

GCN (Graph Convolution Netowrk)

Network approaches have been used many times to combine and amplify signals from individual genes, and have led to remarkable discoveries in biology, including drug discovery, protein function prediction, disease diagnosis, and precision medicine

The main challenge in machine learning on networks is to find a way to extract information about interactions between nodes and to incorporate that information into a machine learning model.

Recent years have seen a surge in approaches that automatically learn to encode network structure into low-dimensional representations, using transformation techniques based on deep learning and nonlinear dimensionality reduction.

Why protin-portun interaction using this approach ?

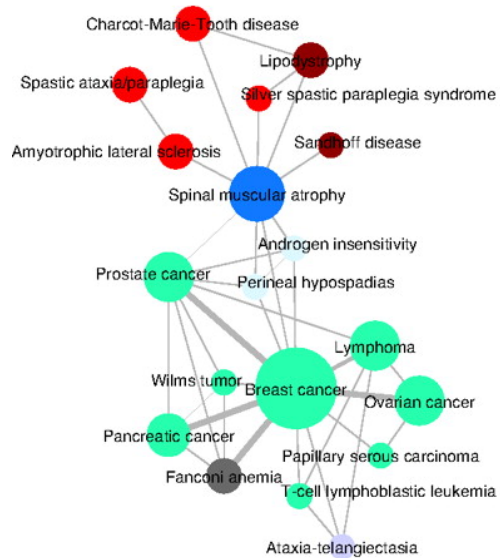
Proteins involved in the same disease tend to interact with each other. How do representation learning methods use this information to detect disease modules? How can we prioritize candidate disease proteins?

Genes associated with similar diseases are likely to interact and have similar expression

Human diesease network

In the following figure we can see that in **HDN** those disease are connected with each other have same genome structure connection which can be seen in **DGN**

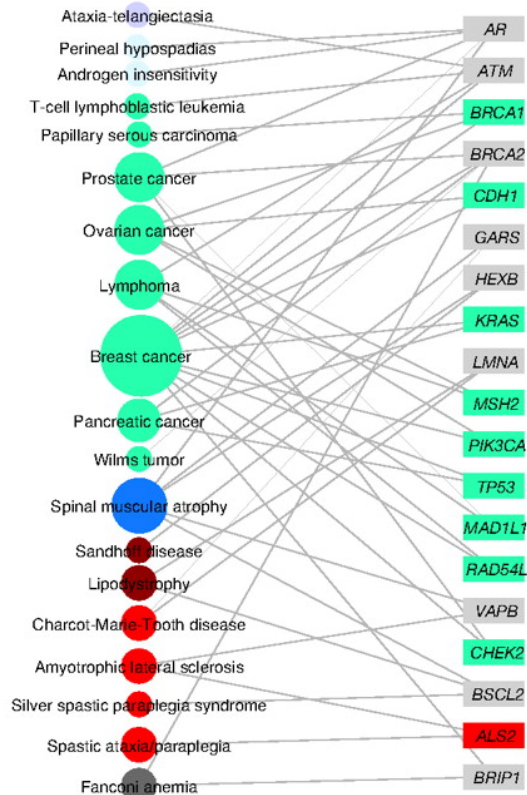
*Human Disease Network
(HDN)*



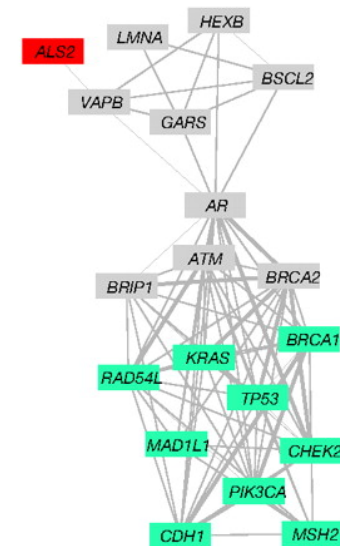
DISEASOME

disease phenome

disease genome



*Disease Gene Network
(DGN)*



Different ways to analyze network

Predict type of given node

Node classification (Classifying the function of proteins in the interactome!)
*an **interactome** is the whole set of molecular interactions in a particular cell*

Predict whether two nodes are linked

Link prediction (Predicting which diseases a new molecule might treat!)

Identify densely linked clusters of nodes

Community detection (Identifying disease proteins in the interactome!)

How similar are two nodes/networks

Network similarity

Network analytics lifecycle include supervised learning

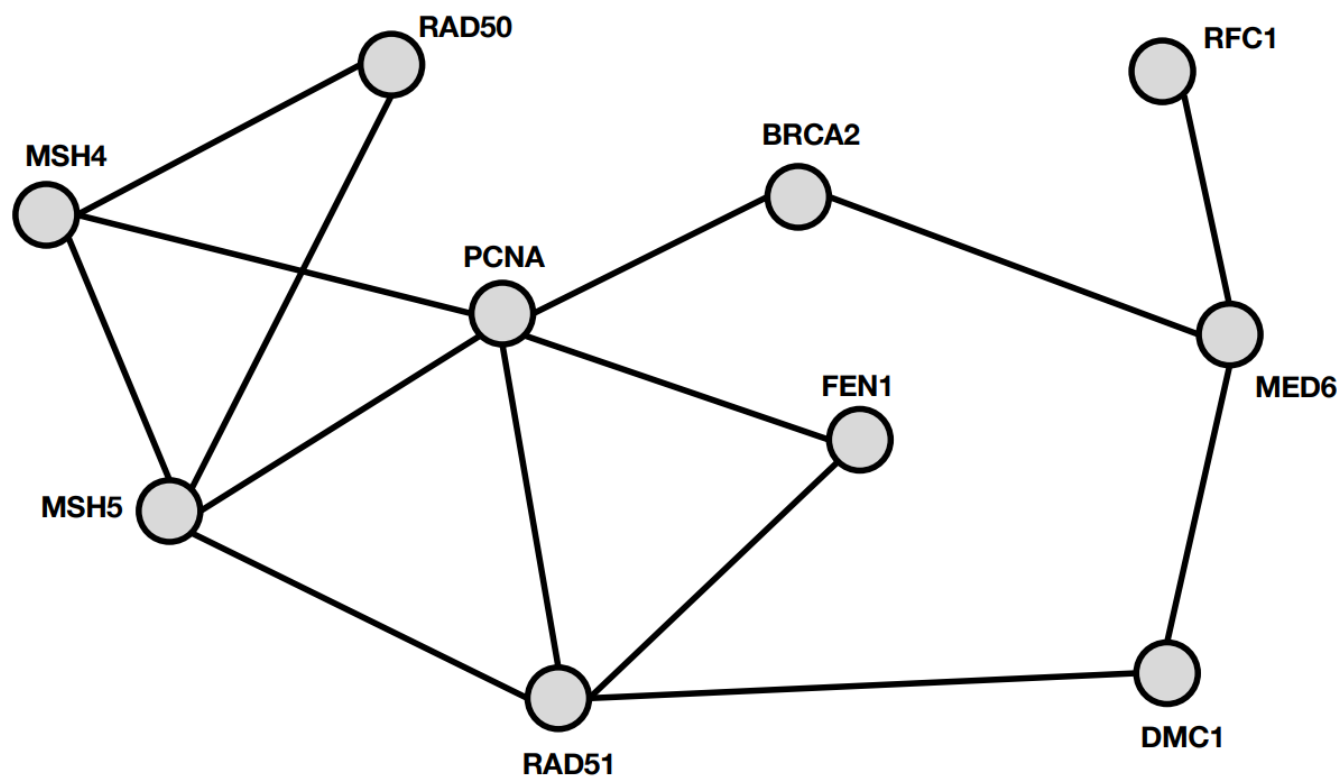
Modern deep learning ToolBox is designed for Grid or Sequence.
It is not good for the unstructured data like Graph networks.

What we have to do is find a **Embedding function F** such that
 $F(\text{Graph Network}) = \text{two dimensional node embedding}$,
Which is one of the node embedding technique

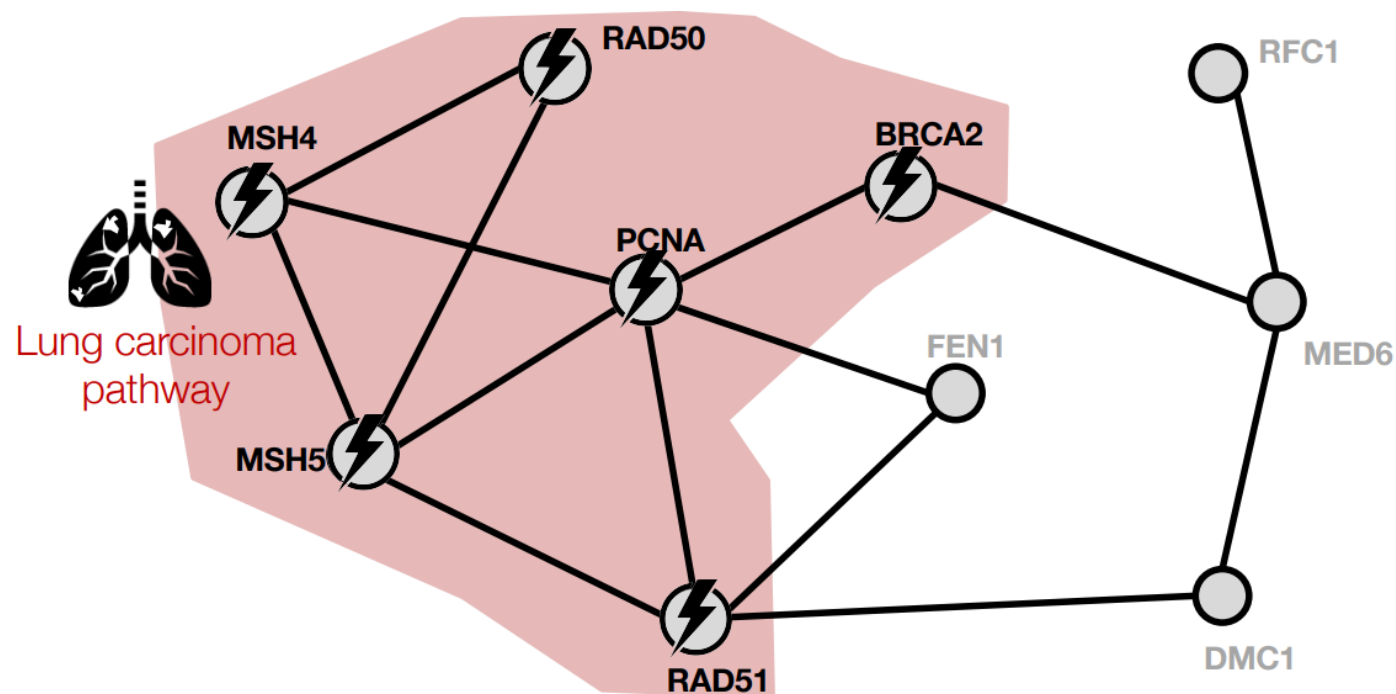
Basically in human body based on the interactome we can decide that what kind of disease person can have. So, If we figure out particular interaction in proteins, disease can be spotted perfectly. Our task is to figure out this interaction.

Look at following example:

This is Human interactome

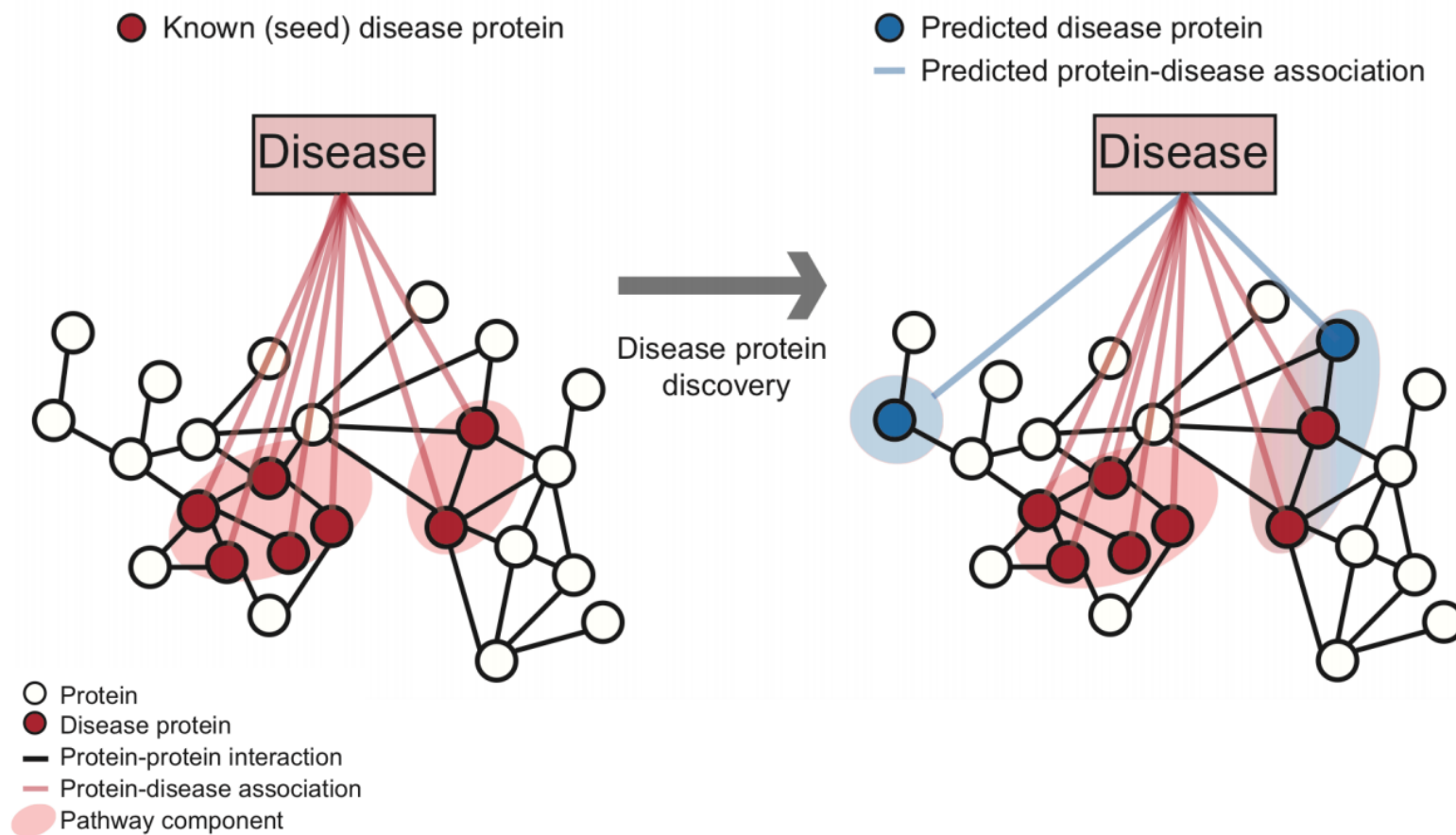


Now, based on these protein interaction, particular disease can be figure out.
here protin **BRAC2** stand for **Breast Cancer 2** simimilarly others are protin molicules.



Now if you will see the interaction of the nodes in specific subgraph then we can figure out disease realated to it.

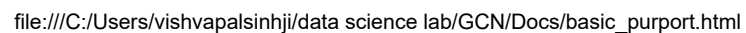
ULTIMATE TASK: We have to findout new protin which is responsible for the disease in same network



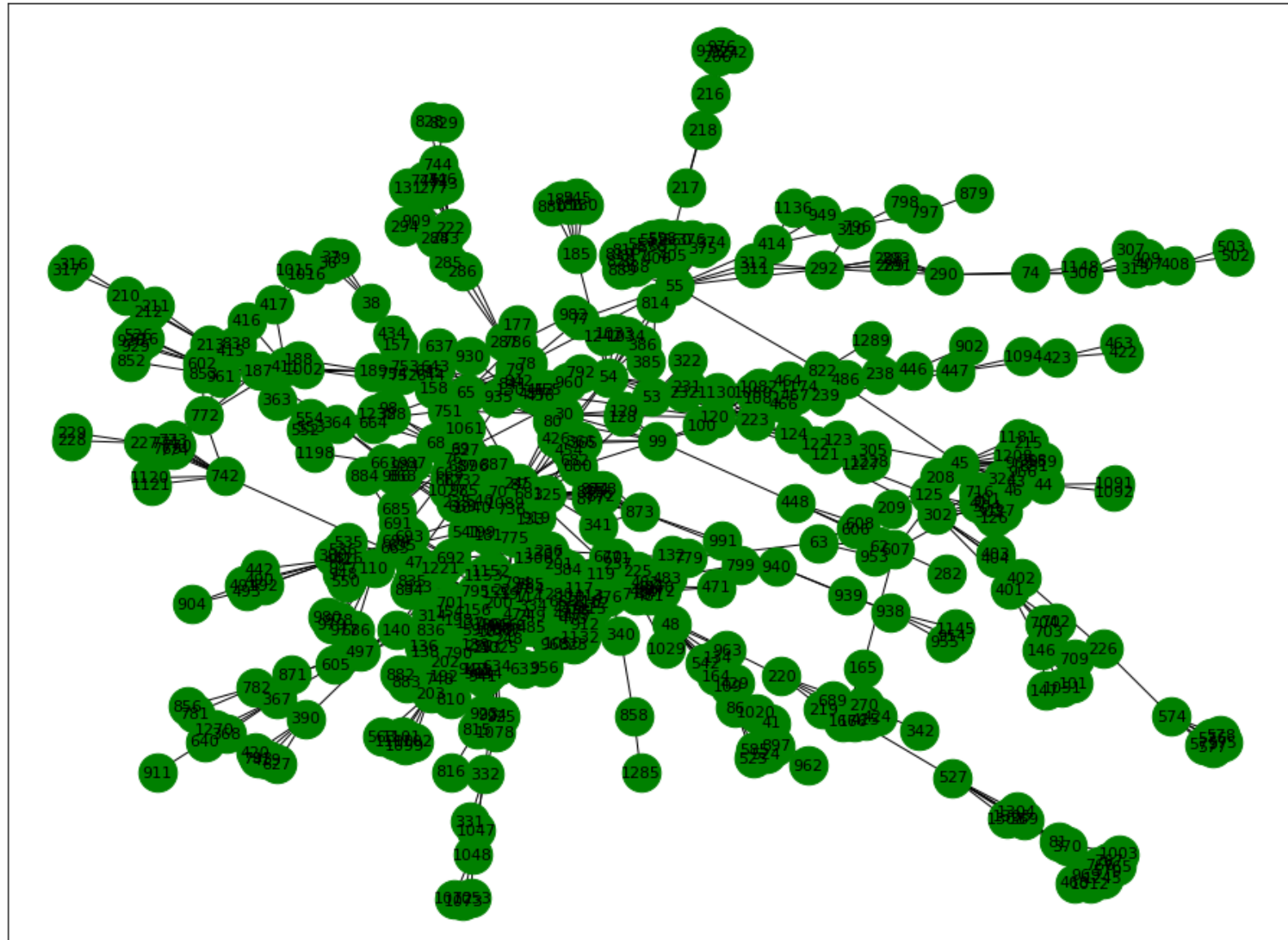
What we can do form the data given to us?

We can see the graph structure in a **yeast.edgelist** (basically protin) as well as graph sturcture of disease from **dissome.edglist**
Further we can also see the node corosponding name into the disease.label

Yeast edgelist



Disease edgelist



So, we have to find out that which proteins combination in subgraph indicate a specific disease.

REFERENCES : [Stanford Tutorial \(http://snap.stanford.edu/deepnetbio-ismb/\)](http://snap.stanford.edu/deepnetbio-ismb/)
(<http://snap.stanford.edu/deepnetbio-ismb/>)

(<http://snap.stanford.edu/deepnetbio-ismb/>)

(<http://snap.stanford.edu/deepnetbio-ismb/>)

In []:

(<http://snap.stanford.edu/deepnetbio-ismb/>)

(<http://snap.stanford.edu/deepnetbio-ismb/>)

(<http://snap.stanford.edu/deepnetbio-ismb/>)