

We want to employ Bayesian network built based on Gene Expression, Gene Genotype and Phenotype data to find the common factors in between different mental diseases as well as different addictions and further, to find commonalities between mental diseases and addictions. We want to have a fully automated pipeline to feed as input GEO data, dbGaP, etc., and build Bayesian networks and compare the networks and as an output get the common factors between diseases. We have two similar pipelines already designed but they are not complete and particularly they lack the ability to compare Bayesian networks

From engineering point of view the process can be divided into following steps:

**Step1:** To deal with two sources of inputs (GEO and dbGaP) and produce the multinet Bayesian networks for each disease. The different inputs need to go through different process (i.e. pruning is different) MulitnetBuilder.Java is already written for making the Bayesian network. It uses Weka’s Jar files.

**Step1.1:** To get the GEO data and produce the multinet Bayesian networks. Neena’s R file is a good start for this. The R file produces the arff files and then we need to use multinetbuildr.java to make the Bayesian network. Some of the R code is irrelevant to our goal. Ignore until line 250. Rest of the code can be helpful. We might want to look for a solution to avoid ignoring too many datasets. Getting the data from server should become smarter meaning automated and precise. Coding this in some other language is a considerable option. The output of the whole step is GEO multinet Bayesian networks files, making it possible to build Bayesian networks for each disease.

**Step 1.2:** To get the dbGaP data and produce the multinet network. In this network nodes are going to be SNP. The amount of data is much more that GEO so we probably need a stricter feature selection. Skanda and Vinnie’s code are related to this.

**Step2:** To either integrate the previous networks into a full multi-net Bayesian networks (Alber Wu paper seems to be helpful) or make biological pathways network on top of the previous networks and compare them. To build the pathways we can use Skanda’s code. Then we have to compare the pathways to find the factors unique to each disease as well as the ones common between diseases.

**Step3:** output the result of the comparisons in form of a document with graphs. Aaron’s code has a nice writerepot.java

Certainly, after these steps we need to run the code for many different inputs and get the material as research result. And we can hopefully write about the results.

**Timeline :**

*Before June 11:*

-reading the background papers. Getting the codes. Analyzing them. Organizing them. Meetings with people.

*June11 until June 17:*

-Reading the R code carefully. (In brief, there are parts that are irrelevant. We might need to find a way to avoid reducing number of experiments in datasets so much, if possible.)

-studying how to deal with GEO dataset without using R.

-talking to Neena and Vinnie on their previous research. (I now have full understading of Neena’s code. I need to work a little more with Vinnie.)

-talking to Milad, Mehran about the project. Milad is all set. Mehran needs further details to talk about.

-Making the time-line.

*June 18 until June 24:*

-Explaining the details of project to Mehran and deciding on which part he can help with.

-Talking to Ali about the project.

-Pinging Vinnie again to upload the documents he promised me on phone.

-trying to talk to Aaron.

*June 25 until July 1:*

-Samira and Farzin and Mehran start working this week.

-Samira to work on step 1.1

-Mehran start with step 1.2.

-I need to go over Skanda’s codes and get a way to use the pathways code.

-Need to explain the codes to Farzin and see what he can start with. Probably step3.

*July 2 until July 8:*

-Amir starts working this week. He will start with step 2.

-Need to discuss with Gil about some different ways we can do this.

-holding weekly meeting with them to make sure things go smoothly.

*July 9 until July 15:*

-step 1 should have made substantial progress.

*July 16 until July 22:*

-Step 1 is hopefully done in this week.

-step 2 should have substantial progress.

-step 3 may be done too.

*July 23 until July 30:*

-Trying to get all the code running smoothly. We can try some samples.

*August:*

Running Experiments. And writing papers.