:10.1093/bfgp/2.4.279.
3.Luo, J. et al., 2013. Model the evolution of protein interaction network assisted with protein age. *Journal of theoretical biology*
4.Calderwood et al., 2007. Epstein-Barr virus and virus human protein interaction maps. *PNAS*, 104(18), pp.7606–11.