**Meeting**

**Tuesday, 23 January 2018**

**Project objective**s: 6-8 pages PNAS

**Topic in biological network:** gene regulatory networks

**Links to Data**:

* This data consists of time-course measurements of gene expression in yeast after heat shock (i.e. exposure to high temperature), which can be found here: (<http://www-genome.stanford.edu/yeast_stress/data.shtml>, there's a download link on the left).
* We could get the gene regulatory network for yeast from here: <http://www.yeastract.com/>, which encodes which transcription factors activate or inhibit which genes.

**Objectives**: analyze time-course gene expression data as a dynamic process occurring in the network, to try to model how the gene expression is "spreading through the regulatory network"

Calculations (models and theory) to start with: merging datasets, but first, check that nodes match. If they don’t match, perhaps exclude them.

2. Create a visual

It would be useful if you meet with each other and try to narrow down the topic a bit --- perhaps in terms of which topics in biological networks you think may be most interesting. It would also be useful to see what data is available. The key goal at the moment is to narrow down the project enough to then start with some calculations (or models and theory, if you go in that direction). A source that may help, as concerns ideas, are the discussions of biological networks in Newman's book. I am not aware of a review article on biological networks specifically --- either one gets more general or more specific.

We’ll be using Python.

Lit Review: ?

Goals for next week (30 january 2018):

1. Lit review
2. See if nodes match in two datasets by merging them.

**Meeting w Professor Porter**

**Thursday, 25 January 2018**

Per Mason, look at lit review. He guesses there will be papers on this.

The models we’ve found are stochastic, or gillespie algorithm.

Per Mason, gillespie is a general term. The main ways of implementing it without bias is w a specific gillespie algorithm.

What flavors do the models have in terms of structure?

They’re a giant system of ODs. There seems to be little data to simulate systems like that (in all but a few cases).

Boolean descriptions? Logic?

Gillespie shouldn’t be too horrible to implement it. There should be some public code.

We should use it! We don’t want to spend time coding if we’re not going to invest in it. Look for pub code.

Alberta physicist @ Penn State whose been doing genetic networks and Boolean things.

James Gleeson has public code on Gillespie algorithm.

We have to send slides by next thursday, february 1 for a 10 minute presentation so the professor sees where we are and so we could get feedback from others.

Presentation is going to be a combo of what we js told Mason and a mini lit review. Have some # of slides that can fit in 10 minutes.

We have 5 groups and 50 minutes, so we have 10 minutes each.

We have to find a meaningful ways to compare genes bc there are only about 5 time points. Look at lit review to see how others have handled this problem.

Mason guesses their solutions from before are far from perfect.

There are also multiple versions of the network. Is it better to keep sparser version where we’re more sure of edges?

Mason advises us (subject to computational time) is to run code on both versions and see which is better. In terms of computation time, it may take time. We have complete and subnetwork. We can run it in entire or subnetwork.

It is about 7000 nodes. These types of simulations should be very doable computationally. There are fast Gillespie methods, and 7k nodes is relatively small.

Do exploration before we do anything definitive.

How should we find good layout of nodes?

Mason says that in Python there are a few things out there. Bc we have biosystem, it’d be nice to see where it is located in space. There are number of network vis tools. The standard methods are pretending the modes are masses and edges are springs and let them relax. Sometimes called spring embedder. Also look at frecturman reingold.

If we can see it in 2d, make them static and then edges might have color based on flow. 8k nodes is at point where it is hard to visualize. We can do analysis on whole set and visualize subset. We can do clustering and visualize straining of network.

B careful not to spend too much time on visualizations and be sure to do analysis. Time-dependent visualization can be really difficult bc some packages can have annoying feature.

Gene expression level will change node itself, so we should keep it 2d and change node color. We’ll prob have to show subset of nodes.

**Meeting**

**Tuesday, 30 January 2018**

We seem to be ahead of the curve wrt what other people in the class are doing, so if we don’t say that much more than what we said to prof Porter, it would probably be ok.

Do we need a stochastic model? Can we get away with a deterministic model?

Perhaps we could each look at a specific model, prepare slides & 2 minute presentation on that model, and then paste together for presentation

We should probably meet on Thursday afternoon/evening to finalize and submit the presentation, and to go over what we’ll say (practice run(s)?)

Need to do a good amnt of introduction (what’s a GRN? What data we have; what we want to do with data, what problems might arise with data)

Perhaps models should take a secondary role (abt ½ presentation)

-- not in depth (“this is a discrete model with boolean values, used in past for these applications” “This is a stochastic model which gives us distributions too”, etc)

-- relatively superficial. Just an overview

Is visualizing the network actually a worthwhile endeavor (maybe not?)

We can talk about which data we should use (Mauricio is talking about two bodies of evidence for edges in the network, one is more permissive than the other, the less permissive one might be more reflective of the actual structure of the network, but it’s possible the other, more correlational one will give us a better simulation. We should try both)

Talk about first step is to look at heat-shock subnetwork

Possibly only a handful of nodes will matter (probably those in the heatshock subnetwork)

Introduction:

-- What is a GRN? (Jason)

-- Network Data

-- Timecourse Data (Mauricio)

Models

-- Logic (binary) (Jason)

-- Deterministic ODES (Joanna)

-- Stochastic DEs (Eli)

-- Hybrid models (Mauricio)

Conclusion (potential problems, etc)

Eli will contact Joanna abt Network Data vs Conclusion

Need to figure out time for thursday (tentatively 5-6pm)

**Meeting**

**02/06/18**

Task before the meeting next week with Mason

1. Get Aa and Ai (the activation and inhibition adjacency matrix)
2. Look at affine dynamical systems
3. Look into simulations a logical models

**Meeting**

**Tuesday, 13 february 2018**

Notes