TOPIC 6
DECIPHERING THE GENETIC FOUNDATIONS OF DISEASES USING NETWORK ANALYSIS

MSC BIG DATA SCIENCE - PG_7
ECS757P - DIGITAL MEDIA AND SOCIAL NETWORKS - 2023/24

ISHWAR JOSHI(230194814)
PRIYA KASHYAP(230194261)
SANCHIT JAIN(230677328)
UTKARSH KUMAR SHARMA(230671946)

INTRODUCTION

- Genetic information forms a complex network where genes and diseases are interconnected.
- Understanding the genetic roots of diseases is crucial to identifying critical genes.
- Analysis of these networks reveals how specific genes contribute to multiple diseases. [1]
- Insights lead to the development of targeted therapies tailored to individual patients.

RESEARCH PROBLEM

Analyzing the “diseasome" dataset, a bipartite graph mapping genetic disorders and disease genes, to understand disease comorbidity and identify potential therapeutic targets through shared genetic foundations.

MOTIVATION

The project aims to study a network known as” diseasomes”, and analyze how genes influence multiple conditions and potential preventions.

This will reveal patterns that offer valuable insights into disease susceptibility, progression, and treatment.

Studying these networks, we can uncover patterns linking genes and diseases, informing treatments and interventions to advance medical research and healthcare.

CHALLENGES

Understanding the complexity of the “diseasome" dataset, and deriving the meaningful insights from the bipartite graph requires a strong understanding of network analysis and domain knowledge.

Addressing the inherent noise and ambiguity in the dataset to ensure accurate analysis and interpretation of genetic associations.

Developing effective algorithms and methodologies for navigating and analyzing the intricate network structure of the “diseasome" dataset.

Reference: [1]

DATASET AND NETWORK PRESENTATION

- Using two datasets (reduced and full) for the analysis which are collected from the bio-diseasome website.
- The data contained is called “diseasome” which contains genes (Chromosomes) and associated diseases.
- Nodes represent genes and disease, and the edges represent their association.
- The Genetic disorder is identified by analyzing the bipartite network of the diseasome dataset.
- To identify the disorder a comprehensive analysis is required along with parameter tuning to reduce the network to the most meaningful state.

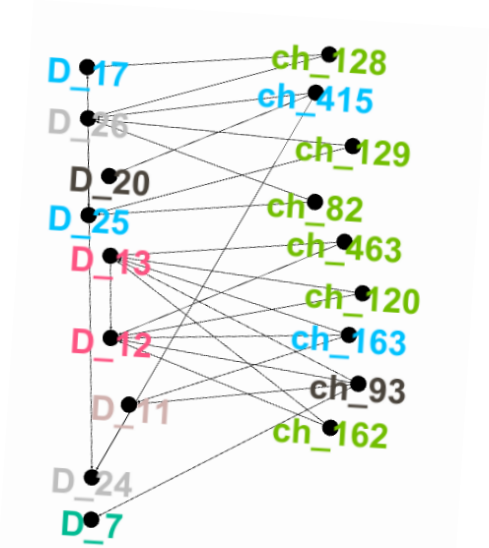
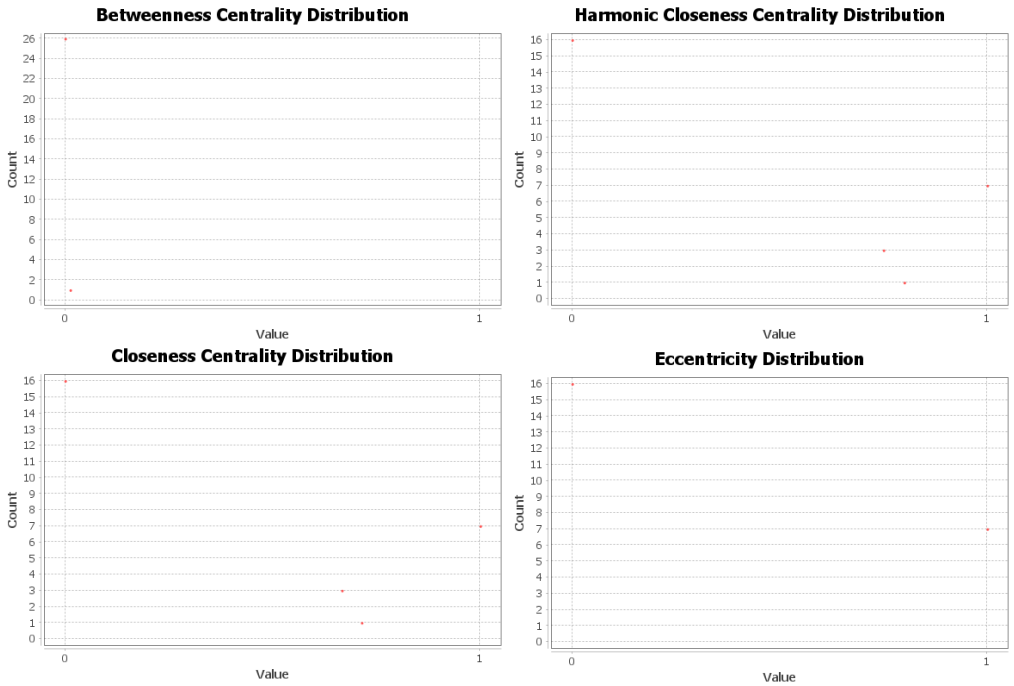


Fig: Bipartite Network of diseasome dataset

Source: Gephi (By PG_7)

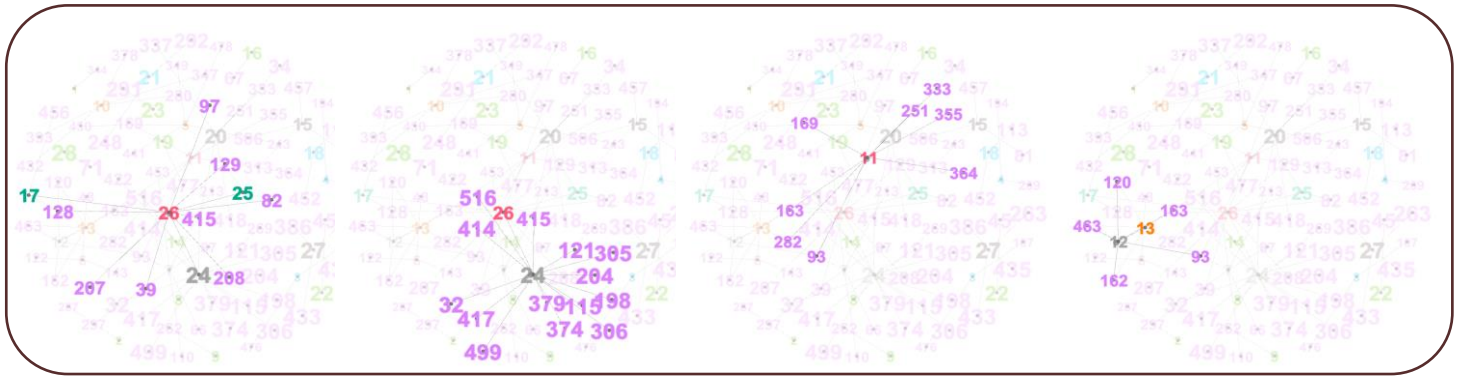
Statistics	Values
Average Clustering Coefficient	0.192
Average Degree	0.963
Modularity	0.500
Modularity with resolution	0.500
Number of Communities	12
Network Interpretation	directed
Number of iterations	100
Sum change	0.003095184399684897

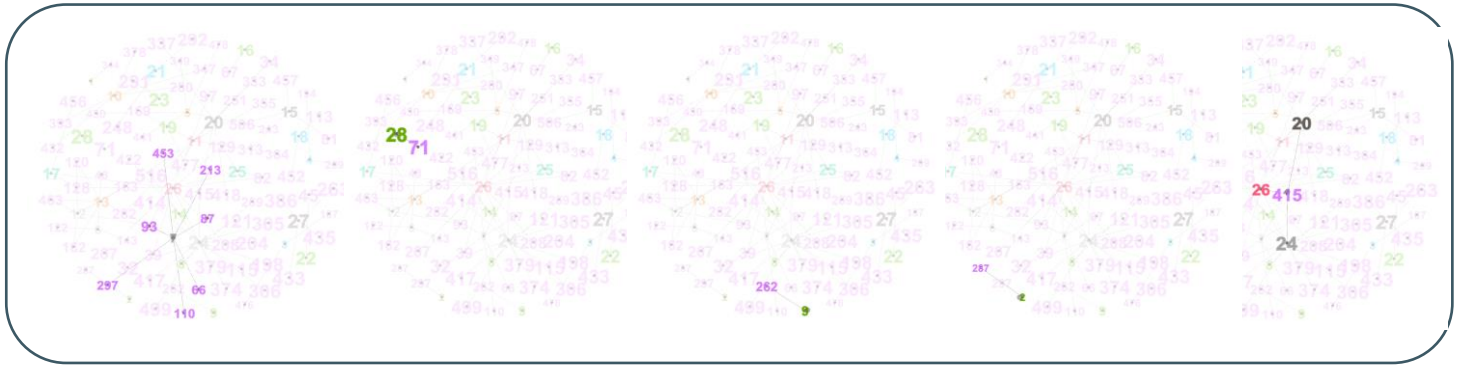


BASIC NETWORK STATISTICS

- Visualization of the bipartite network helps in calculation of required basic network statistics.
- Degree distribution, Clustering Coefficient, Modularity, Centrality measure(Eigen Vector).
- The bipartite network is defined by Jean-Loup Guillaume as, "A bipartite network is a triple G = (T,⊥,E) where T and ⊥ are two disjoint sets of nodes, respectively the top and bottom nodes, and E ⊆ T×⊥ is the set of links of the network."[2]

Key findings: -





	High	Low
Eigen Vector* Centrality	D26	D28
Clustering Coefficient	D24	D9,2

D12 and 13 have equal centrality

Reference: [2]

NETWORK ANALYSIS

DECIPHERING THE GENETIC FOUNDATIONS OF DISEASES

OBJECTIVE

- Analyze the modular structure of the diseasome network through community detection algorithms.
- Identify clusters of diseases, revealing distinct communities within the network.
- Analyze community characteristics such as size, density and connectivity between communities.
- Biological implications of identified communities, and investigating how genes within the same community may contribute to similar diseases or share genetic causes.

METHODOLOGY

- Construct a network using the entire diseasome dataset and then partition it into communities.
- Analyze the community sizes, density, and bridge genes or diseases between different communities.
- Examine the functional significance of genes within communities, and discuss potential implications for disease diagnosis, treatment, and drug development based on identified genetic associations.
- Finding a stable network using gene knockout simulations identifying critical genes and evaluating the network's robustness to perturbations.

OUTCOME

- Constructing the network and analyzing the distribution will give us details regarding the degree distribution, clustering coefficient, and modularity.
- Performing gene knockout simulations will help us identify critical genes, and create a stable network.
- Network motifs will help us identify biologically significant patterns and using diffusion models will help us to simulate the propagation of effects through the network.

REFERENCE

The human disease network
https://www.pnas.org/doi/epdf/10.1073/pnas.0701361104

Bipartite structure of all complex networks
https://www.sciencedirect.com/science/article/abs/pii/S0020019004000754