# CONNECTOMES GENERATIVE MODELS

Ayagoz Musabaeva, Anvar Kurmukov, Yulia Dodonova

## 0. PROBLEM

1. Standard data set consist of dozens - hundreds of objects
2. Data dimension is hundreds - thousands features

Solving classification task “as is” lead us to either overfitting or very strange validation approaches. So we want to improve classification model, or at least increase it robustness by :

1. Reducing data set dimension space (smart feature selection)
2. Increasing size of data set **(this)**

## 0. WHAT ARE WE TALKING ABOUT?


GOALS :

- Improve classification model (could be real)
  - Data augmentation
    - Non-Embedded graph generators
    - Embedded graph generators
  - Data normalization
    - Small-world approach
- Understanding brain structure (ultimate bonus)

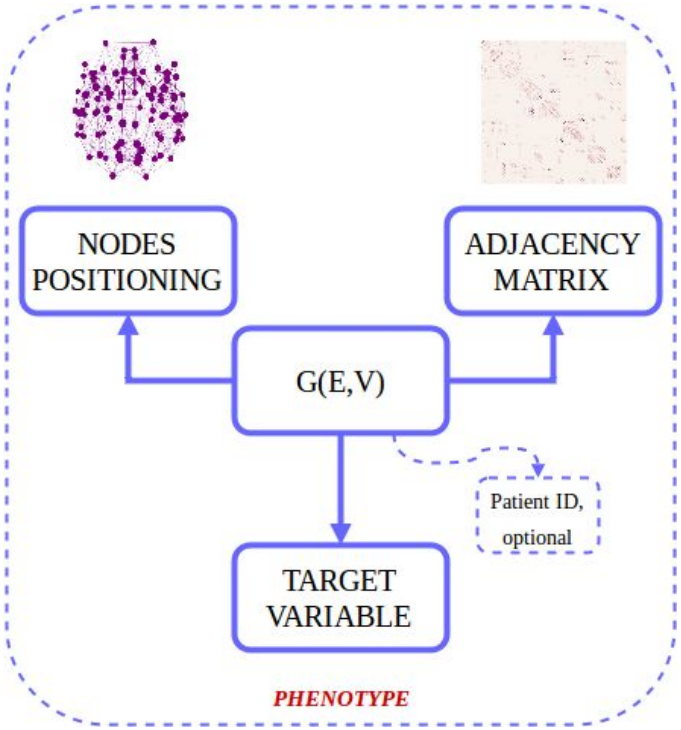
# 1. DATA AUGMENTATION

“It’s all about data!”

<u>What we have :</u>

Set of weighted, strongly connected graphs with equal set of vertices (V) and different set of edges (E). For each graph G(V,E) we have:

- Weighted adjacency matrix, A
- Set of geometrical nodes positions (x, y, z), C
- Target variable (Normal or Pathological), T
- *Optional : patient indicator (in case of multiple connectomes from single patient), ID*



What we want : *(spoiler Data augmentation)*

We want to generate new graph, based on existing ones.
Following questions may arise here:

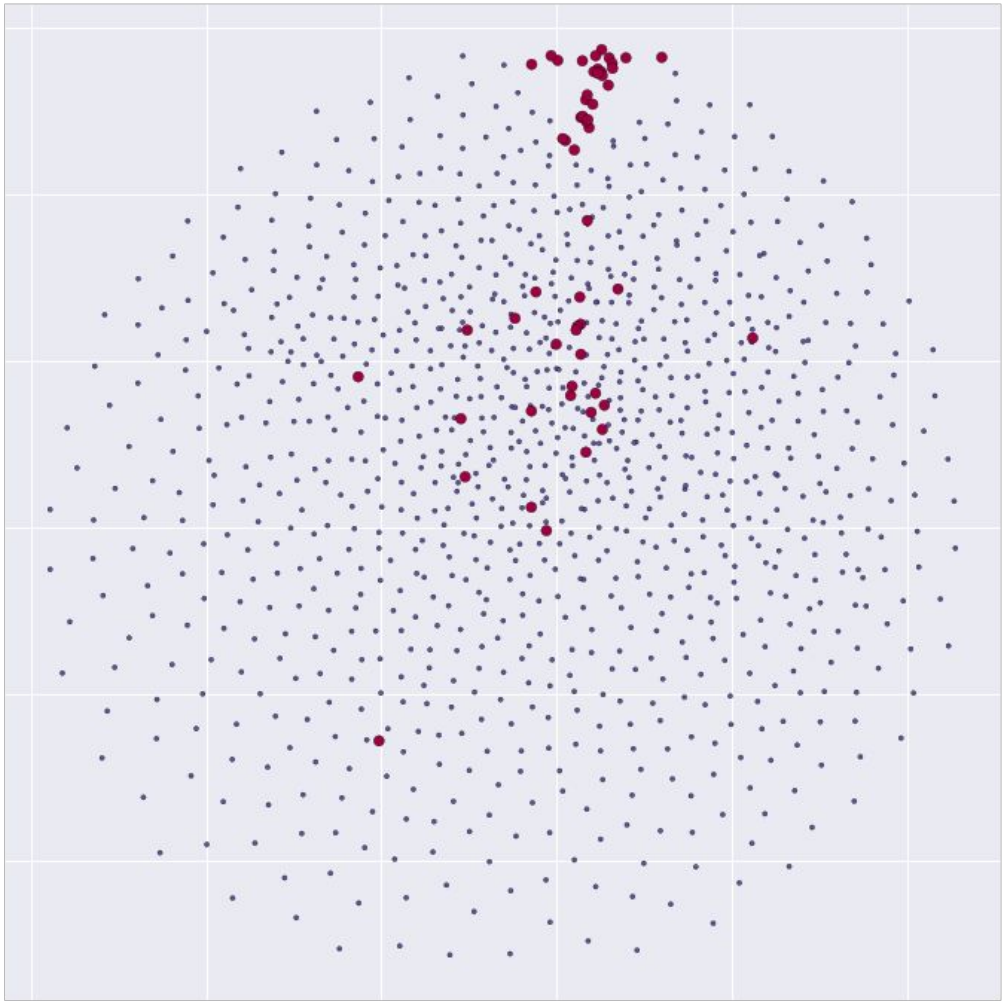
- What type of model should we use?
  - Non-Embedded
  - Embedded
  - *More? Or another type classification?*
- What subset of existing graphs should we use to generate new one? (*)
  - Single graph
  - Single patient (if possible)
  - Subset of graphs based on target variable
  - Whole set of existing graphs
- Do we need weighted graphs?
- What about generating not graphs as Adjacency Matrices but graphs Laplacians?
- Once we have new graph, how to choose its target variable?
- Once we have set of new graphs, how to validate their goodness?
- *more?*

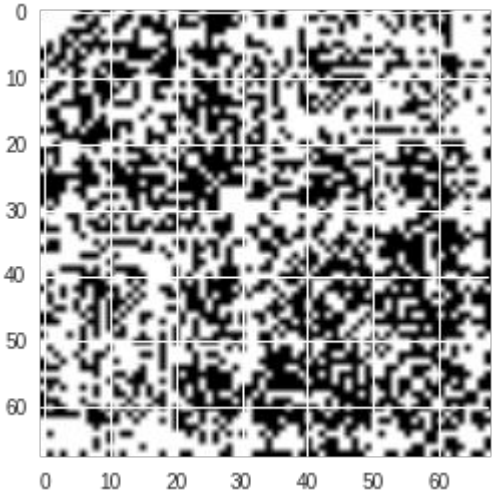
## 2. WHAT OTHERS DID?

List of existing approaches (I’ll show you some articles right now)
So someone could say couple of words about existing approaches

## 3.HOW TO OBTAIN MODEL PARAMETERS

In our previous papers we show that classification based on similarity of graphs partition works well, moreover some algorithms gives us partitions comparable with graph geometry
(geometrical nodes positions), without any knowledge about it. Which means, that partitions of Healthy are more or less different from partitions of Pathological patients (e.i. Alzheimer). This gives us an opportunity to extract some structural parameters from given partitions in order to use them for generative model (i.e. in-out community degrees, average community size etc.). Generating networks with similar structure to some phenotype may be consistent.
Again (*) is open.




0
10
20
30
40
50
60
0
10
20
30
40
50
60