Application of Node2vec: Predict Optimized Treatment for Depression CS224W Project Report

Minakshi Mukherjee: adaboost@stanford.edu
Suvasis Mukherjee : suvasism@stanford.edu

I. *Abstract*

For the past 60 years, the anxiety and depression medications are prescribed to patients based on The Hamilton Depression Rating Scale (HDRS)[1]. The HDRS[1] does not take into account the neuro biomarkers as it is very expensive to do FMRI on all patients. Goal of this project is to identify clinically applicable imaging biomarkers and establish intrinsic functional connectivity to predict efficacies of three antidepressants: Sertraline, Venlafaxine, Escitalopram from the small dataset of 128 patients collected from Williams PanLab, Precision Psychiatry and Translational Neuroscience, Stanford Medicine iSPOT-D project. There is a need for markers that are predictive of remission and guide classification and treatment choices in the development of a brain-based taxonomy for major depressive disorder (MDD) that affect millions of Americans.

II. *Introduction*

Our project analyzes the iSPOT-D dataset for 128 patients with images from functional magnetic resonance imaging(FMRI) data, uses different Graph Analysis techniques and computes the functional scores based on multiple brain image attributes. We compare the correlation between functional score and Hamilton Score to predict the antidepressants linked to different brain attributes. The topological structure of functional brain network plays an important role in major depressive disorder(MDD). We built a network using these highly connected and mostly unexplored interdependent components, explored the dataset using some of the common network construction techniques to obtain network statistics like density, cluster coefficient and took a deep dive into community detection.

We created patient nodes in our network graph where each node contains the feature attribute related to multiple social bio-markers based on FMRI data as well as the antidepressants taken by them. Hence, this patient network is associated with rich attributes. Our goal is to find the social network embedding. We projected the patient information into a low-dimensional embedding space. Since the network structure and feature offer different sources of information, it is crucial to capture both of them to learn a comprehensive representation of each patient feature. We took into account both homophily and the network structure to get more informative node representation. Our objective was task-independent feature learning, it is an unsupervised problem. There are no fixed node ordering or reference point. We used embedding methods that preserve both the structural proximity and attribute proximity of social network.

We denote a patient network as G = (U, E, A), where U = $\{u_1, ..., u_M\}$ denotes the patients, E = $\{eij\}$ denotes the links between the patients i and j, and A = $\{Ai\}$ denotes the attributes of the patient i. We created undirected and unweighted graph, so each edge $\{eij\}$, connecting patient i and patient j is associated with a weight = 1. For structural proximity we used the nodes u_i and u_j with a link e_{ij} between them. We applied node2vec that controls the random walk by balancing the breadth-first sampling (BFS) and depth-first sampling (DFS) to generate the embedding. For attribute proximity, we meant the proximity of the nodes represented by the patients using all the feature attributes. The attribute intersection of patient i and patient j, denoted by A_i and A_j gives the attribute proximity of the nodes u_i and u_j. By enforcing the constraint of attribute proximity, we can model the attribute homophily because the patients with similar attributes will be

placed close to each other in the embedding space.

III. *Related Work*

A. *Functional Score*

The objective is to create a functional score for patient by leveraging the network structure and rich information available in the dataset. We used the word "feature vector" to denote the patient's clinical biomarkers. Our iSPOT-D dataset contains several clinical biomarkers related to patient, e.g. social and occupational functioning assessment scale(SOFAS)[8], brain scan data from brain regions Amygdala [9], Insula [10], Nac (nucleus accumbens) [11] known to control human behavior and other social attributes of a patient like, age, gender and education. Functional score is dictated by these attributes. Functional score takes into account both the structural proximity and the feature vector proximity of the patient node in the graph.

In this section we plan to summarize patient attributes and network embedding method like node2vec.

At the outset we tried to understand the homophily effect among the patients in the dataset. The homophily principle, "birds of a feather flock together" is one of the most striking and robust empirical regularities of social life [7]. Hence, in graph analysis, nodes that are highly interconnected and cluster together should embed near each other. SOFAS[8] captures patient's level of social and occupational functioning and is not directly influenced by the overall severity of the individual's psychological symptoms. Patient's brain scan data studies functionally central structure between amygdala [9], basal ganglia, mesolimbic dopaminergic regions, mediodorsal thalamus and prefrontal cortex, the nucleus accumbens[10] appears to play a modulative role in the flow of the information from the amygdaloid complex to these regions. Dopamine is a major neurotransmitter of the nucleus accumbens and this nucleus has a modulative function to the amygdala-basal[9] ganglia-prefrontal cortex circuit. Together with the prefrontal cortex and amygdala[9], nucleus accumbens[11] consists of a part of the cerebral circuit which regulates functions associated with effort. It is anatomically located in a unique way to serve emotional and behavioral components of feelings. It is considered as a neural interface between motivation and action, having a key-role in food intake, reward-motivated behavior, stress-related behavior and substance-dependence. It is involved in several cognitive, emotional and psychomotor functions, altered in some psychopathology. Moreover it is involved in some of the commonest and most severe psychiatric disorders, such as depression, schizophrenia, obsessive-compulsive disorder and other anxiety disorders, as well as in addiction, including drugs abuse, alcoholism and smoking. The feature vector of the patient reveals a significant detail which is not accommodated in the Hamilton score. We tried to embed nodes from the same network community and from the same structural roles in the graph(e.g., hubs) closely together.

B. *Network Embedding*

We investigated some earlier works on unsupervised learning algorithm that computes low dimensionality and neighborhood preserving embeddings of high dimensional data. Local Linear Embedding (LLE)[12] and Laplacian Eigenmap[13] first transform data into an affinity graph based on the feature vectors of nodes (e.g., k-nearest neighbors of nodes) and then embed the graph by solving the leading eigen vectors of the affinity matrix. Node2vec[14] and DeepWalk[15] are some of the recent works focused more on embedding an existing network into a low-dimensional vector space to facilitate further analysis and achieve better performance than those earlier works. In node2vec [14] the authors modified the way of generating node sequences by balancing BFS and DFS, and achieved performance improvements. However, all these methods only leverage network structure. Patient profile contains valuable attribute information. Purely structure-based methods fail to capture such valuable information, this leads to less informative embeddings.

C. *Network enhancement(NE) as a general method to denoise weighted biological networks*

Denoising dataset is necessary before analysis. This paper by Jure Leskovec et al.[3] explores a mathematical approach to extract noise from undirected weighted graph. It intends to replace row-normalized transition matrix with a more robust symmetric Positive Semi Definite(PSD) doubly stochastic matrix. The NE diffusion

technique preserves the eigenvectors and increases the eigengaps for the large eigenvalues. The re-weighting is helpful when the noise in the network is present in the eigen direction where the eigenvalues are small. This has advantage over PCA technique where the eigen spectrum is truncated at a certain threshold. NE defusion technique helps in reducing network noise and offers better quality network performance analysis.

The denoising algorithm presented in the above paper treats all the nodes as independent and identically distributed(i.i.d), hence small subset of high confidence nodes are ignored. However, the algorithm can take advantage of the small amount of accurately labeled data to denoise networks. The paper does not discuss mechanism to extract accurately labeled nodes with high confidence.

Initially, we thought to improve the algorithm on this deficiency because we have a very through clinical data with all the features presented, hence, we cannot make i.i.d assumptions when there are obviously socially correlated factors that contribute to depression. Finally, we used node2vec to identify feature embedding instead of using the algorithm presented in the paper.

IV. *Methods and Algorithm*

Patient networks are more than just links; patients biomarkers are very expensive information and provides a rich set for patient feature vectors. To learn more informative representations for patients, it is essential to capture the attribute information.

In order to create a new functional index, we will develop a functional/social score of the patient based on embedding methods that preserve both the structural proximity and attribute proximity of patient network.

Structural Proximity denotes the proximity of patients that is evidenced by links. For nodes $\{u_i\}$ and $\{u_j\}$ representing patients i and j, if there exists a link $e_{\{}ij\}$ between them, it indicates the direct proximity; on the other hand, if $\{u_j\}$ is within the context of $\{u_i\}$, it indicates the indirect proximity. In our method, we apply the walking procedure proposed by node2vec [14], which controls the random walk by balancing the breadth-first sampling (BFS) and depth-first sampling (DFS). We used the term neighbors to denote both the first-order neighbors and the nodes in the same context for simplicity.

Feature Proximity denotes the proximity of patients that is evidenced by features. The feature intersection of $\{A_i\}$ and $\{A_j\}$ for patients i and j indicates the feature proximity of nodes $\{u_i\}$ and $\{u_j\}$. By enforcing the constraint of feature proximity, we can model the feature closeness effect, as patients with similar features will be placed close to each other in the embedding space.

Network structures uses only the patient ID which can be represented in a M-dimentional sparse vector with the 1 at its i^{th} element and 0 elsewhere. The structural proximity is a function f which maps 2 nodes u_i and u_j for patients i and j to their estimated proximity scores.

Probability that node u_j is connected to node u_i is

$$p(u_j|u_i) = \frac{exp(f(u_i, u_j))}{\Sigma_{k=1}^{M} exp(f(u_i, u_k))} \tag{1}$$

Structural proximity of a node u_i with respect to all its neighbors $j \in N_i$ is the conditional probability of a node set N_i given node n_i, denoted as $p(N_i|u_i)$.

$$p(N_i|u_i) = \prod_{j \in N_i} p(u_j|u_i) \tag{2}$$

$j \in N_i$ where $N_i = \{set\ of\ neighbors\ of\ u_i\}$.

Global structural proximity is given by the likelihood function for the global structure:

$$l = \prod_{i=1}^{M} p(N_i|u_i) = \prod_{i=1}^{M} \prod_{j \in N_i} p(u_j|u_i) \tag{3}$$

We calculated the pairwise proximity $f(u_i, u_j)$ between patient nodes u_i u_j as an inner product of the embeddings of the feature vectors of patients i and j. The feature vector consists of 11 normalized attributes, some of the important ones are: 3 antidepressants Sertraline, Venlafaxine, escitalopram, 5 FMRI brain scan data from brain region Amygdala, Insula and Nucleus Accumbens, 3 social and occupational functioning assessment scale(SOFAS) scores.

By using node2vec, we calculated embeddings emb(u_i) and emb(u_j) for patient nodes u_i and u_j .
$f(u_i)$ = feature vector of node u_i for patient i
$f(u_j)$ = feature vector of node u_j for patient j

From equation (3),

$$l = \prod_{i=1}^{M} \prod_{j \in N_i} p(u_j|u_i) = \prod_{i=1}^{M} \prod_{j \in N_i} \frac{exp(f(u_i, u_j))}{\sum_{k=1}^{M} exp(f(u_i, u_k))} \tag{4}$$

where

$$f(u_i, u_j) = f(u_i)^T f(u_j) \tag{5}$$

We maximize, the conditional link probability over all the nodes with respect to all the parameters Θ.

$$\Theta^* = argmax_\Theta \prod_{i=1}^{M} \prod_{j \in N_i} log \frac{exp(f(u_i, u_j))}{\sum_{k=1}^{M} exp(f(u_i, u_k)} \tag{6}$$

$$\Theta^* = argmax_\Theta \sum_{u_i \in M} \sum_{j \in N_i} log \frac{exp(f(u_i, u_j))}{\sum_{k=1}^{M} exp(f(u_i, u_k)} \tag{7}$$

The optimization problem in Equation-7 has two effects:
1.to enhance the similarity between any u_i and these $u \in N_i$
2.to weaken the similarity between any u_i and these $u \in N_i$.

Critique:
First problem of the model:

Equation(7) assumes that if two nodes representing the patient IDs are not linked together, they are dissimilar, but that is not necessarily true.

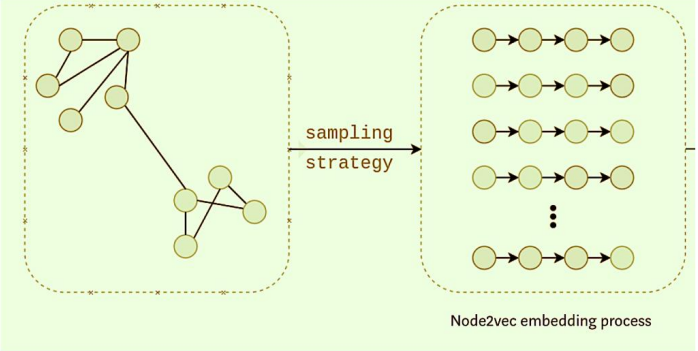
Second problem of the model:

This is linked to the calculation of the normalization constant in equation (7). In order to calculate a single probability, we need to go through all combinations of patient IDs in the network and that is NP-hard.

Due to the above two complexities, our algorithm calculates the functional score based on pairwise proximity $f(u_i, u_j)$ which is easy to derive using node2vec.

Algorithm:

Our objective is to feed quality embeddings into the algorithm. This adds knowledge to the data and thus makes the task to train the model easier. node2vec helps in extracting meaningful embeddings. The embeddings are learnt using a skip-gram neural network model[16]. node2vec uses word2vec framework to train a simple neural network with one hidden layer and provides the output probabilities of the nearby node using softmax classifier. The notion of "nearby" is implemented using the "window size" parameter of node2vec. We choose "window size"=10 to keep it computationally efficient for our data size; so it will search 5 nodes before and 5 nodes after and provide the embeddings for 10 nodes.



There are two hyperpaerameters in node2vec algorithm:

Return parameter p:
It controls the likelihood of immediately revisiting a node in the walk.
If $p > max(q, 1)$,
it is less likely to sample an already visited node and avoids 2-hop redundancy in sampling.
If $p < min(q, 1)$,
it backtracks a step and keep the walk local.

In-out parameter q:
If $q > 1$
it does inward exploration,Local view and BFS behavior If $q < 1$
it does outward exploration,Global view and DFS behavior

Summary of our project's algorithm

1.**Generate the undirected** and unweighted graph from the patient data set where each patient ID is a node. We have nodes $u_1, u_2, .., u_{128}$.

2.**Generate the feature** vectors $f(u_1), f(u_2), .., f(u_{128})$ associated to nodes $u_1, u_2, .., u_{128}$ with 13 features: age,gender,3 antidepressants: Sertraline, Venlafaxine, escitalopram, 5 FMRI brain scan data from Amygdala, Insula and Nucleus Accumbens and 3 social and occupational functioning assessment scale(SOFAS)[8] scores.

3.**Use node2vec** and generate embeddings $emb(u_1), emb(u_2), .., emb(u_{128})$ with window size=10 associated to nodes $u_1, u_2, .., u_{128}$, where $emb(u_i)$ is a vector of length 10 consisting of the embediings for node u_i.
We have used hyperparameters p=10 and q=.1 to look into homophily.

4.**Calculate Functional** Score:

For each node u_i, calculate the inner product of $f(u_i, u_k)$, where k iterates through all the embedding nodes found in step(3) above. Since "window size"=10, we will get 10 of these inner products. We averaged all the 10 inner products and output as functional score of node u_i.

Pearson correlation coefficient

Functions of Correlation Coefficient has been used extensively in psychological research, because scale-free measure of association is very important in the areas of psychology to understand effectiveness of a measure.

After getting the functional scores from all the patient nodes, we wanted to understand the association between Hamilton Score and the functional score as well as the association between SOFAS score and the functional score. Hence, we calculated two sets of Pearson correlation coefficients.

X: vector of hamilton scores for all the patients
Z: vector of SOFAS scores for all the patients
Y: vector of functional scores for all the patients as found based on our algorithm

$$\rho(X,Y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$
$$\rho(Z,Y) = \frac{\sum(z_i-\bar{z})(y_i-\bar{y})}{\sqrt{\sum(z_i-\bar{z})^2\sum(y_i-\bar{y})^2}}$$

$\rho(X,Y)$ denotes a numerical measure of dependence or association between X and Y.

Similarly, $\rho(Z,Y)$ denotes a numerical measure of dependence or association between Z and Y.

We calculated the correlation coefficent between the Hamilton score and Functional score.
We also calculated the correlation coefficient between SOFAS score and Functional score.

TABLE I
CORRELATION COEFFICIENT

Between Hamilton score and functional score	$r = .56$
Between SOFAS score and functional score	$r = .78$

Usefulness of the above Correlation metric

1.Correlation helps in predicting one quantity from another.
2.Correlation might indicate the presence of a causal relationship.
3.Correlation is a statistical measure that describes the association between random variables.

We saw that the correlation coefficient between SOFAS score and functional score is higher than the correlation coefficient between Hamilton score and functional score. SOFAS score focuses exclusively on the individual's level of social and occupational functioning and is not directly influenced by the overall severity of the individual's psychological symptoms. The Hamilton(HDRS)[1] scale was originally developed for hospital inpatients, thus the emphasis is more on melancholic and physical symptoms of depression as opposed to age,gender,education and other social attributes.

Hence, we believe, our functional score based on social and brain FMRI data establishes a perfect bridge between SOFAS score and Hamilton Score as it takes into account the social attributes of the patient as well as the overall psychological

symptoms based on brain scan data, hence this is more representative of patient's overall wellbeing.

Mixture Model

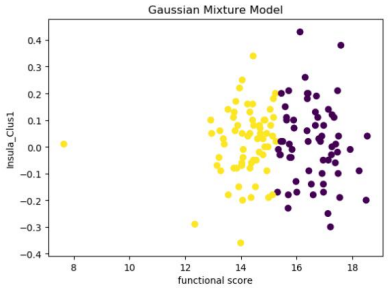
In order to understand the meaning of the correlation coefficient with respect to the structure of each of the brain scan data, we deep dive further using Mixture Models.

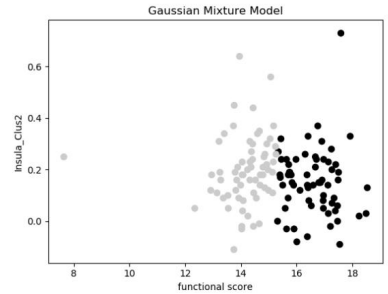
Assumption

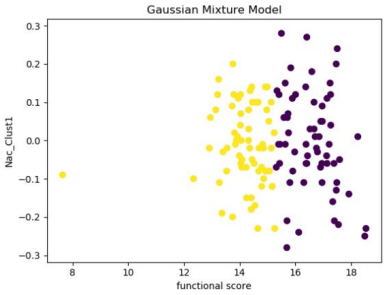
A distribution f is a mixture of K component distributions $f_1, f_2, ...f_K$ if
$f = \sum_{i=1}^{K} \lambda_k f_k$.
λ_k are the mixing weights, $\lambda_k > 0$, $\sum \lambda_k = 1$ Here we assume, $f_1, f_2, ...f_K$ follow Gaussian. In the above, f $\in$ a complete stochastic model, first we pick a distribution, with probabilities given by the mixing weights, and then generate one observation according to that distribution.
Symbolically,

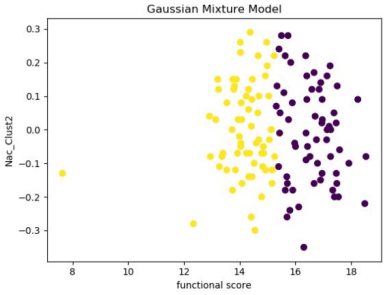
$$Z \sim Mult(\lambda_1, \lambda_2, ..., \lambda_K)$$
$$X|Z \sim f_Z$$

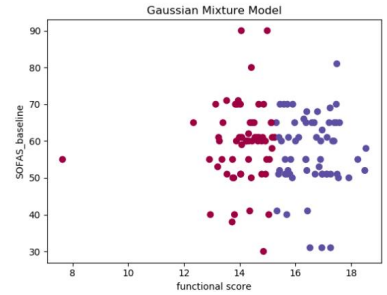
We ran different Gaussian Mixture models using our functional score and brain data and it reveals that the feature dataset indeed follow Gaussian and we can separate them clearly using Gaussian Mixture Model.

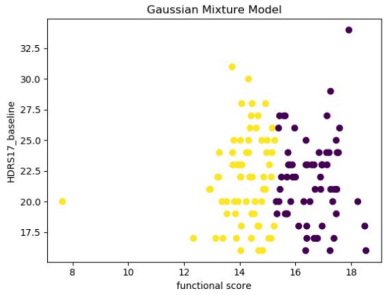












Other Statistics

We calculated few other statistics for our dataset.

Clustering coefficient of node i:

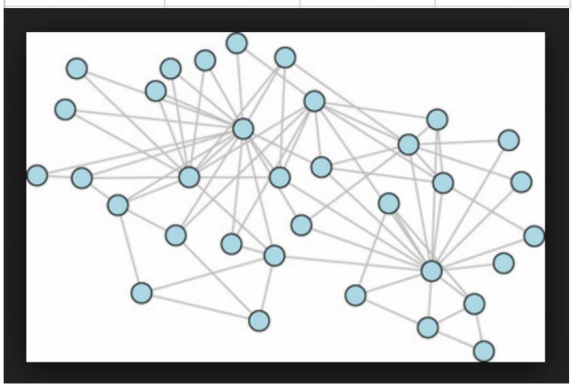
$$C_i = \frac{2 * r_i}{k_i(k_i - 1)}$$

r_i is the number of triangles around a node i and k_i is the degree of node i.
We did hierarchical clustering of 128 nodes and

we found the total number of unique clusters is 2 using total number of iterations 113 and mincut. We used Jaccard similarity value for clustering.

Clique-set: A clique is a subgraph containing vertices that connect to each other. If a graph contains edges that represent functional connectivity, then cliques from this graph would represent patients that behave similar with respect to the social attributes. We were looking for the set of 3-vertex or higher cliques to assess functionally similar networks for our dataset.
Here are the values of the metrics from the patient graph:

Clustering Coeff	Betweenness	PageRank	Eigenvector	Authority
0.85125	0.044	0.00237	0.2167	0.2167



V. *Results and Findings*

Based on the above analysis on the dataset, we got a comprehensive understanding of the characteristics of the patient nodes. The nodes capture the social bio markers behind depression symptoms. This functional score signifies a social score for each patient with respect to the three antidepressants. Strong correlation coefficient validates the association between functional score and the HDRS17 baseline (Hamilton score). Also, correlation coefficient validates strong association between SOFAS baseline and the functional score. HDRS17 baseline or Hamilton score and the SOFAS baseline scores are subjective in nature. These scores are determined by the healthcare professional's assessment of the patient. Whereas the functional score is computed by taking into account patient's non subjective elements like FMRI brain scan data, age, education and medication. Strong correlation between subjective scores like SOFAS baseline/HDRS17 baseline and functional score indicate the assessment of the healthcare professionals is accurate. Functional scores can predict the medication requirement of the patient.

Our dataset is very small as it is based on patient FMRI data,hence we applied the specific techniques that will provide results with moderately high accuracy.

The high clustering coefficient of 0.85 for the patient network suggests that if two patients clinical biomarkers are similar and they are taking the same antidepressant and if a third patient's biomarker matches with these two,then we can draw same conclusion with high probability that the third patient will benefit from the same antidepressant.

We did not make any i.i.d assumptions for any of our model as we expected high correlation between the social attributes and our assumptions are validated by the strong correlation coefficient found above.

We used Node2vec framework for learning vertex embeddings. This means learning a mapping of vertices to euclidean space that maximizes the likelihood of preserving network neighbourhoods of vertices. In node2vec, while sampling neighborhoods of a source patient node, we used Breadth-first Sampling (BFS) where the neighborhood was restricted to nodes that are immediate neighbors of the source patient node. Hence, we used the homophily hypothesis to search for nodes that are highly interconnected and belong to similar network clusters or communities and the embedding vectors provided those closely connected nodes.

VI. *Future Enhancements*

In our algorithm, the proximity of two nodes is modeled as the inner product of the embedding of feature vectors. However, it is known that simply the inner product of embedding vectors can limit the models representation ability and incur large ranking loss[5]. To capture the complex non-linearities of real-world networks, we would like to model the pairwise proximity of nodes by adopting a deep neural network architecture.

In future, we would like to enhance the Embedding layer as follows: it will consist of two fully connected components where one component is the one-hot patient ID vector that captures structural information of the graph network and the other component encodes the generic feature vector. The

embedding layer will be fed into multilayer perceptron which is neural network's hidden layer and the output vector of the last hidden layer will be transformed into probability vector which we will use to generate functional score for each patient node.

In our project, we just used BFS sampling in node2vec, we like to incorporate DFS sampling strategy where the neighborhood will contain nodes sequentially sampled at increasing distances from the source patient node. Hence, we will use structural equivalence hypothesis to embed nodes that have similar structural roles in networks. Unlike homophily, structural equivalence does not emphasize connectivity; nodes could be far apart in the network and still have the same structural role and this will be representative of a robust patient network and real networks commonly exhibit both behaviors where some nodes exhibit homophily while others reflect structural equivalence.

VII. *Github link*

The following github repo contains a link of the code and a copy of iSPOT-D dataset obtained from Dr.Adina Fischer,MD,PhD, a resident Stanford Psychiatry physician and a T32-funded postdoctoral fellow under the mentorship of Professor Leanne Williams and Professor Alan Schatzberg, Williams PanLab, Precision Psychiatry and Translational Neuroscience.

https://github.com/suvasis/cs224wproject

REFERENCES

[1] https://dcf.psychiatry.ufl.edu/files/2011/05/HAMILTON-DEPRESSION.pdf

[2] Community detection in graphs - Santo Fortunato Complex Networks and Systems LagrangeLaboratory, ISI Foundation, Viale S. Sever https://journals.aps.org/pre/abstract/10.1103/PhysRevE.80.056117

[3] Network enhancement as a general method to denoise weighted biological networks by Bo Wang,Armin Pourshafeie, Marinka Zitnik, Junjie Zhu, Carlos D. Bustamante,Serafim Batzoglou, JureLeskovec https://www.nature.com/articles/s41467-018-05469-x

[4] Spectral Clustering: Analysis and an algorithm bu Andrew Ng, Michael Jordan, Yair Weiss http://snap.stanford.edu/class/cs224w-readings/ng01spectralcluster.pdf

[5] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, Neural collaborative filtering, in WWW, 2017.

[6] BIRDS OF A FEATHER: Homophily http://aris.ss.uci.edu/~lin/52.pdf

[7] SOFAS https://kenniscentrum-kjp.nl/wp-content/uploads/2018/04/Social-Occupational-Functioning-Assessment-Scale-SOFAS.pdf

[8] amygdala cluster https://www.ncbi.nlm.nih.gov/pubmed/27669407

[9] Insula https://www.neuroscientificallychallenged.com/blog/2013/05/what-is-insula

[10] nucleus accumbens https://www.neuroscientificallychallenged.com/blog/2014/6/11/know-your-brain-nucleus-accumbens

[11] Locally Linear Embedding https://cs.nyu.edu/~roweis/lle/papers/lleintro.pdf

[12] Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering https://papers.nips.cc/paper/1961-laplacian-eigenmaps-and-spectral-\\techniques-for-embedding-and-clustering

[13] node2vec A. Grover, J. Leskovec. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016 https://arxiv.org/abs/1607.00653

[14] DeepWalk: Online Learning of Social Representations Bryan Perozzi, Rami Al-Rfou, Steven Skiena https://arxiv.org/abs/1403.6652

[15] Word2Vec Tutorial - The Skip-Gram Model// http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

[16] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, Neural collaborative filtering, in WWW, 2017. https://arxiv.org/pdf/1705.04969.pdf/

[17] Network enhancement as a general method to denoise weighted biological networks by Bo Wang,Armin Pourshafeie, Marinka Zitnik, Junjie Zhu, Carlos D. Bustamante,Serafim Batzoglou, JureLeskovec https://www.nature.com/articles/s41467-018-05469-x