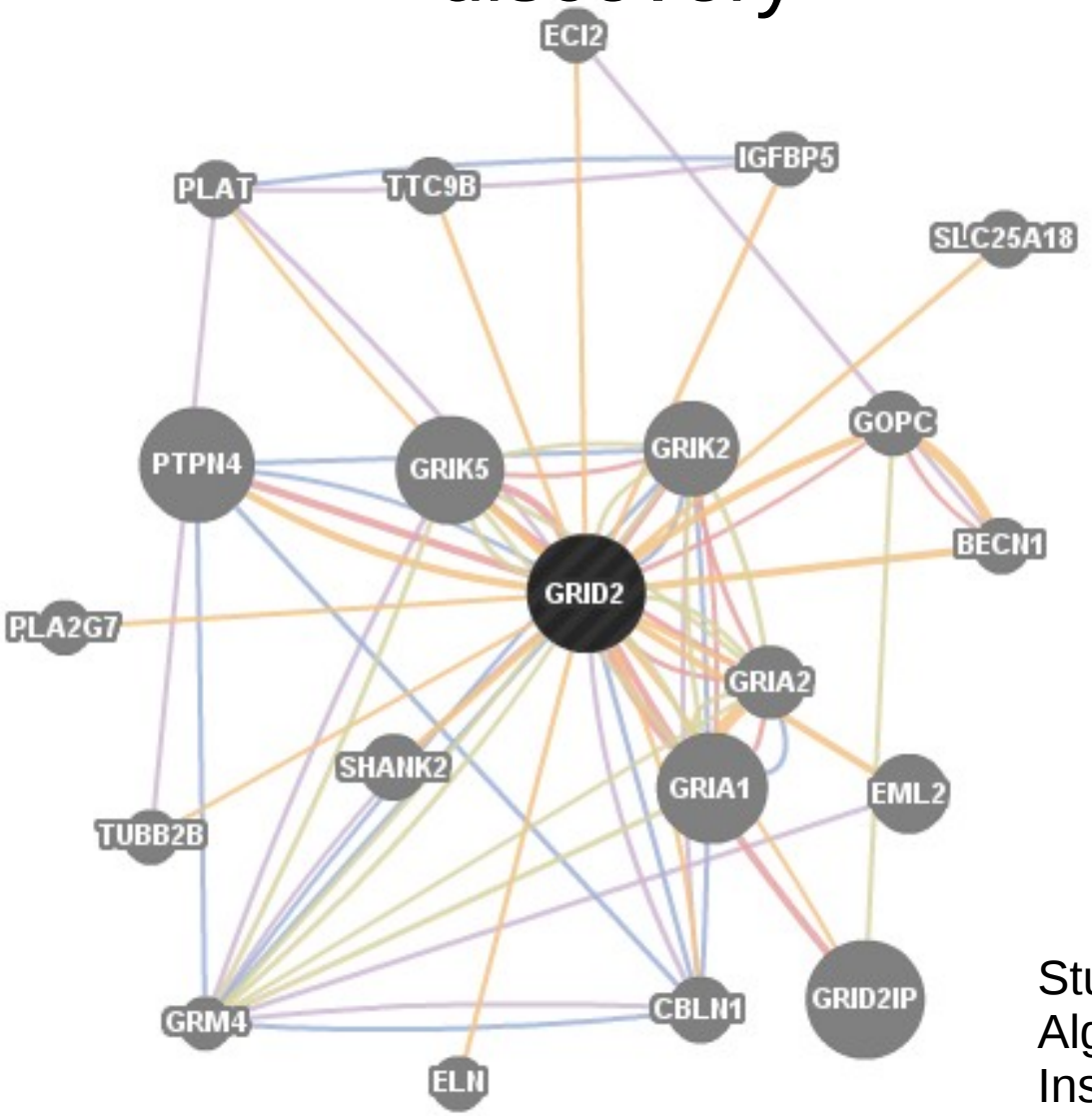
Machine learning in disease subnetwork discovery



Student: Bogdan Kirillov,
Algorithmic Bioinformatics,
Institute of Bioinformatics

Research supervisor: Son Pham, Salk Institute

Why?

- The causes of most neuronal diseases are still unknown
- Neuronal diseases (f. e. autism) are severe and currently incurable
- We have a large whole genome sequencing dataset
- We can explore it with graph theory and machine learning

The Data

- A large (~PetaB) dataset of whole genome sequencing data
- A genemania.org homo sapiens combined gene interaction network

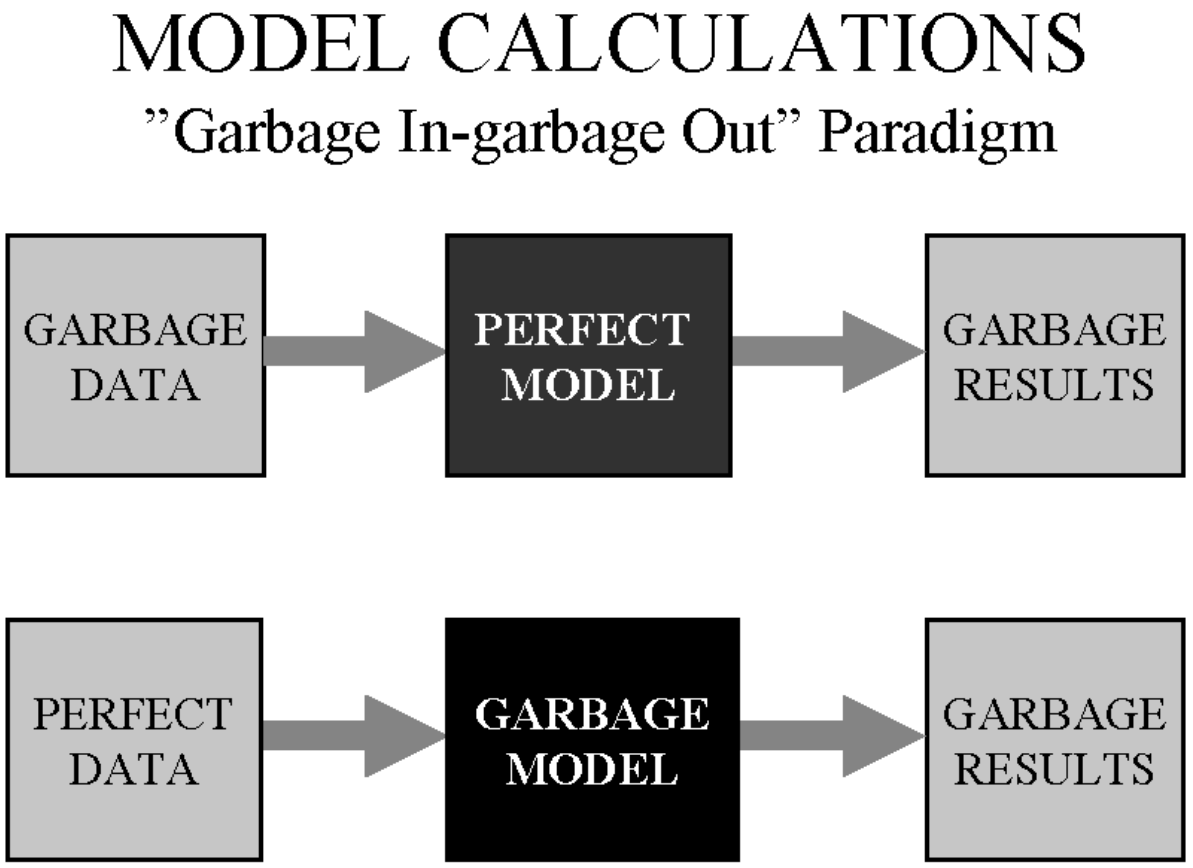
  Directed multigraph with 19264 nodes and 7290094 edges

```
>>> print(c.graph.summary())
IGRAPH DNW- 19264 7290094 --
+ attr: name (v), group (e), network (e), weight (e)
>>>
```

What needs to be done?

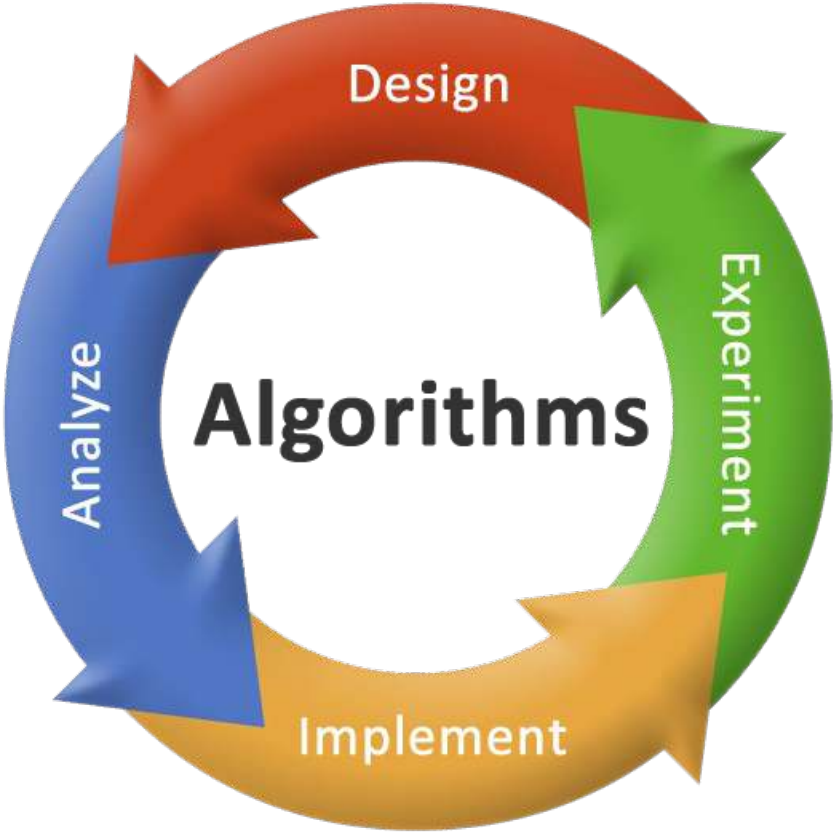
- Gene interaction network verification

  We have to check whether the network is correct, because of the first rule of data analysis: check your data.



What needs to be done?

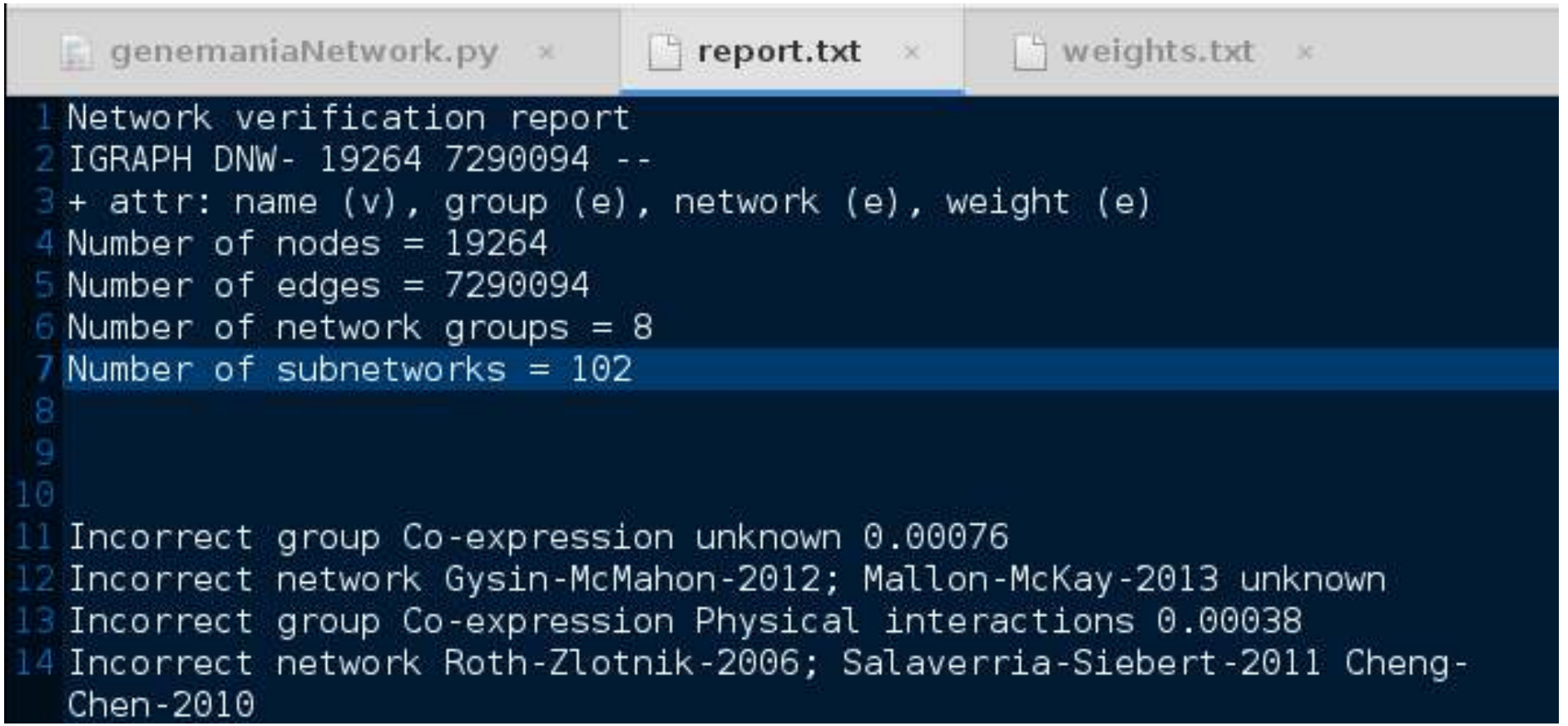
- Algorithm design and implementation



What was done?

- Genemania network verifier

  https://github.com/kspham/diseasenetwork



These “Incorrect network” and “Incorrect groups” are yet to deal with

Plan till December

- Get rid of incorrects in the verifier
- Design and implement an algorithm