# **Deciphering the Genetic Foundations of Diseases using Network Analysis - 2024**

**Context and project description**

In the age of personalized medicine, understanding the genetic underpinnings of diseases is paramount. The complex interplay between genes and diseases forms a vast network, where nodes represent genetic entities and diseases, and edges denote associations between them. This network encapsulates critical information about the susceptibility, progression, and treatment of diseases, making its analysis fundamental for advancements in medical research and healthcare.

The study of these networks, often referred to as "***diseasomes***," sheds light on how specific genes contribute to multiple diseases, revealing patterns that are crucial for the development of targeted therapies. The challenge lies in the complexity and scale of the data, which encompass thousands of genes and disease associations. This complexity requires sophisticated network analysis techniques to decipher the intricate web of interactions and to identify key genetic factors that could lead to breakthroughs in understanding and treating diseases (Barabási et al., 2011; Goh et al., 2007).

**Network dataset**

For this project, the focus will be on analysing the "Diseasome" dataset obtained from the Network Repository (see link and description below). This pivotal dataset forms a comprehensive bipartite graph mapping the relationships between genetic disorders and disease genes. It encompasses nodes representing both diseases and genes, linked by edges that denote known genetic associations. With its extensive coverage, the "Diseasome" dataset includes a wide array of diseases and genetic links, making it an invaluable resource for exploring the genetic underpinnings of disease comorbidity and identifying potential therapeutic targets. The dataset's structure facilitates a network-based exploration of how diseases are interconnected through shared genetic foundations, providing a unique lens through which to examine human health and disease.

**Tasks**

**Task 1**  **(Network Analysis and Visualisation)**

* **a)** Construct the bipartite network for the **reduced version of the Diseasome dataset** provided on the topic page (file “bio-diseasome\_reduced.mtx.txt”).
* **b)** Compute four key network statistics: degree distribution, clustering coefficient, modularity, and one centrality measure (either Degree, Betweenness, Closeness, or Eigenvector).
* **c)** Visualize the network structure using Gephi or Python's matplotlib library for a graphical representation of the diseasome network, highlighting key nodes and communities.
* **d)** Interpret the findings from the computed statistics and visualisation and discuss their significance in understanding the genetic and disease associations within the diseasome network. This includes identifying genes that play a central role in multiple diseases and discussing the implications for understanding disease mechanisms and potential therapeutic targets.

**Task 2 - Community Detection and Analysis**

* **a)** Construct the entire network **using the full Diseasome dataset** and employ community detection algorithms to partition the network into communities, focusing on revealing clusters of diseases and genes that exhibit higher internal connectivity. This task will help you to understand the modular structure of the Diseasome network.
* **b)** Analyse the characteristics of these communities, including the size of each community, the density of connections within communities, and the genes or diseases that serve as bridges between communities. This step will help you to understand the modular structure of the Diseasome network.
* **c)** Discuss the potential biological implications of the identified communities, focusing on how genes within the same community may contribute to similar diseases or how diseases within a community may share genetic causes.

**Task 3 - Network Dynamics and Robustness Analysis**

* a) Perform gene knockout simulations to study their impact on network stability and disease associations
* **b)** Identify critical genes or diseases using centrality measures, assessing their importance in maintaining network integrity.
* **c)** Evaluate the network's robustness to perturbations, comparing its resilience to that of random and scale-free networks.

**Task 4 - Advanced Network Analysis and Hypothesis Generation (PG only)**

* **a)** Apply advanced metrics like PageRank and explore network motifs to uncover biologically significant patterns.
* **b)** Utilize diffusion models or epidemic spreading models to simulate the propagation of effects through the network.
* **c)** Generate hypotheses on gene-disease relationships based on advanced analysis, proposing new therapeutic or diagnostic approaches.
* **d)** Employ advanced visualization techniques to illustrate identified patterns, interpreting their implications for genetic influences on diseases.

**References**

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nature Reviews Genetics.

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. Proceedings of the National Academy of Sciences.

Ideker, T., & Krogan, N. J. (2012). Differential network biology. Molecular Systems Biology.

**Dataset**

The "Diseasome" dataset can be downloaded from the Network Repository using the following steps:

* Visit the webpage:<https://networkrepository.com/bio-diseasome.php>
* Click on the "Download" button or link provided on the page to obtain the dataset.
* The dataset is typically compressed; extract the files using a file archiver program.

Even though the "Diseasome" dataset is anonymised and publicly available, it's crucial to handle genetic data ethically, respecting privacy due to its sensitive nature.

**Resources**

**Gephi Tutorials**: Gephi provides a range of tutorials and documentation to help students get started with network visualization and analysis. These tutorials cover topics like data import, layout, and visualization customisation.

**Python NetworkX Library**: NetworkX is a Python library for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. It provides extensive documentation and examples.