Part 1: Collect and Describe Data

**Paper**: Identification of critical connectors in the directed reaction-centric graphs of microbial metabolic networks

**Authors**: Eun-Youn Kim, Daniel Ashlock & Sung Ho Yoon

**Group**: Prathamesh Chati, Beniamin Krupkin, Minna Apostolova

**Data Collection**

The data used are metabolic graphs from five microbial species. Each graph is represented as a directed adjacency matrix where 1 indicates and edge and 0 indicates not edge between pairs of genes. The five graph adjacency matrices were downloaded from the “Additional Files” section of the paper along with the supplementary tables. The first supplementary table includes scale-free properties for each of the five graphs provided.

**Data Description**

**Table S1. Modularity and scale-freeness of the reaction-centric metabolic networks**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Strain** | **Modularity (*P*-value)** | **Degree distributions** (**Pr(*k*) ≈ *k*-γ )** | | |
| **Total degree (γ*tota*l)** | **In- degree**  **(γ in** **)** | **Out-degree**  **(γ*out*)** |
| ***E. coli* (iJO1366)** | 0.6103 (0) | -1.29 | -1.32 | -1.50 |
| ***B. subtilis* (iYO844)** | 0.5622 (0) | -1.03 | -1.10 | -1.12 |
| ***G. metallireducens* (iAF987)** | 0.5179 (0) | -1.03 | -1.14 | -1.11 |
| ***K. pneumoniae* (iYL1228)** | 0.6266 (0) | -1.21 | -1.31 | -1.47 |
| ***S. cerevisiae* (iMM904)** | 0.5705 (0) | -0.99 | -1.08 | -1.16 |

The supplementary table above was taken directly from the manuscript and describes the parameters for the scale-free network for each of the five strains queried in the study. Since this is a directed graph, the in- and out-degree can be computed for each network.

**Graph Properties Table**

The following table was generated using the adjacency matrices provided in the “Additional Files” section. It describes the above supplementary table in more detail, specifically looking at the directed and undirected characteristics of the network. These features were calculated using the NetworkX package in Python. Specifically, it extends Table S1 by providing count level information on the nodes and edges.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Strain** | **Total Nodes** | **Degree distributions** | | |
| **Total degree** | **In- degree** | **Out-degree** |
| ***E. coli* (iJO1366)** | 1251 | 9099 | 3866 | 5233 |
| ***B. subtilis* (iYO844)** | 748 | 6489 | 2964 | 3525 |
| ***G. metallireducens* (iAF987)** | 900 | 8049 | 4567 | 3482 |
| ***K. pneumoniae* (iYL1228)** | 1137 | 8084 | 3750 | 4344 |
| ***S. cerevisiae* (iMM904)** | 881 | 10460 | 5812 | 4648 |

The adjacency matrix is binary and contains 1 for an edge between two nodes and a 0 for no edge. It is directed, so the matrix is square but not symmetrical. There are five total variables for each strain and the graph properties are described in the table. There is no missing data for the graphs. The original graphs were sourced from separate sources. Here, the nodes represent genes within each graph. The in-degree is considered as the top half of the square adjacency matrix while the out-degree is considered as the bottom half (half is classified as along the diagonal of self-edges).

The nodes and edges are summarized visually in the figures below. The first figure shows the degree distribution while the second figure shows number of nodes relative to edges.

A screenshot of a computer

Description automatically generated

A group of orange and blue bars

Description automatically generated