

# GraphLinker

A Visual Comparative Environment of Genomic and Metabolic Networks

Project Update

Niels Hanson

Monday, November 21 2011

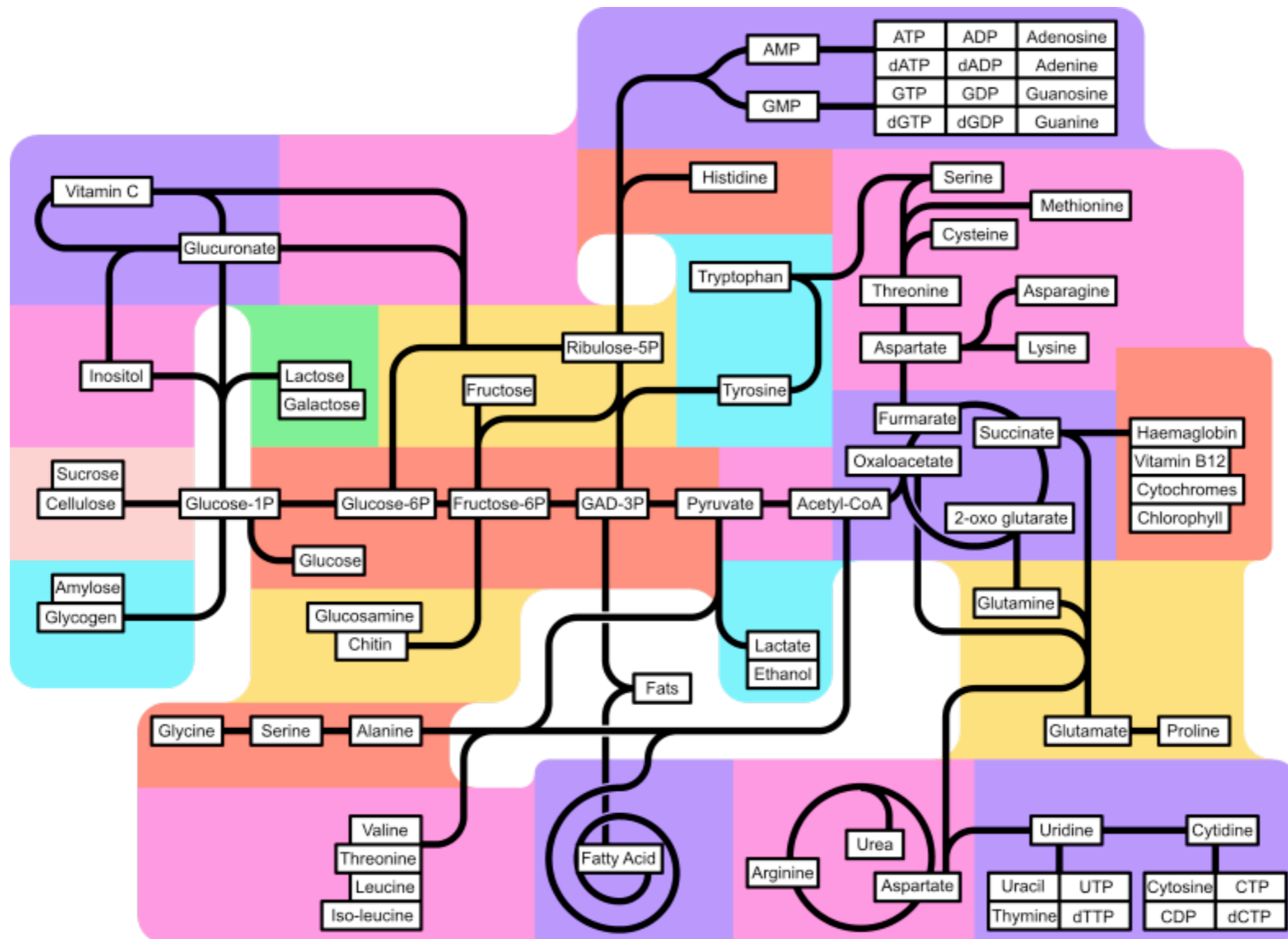
# Microbacterial Communities

- diversity and dynamics largely unexplored
- affect virtually every environment on the planet  
(your gut, fisheries, natural gas production)
- notion of bacterial machines  
— make materials, food products, fuels

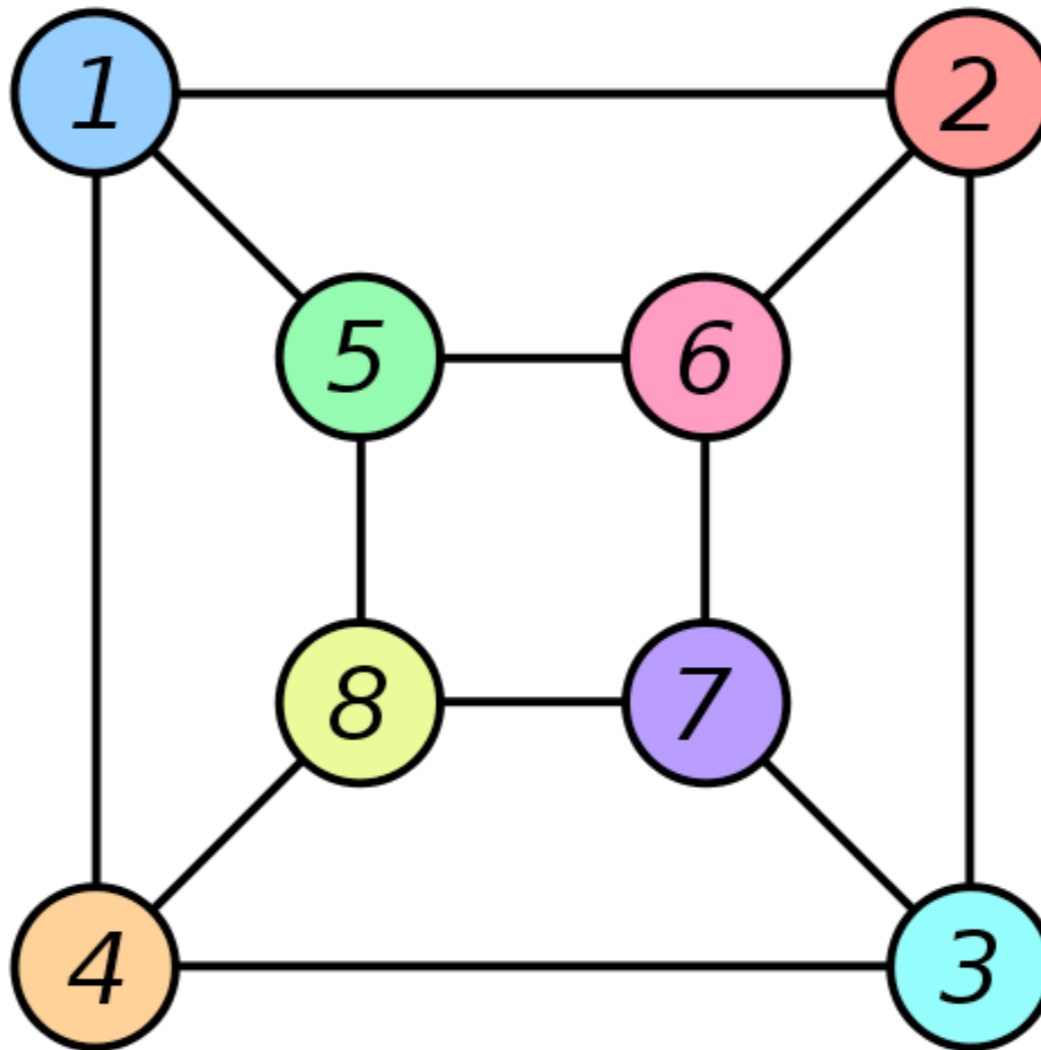
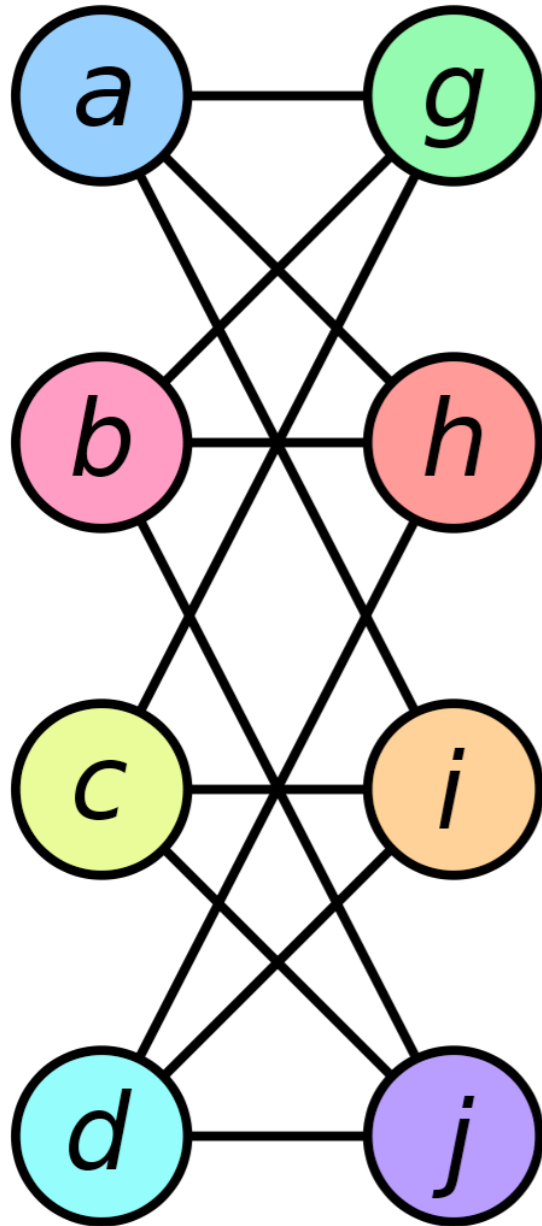
# Two Networks at Play

- Metabolic pathways and taxa (species)
- Unlike higher-order organisms, metabolism is shared between species
- Like to infer which taxa are participating in each network
- Problem can be mapped one of the classic Garry & Johnson problems, Graph Isomorphism

# Metabolic Pathways



# Graph Isomorphism



$a = 1$   
 $b = 6$   
 $c = 8$   
 $d = 3$   
 $g = 5$   
 $h = 2$   
 $i = 4$   
 $j = 7$

# Graph Isomorphism

- is NP-Complete, though approximation algorithms exist for practical use
- these algorithms give a measure of associativity on repeated runs
- a visual encoding of graph mapping may allow the use to compare two related graphs

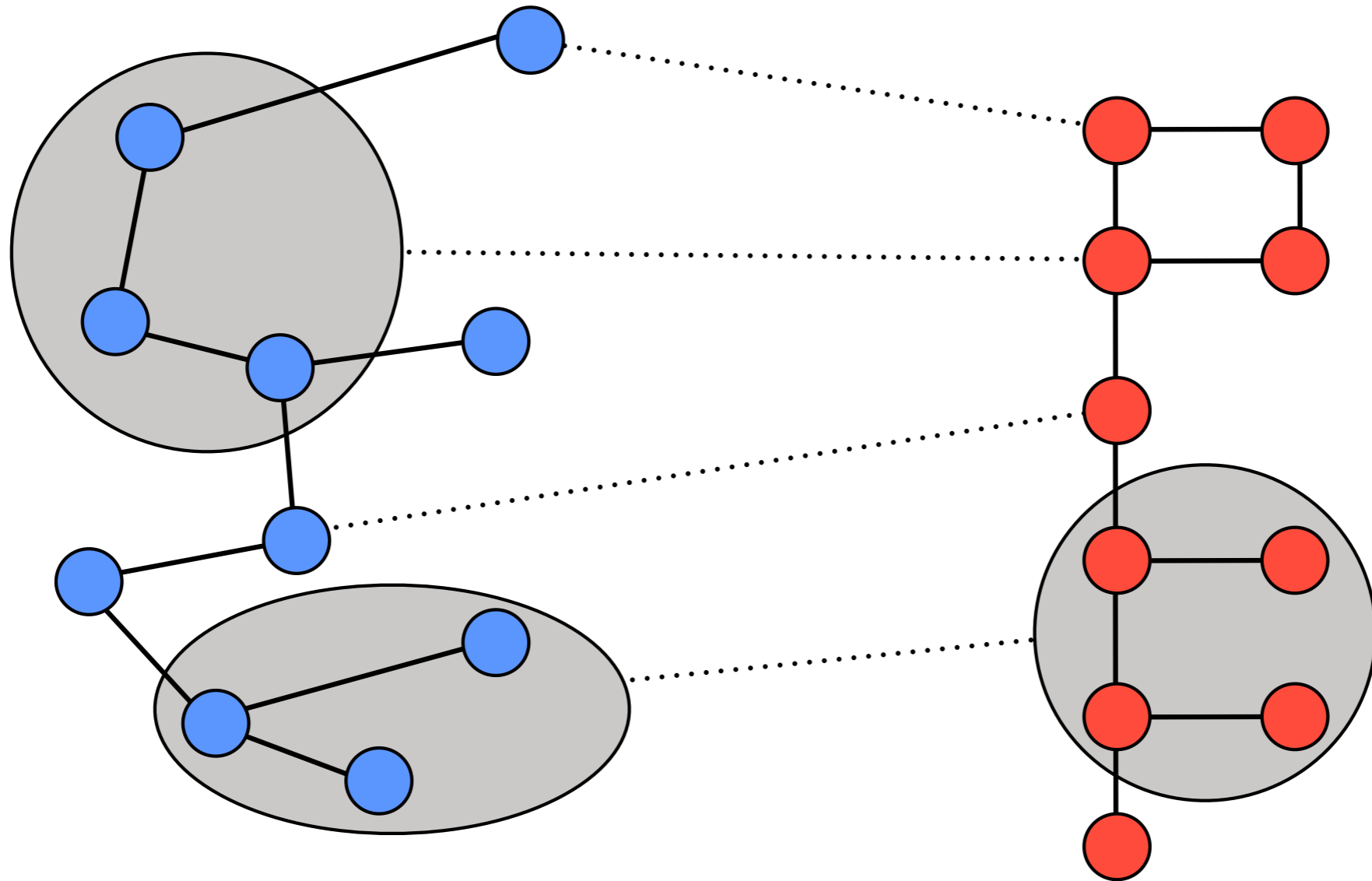
# Data

- Taxa Presence/Absence  
— Fisher's Exact test P-value
- Pathway Presence/Absence  
— Fisher's Exact of SRI<sup>1</sup> PathoLogic
- Linkage values via bootstrap of algorithm by Schmidt *et. al.*<sup>2</sup>  
— values normalized between 0 and 1

<sup>1</sup>Stanford Research Institute (SRI)

<sup>2</sup>Schmidt, D.C. & Druffel, L.E. A Fast Backtracking Algorithm to Test Directed Graphs for Isomorphism Using Distance Matrices. J.ACM 23, 433-445 (1976).

# Visual Encoding



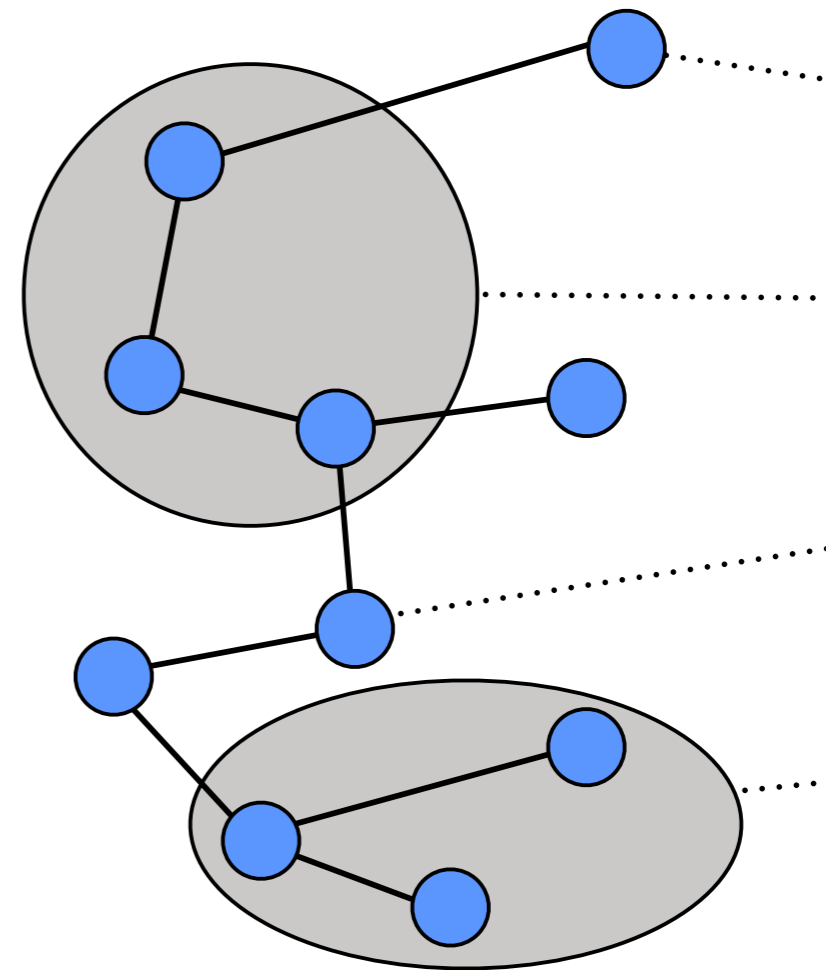
**OTUs (Taxa)**

**Metabolic Pathway**

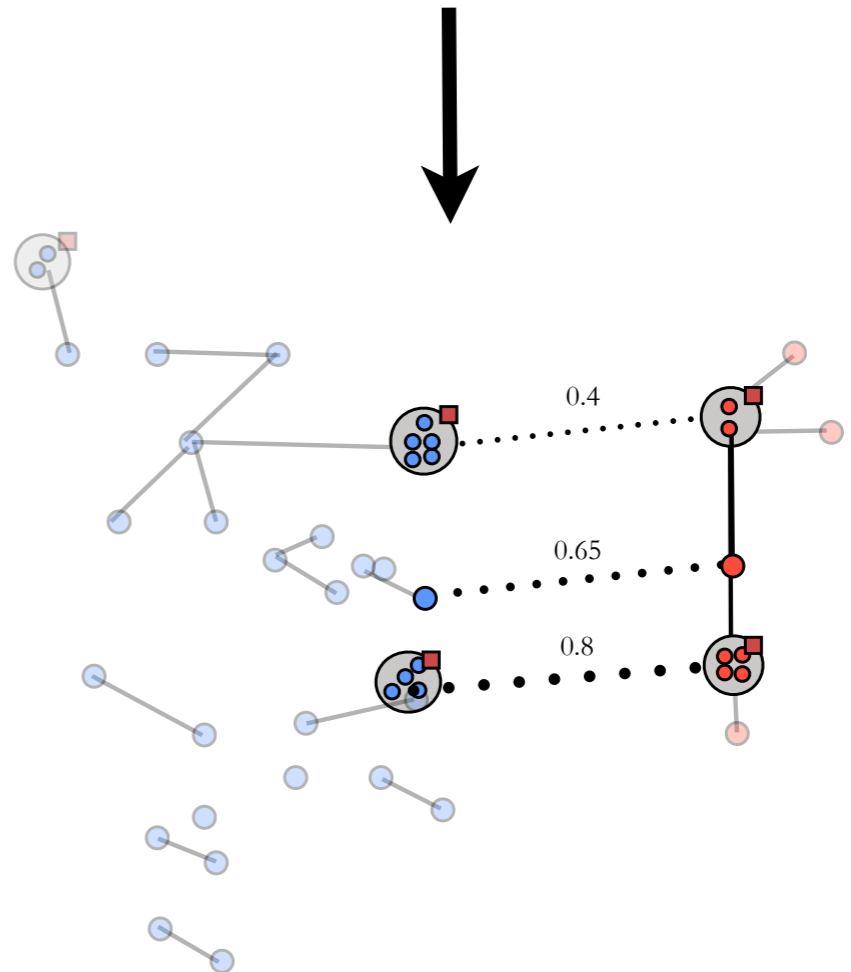
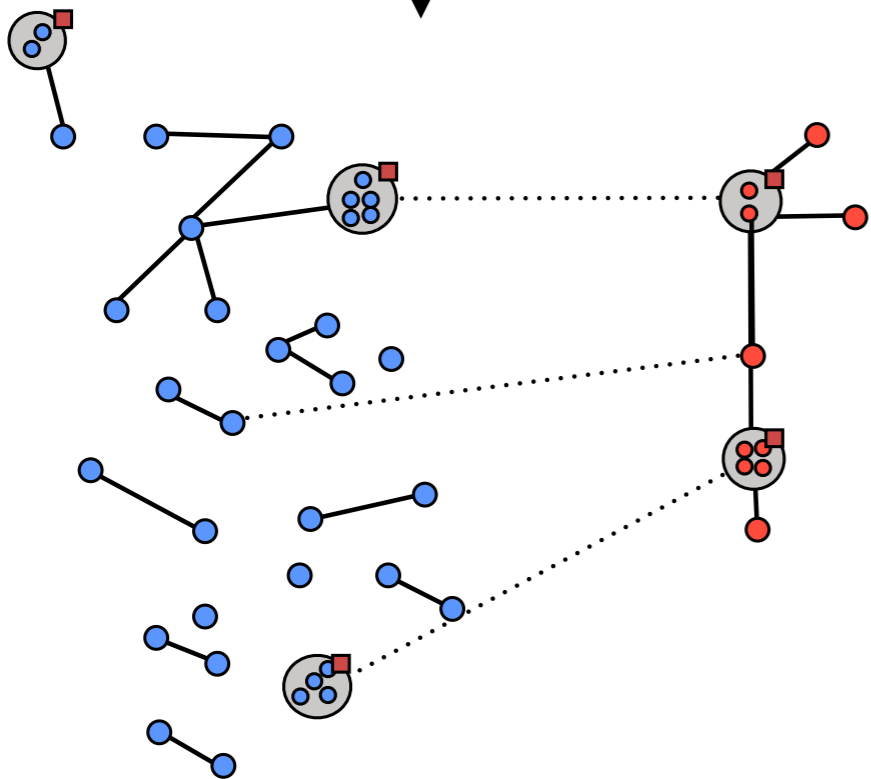
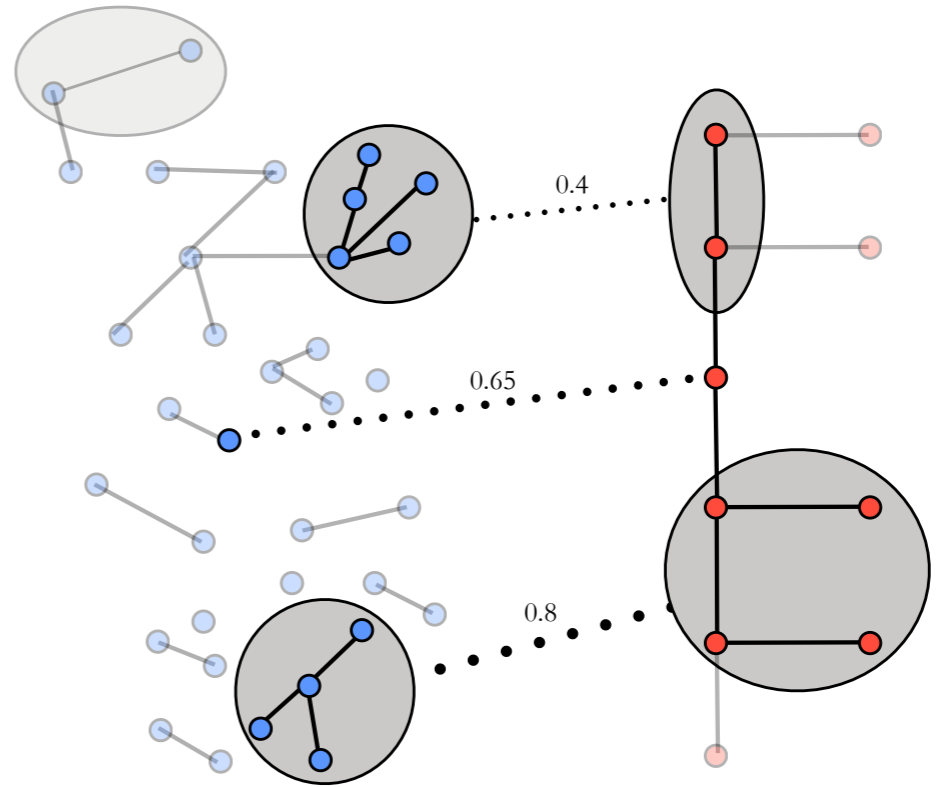
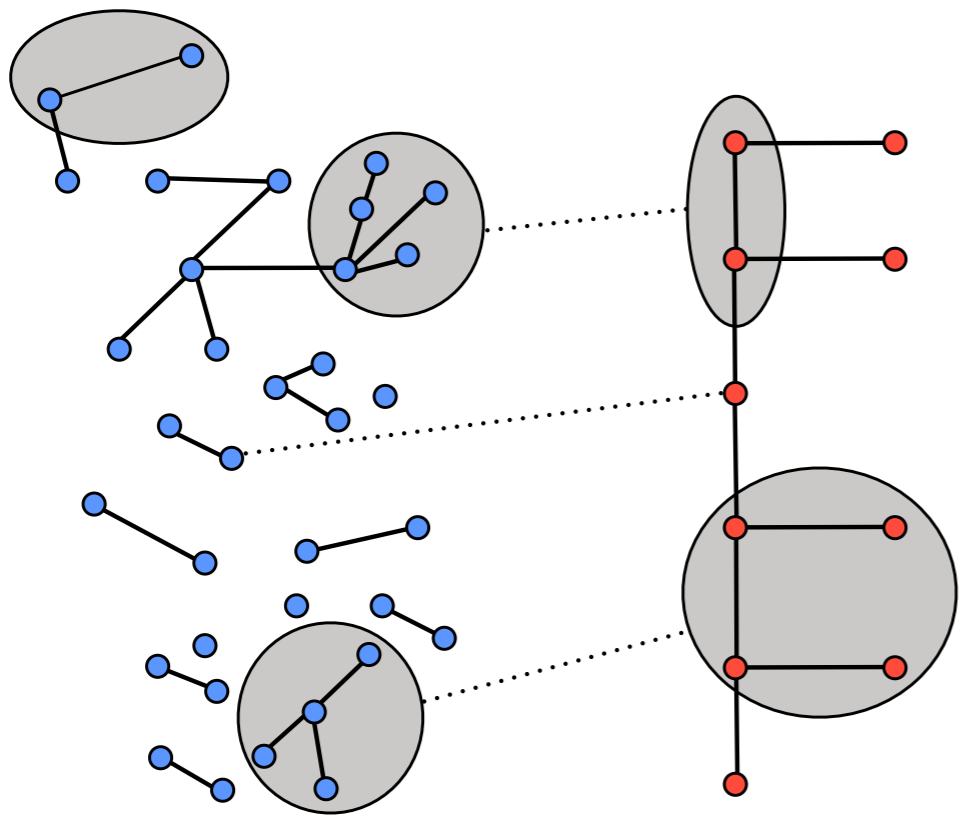


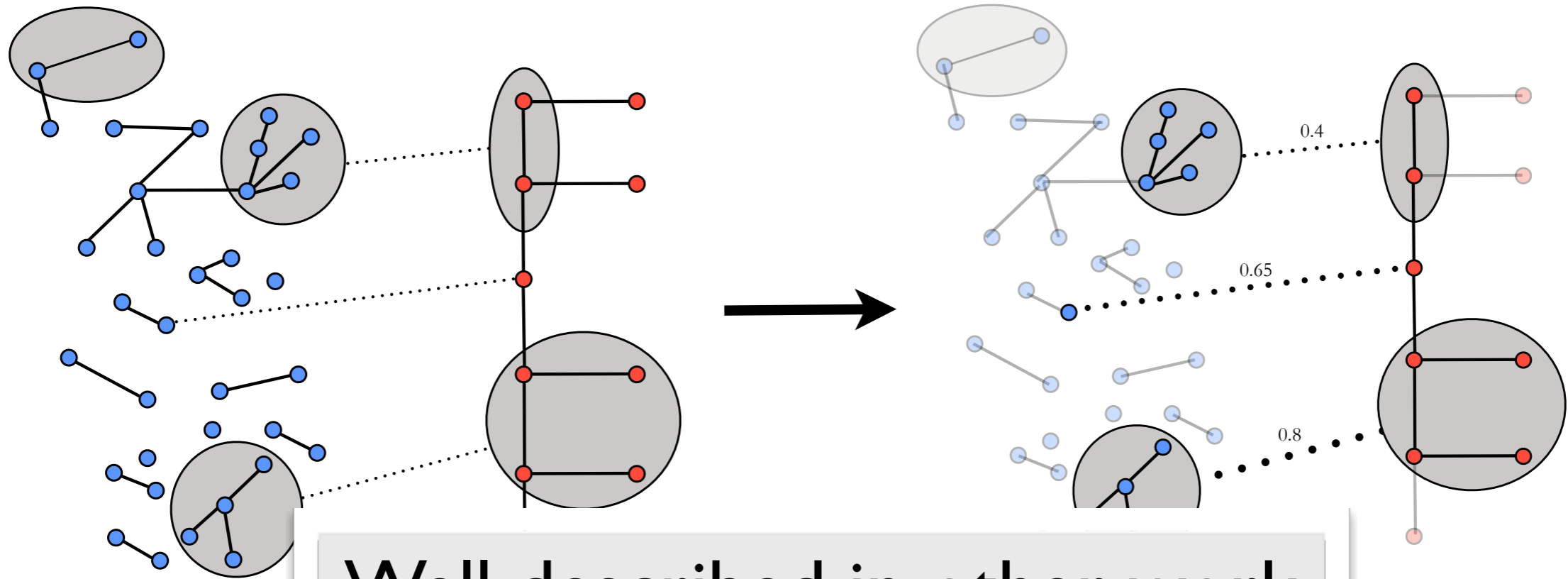
# Visual Encoding

- position to separate two graphs
- edges within and meta-edges between graphs
- clusters form meta-nodes (containment)
- meta-edge transparency encodes strength of relationship (0 to 1)

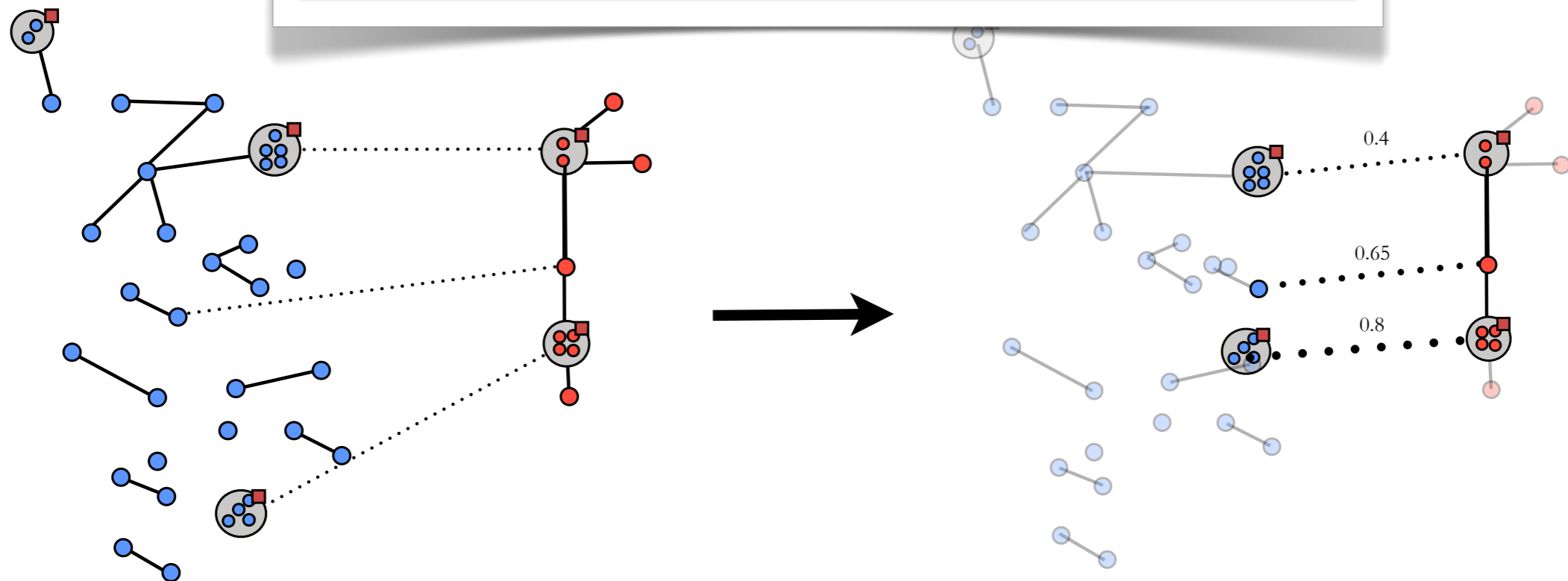


OTUs (Taxa)





Well described in other work  
Archambault et. al.





# Implementation

- Cytoscape - platform for complex network analysis [www.cytoscape.org/](http://www.cytoscape.org/)
- implement as a network analysis plugin
- software under active development
- number of different things (layouts, visual, encoding)
- can be difficult to work with framework

# Implementation



- Guarantees obsolescence
- Complete structure rework
- Should I move to Processing?

# Timeline

## Milestone

	Date
1. Become accustomed with the most recent version of Cytoscape, implementation of previous plug-ins etc.	October 29th
2. Adding edge weights to edges, displaying two separate graphs at the same time.	November 1st
3. Cluster nodes by containment.	November 4th
4. Linking nodes on two graphs by edges of a different style. Change edge width due to weight of association.	November 6th
5. Bring meta edges and associated nodes to the centre, sort vertically by edge weight.	November 8th
6. 'Ghost' or remove nodes and edges not associated with meta-edges.	November 10th
7. Have basic demonstration of functionality ready to go for project update presentation.	November 12th
8. If four of the last 7 tasks are not complete, change to 'Plan B' of implementing basic functionality in Processing.	November 15th
9. Basic Pan & Zoom should work. Keep highlighted nodes in the main view at a tangential angle.	November 18th
10. Have all meta-edge functionality working. Move on to extra features of multi-nested hierarchical clustering.	November 20th
11. Collapsing and expanding clusters or meta-nodes.	November 22nd
12. Implement occluding avoidance algorithm when expanding or collapsing nodes	November 25th
13. Final extras and features isolated. Have second demonstration model ready to present.	November 30th
14. Begin final writeup. Start replacing theoretical data with legitimate calls from biological sources.	December 1st

**Questions?**