PaIntDB: Visualizing Protein-Protein Interaction Networks in *P. aeruginosa*

CPSC 547 - Final Project

Javier Castillo-Arnemann

December 10, 2019

Hancock Lab

UBC

Background

- *Pseudomonas aeruginosa* is a multi-drug resistant pathogen involved in cystic fibrosis and other diseases.
    - Antibiotic resistance has gotten worse and will continue to do so.
- Antibiotic resistance results from the complex interactions between thousands of genes and gene products.
- A systems-level understanding of biological function is necessary to see the broader picture (looking at groups of genes instead of individual genes).
- PaIntDB (**P**seudomonas **a**eruginosa **Int**eractions **D**ata**B**ase) allows researchers to upload a long list of genes and explore their interactions.

Exploring Large Experimental Datasets

1. Upload list of genes (usually >1000) with optional experimental data.

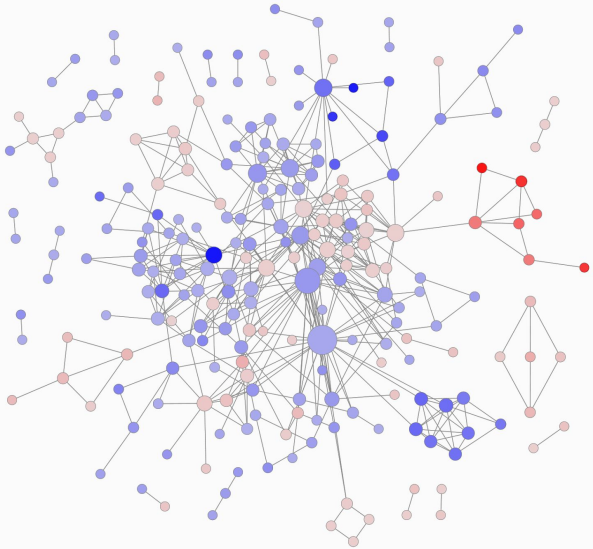
gene	baseMean	log2FoldChange	lfcSE	stat
PA3326	2141.749	-3.446	0.163	-21.173
PA3327	3569.542	-4.395	0.217	-20.232
PA3330	749.070	-4.722	0.274	-17.214
PA3333	755.576	-4.440	0.260	-17.062
PA3332	424.146	-4.497	0.264	-17.010
PA3329	859.657	-4.461	0.268	-16.648
PA3334	299.896	-4.409	0.279	-15.821
PA3331	1161.924	-4.650	0.301	-15.437
PA3328	713.058	-4.208	0.276	-15.259
PA3335	513.662	-3.845	0.253	-15.167
PA3724	776.790	-2.963	0.204	-14.534
PA3336	442.495	-4.008	0.276	-14.519
PA2840	1650.841	2.440	0.170	14.393
PA2796	696.220	-1.838	0.141	-13.049

2. Map genes to interaction database and generate a network.

Additional info:

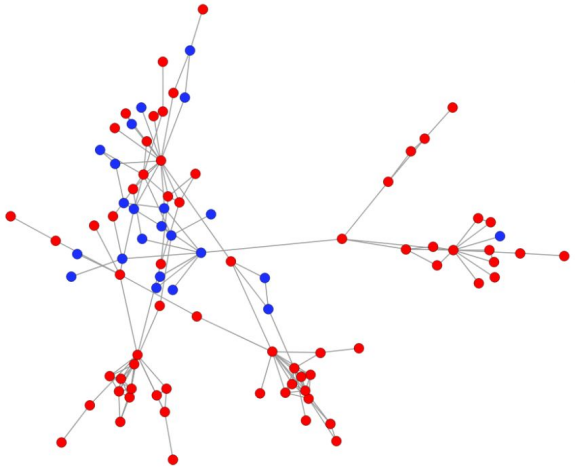
- Protein location
- Functional terms
- Accession Numbers

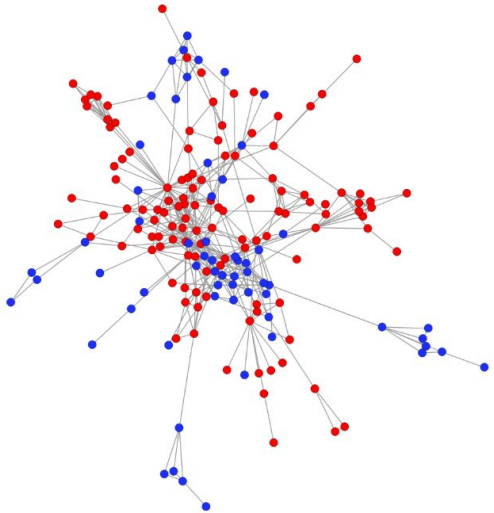
3. Visualize and explore network.

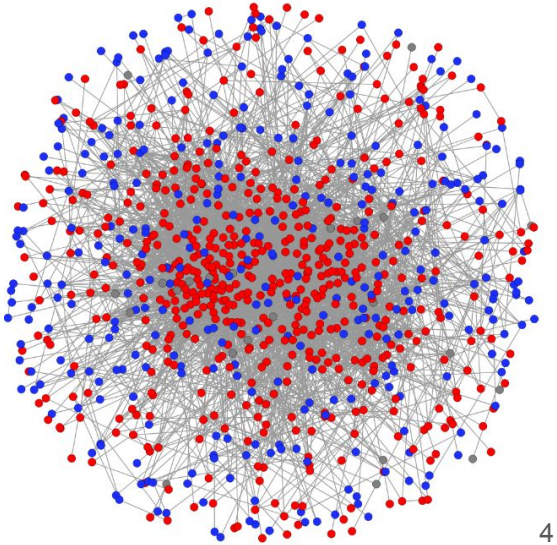


Visualization Objectives

- Create intuitive, user-friendly vis tool for biologists, in contrast to Cytoscape and NetworkAnalyst.
- Deal with the “hairball” problem as networks get bigger.

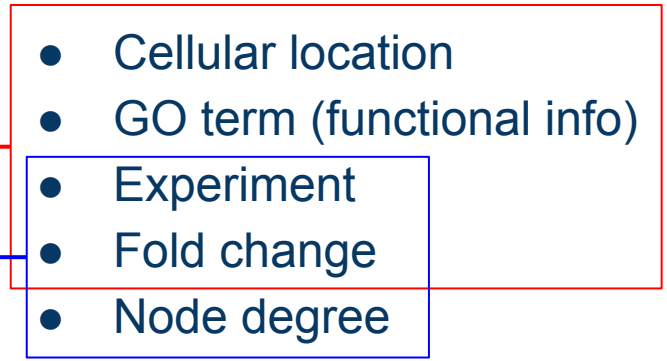






Framework

- **What?** Undirected network in node-link representation with associated attributes.
- **Why?** Discover, annotate, search and identify groups of genes of biological significance and generate new hypotheses.
- **How?** Filter nodes with attributes and encode some visually. One view for the whole network and one for the filtered network. Table view for selected nodes.



Attributes

- **Experiment:** Categorical, 3 levels, mapped to hue.
- **Differential Expression:** derived from fold change, 2 levels, mapped to hue.
- **Node Degree:** depends on network, mapped to node size.

Filters:

- **Experiment, Differential Expression**
- **Localization:** Categorical, 8 levels.
- **Enriched GO Terms:** depends on network, have associated p-values.

Color Mapping

- (●) Significance Source
- ( ) Differential Expression

RNASeq
TnSeq
Both

Select nodes

▶ By source of interest

▼ By localization

- [ ] Unknown
- [x] Cytoplasmic
- [ ] Cytoplasmic Membrane
- [ ] Outer Membrane
- [ ] Outer Membrane Vesicle
- [ ] Extracellular
- [ ] Unknown (This protein may have multiple localization sites)
- [ ] Periplasmic

▶ By differential expression

▼ By enriched GO term

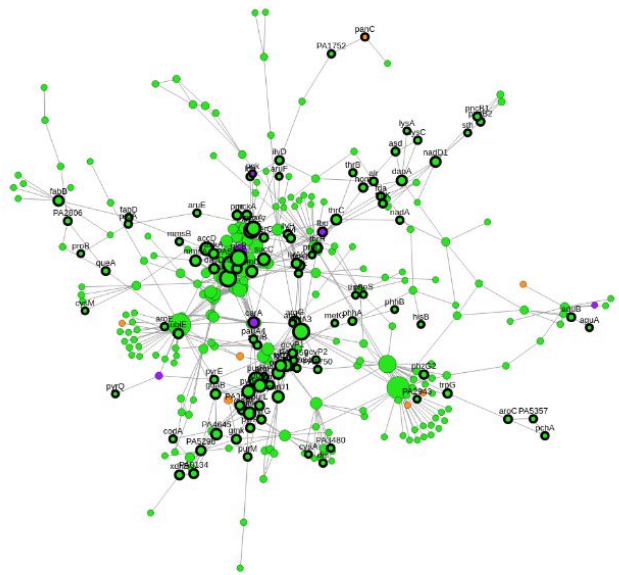
× small molecule metabolic process

Make Selection

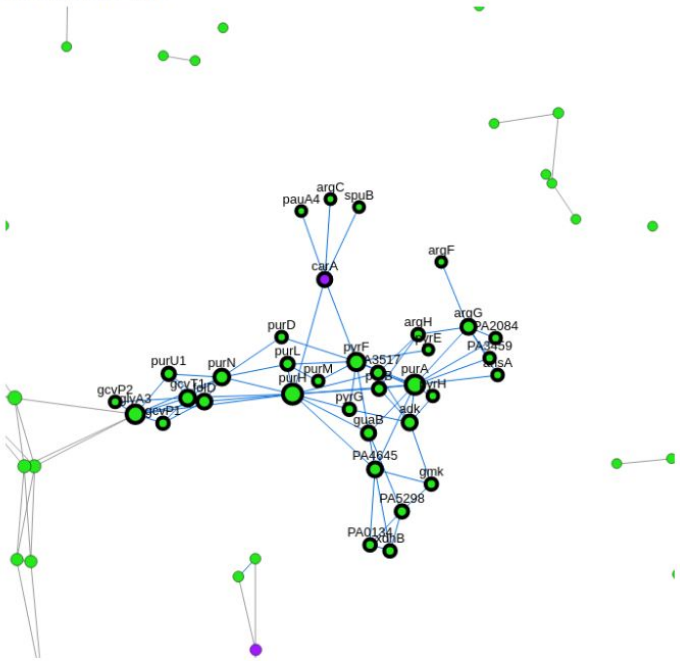
Selected 109 out of 327 nodes

Make Sub-Network

Full Network View



Sub-network View



Selected Node Details

Locus Tag	Short Name	Descripton	Log2 Fold Change	Adjusted p-value	NCBI Accession #	UniProtKB Accession #
PA3537	argF	carbamoyltransferase, anabolic	1.69	9.69e-7	NP_252227.1	P11724
PA4758	carA	carbamoyl-phosphate synthase small chain	1.35	3.17e-7	NP_253446.1	P38098
PA5263	argH	argininosuccinate lyase	0.95	7e-7	NP_253950.1	P50987
PA0134	PA0134	probable guanine deaminase	-0.83	0.00681	NP_248824.1	Q9I6Z8
PA1523	xdhB	xanthine dehydrogenase	-1.72	0.0132	NP_250214.1	Q9I3J0
PA5298	PA5298	xanthine phosphoribosyltransferas	1.29	0.00669	NP_253985.1	Q9HTQ6

Conclusions and Future Work

- Having hundreds of nodes at the same time in a single view is not useful for exploration, even with clustering.
- Database biological info very useful to filter networks with prior biological knowledge.
- Layout parameters should change depending on network size/topology.
- Implement algorithm to extract minimally-connected networks.
- Explore network statistics to identify important topological features.
- Use GO term hierarchy to design better filters.
- Add more filters by scraping other databases.
- Explore ways to show the user the filter search space.