A3:

Date of announcement: 5:00 pm IST, October 09, 2021 (Saturday)
Date of submission: 11:59 pm IST, October 25, 2021 (Monday)
Summary: The assignment is on browser-based visualizations of networks, trees, and multivariate data.

Libraries to be used: Plotly or D3.js
Diseases related datasets:
* Network data - diseasome: https://networkrepository.com/bio-diseasome.php
* Network data - : https://networkrepository.com/infect-dublin.php
* Multivariate data: https://catalog.data.gov/dataset/ah-sickle-cell-disease-provisional-death-counts-2019-2020-3e71c
* Multivariate data: https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi

Visualizations to generate:

1. Node-link diagrams with two different layouts, namely, force-directed, circular, including node-labels
2. Matrix visualization of the node-link diagram generated in #1, including node-labels
3. Treemap
4. Parallel coordinates plot with axis swapping and brushing.

Tasks:

1. Pick at least one network and one multivariate dataset from the aforementioned list.
2. Decide how you would like to use the dataset for visualization -- directly or after remodeling.
3. Prepare datasets for generating visualizations in the aforementioned list, such that the prepared datasets include at least a network, a tree, and a table.
4. Implement the visualizations given in the aforementioned list.

Notes:

- The network dataset webpages in the networkrepository contain node-link diagram visualizations. These could be used as references for the visualizations you generate. Fig.3 is the reference visualization of bio-diseasome and Fig.4 is that of infect-dublin.

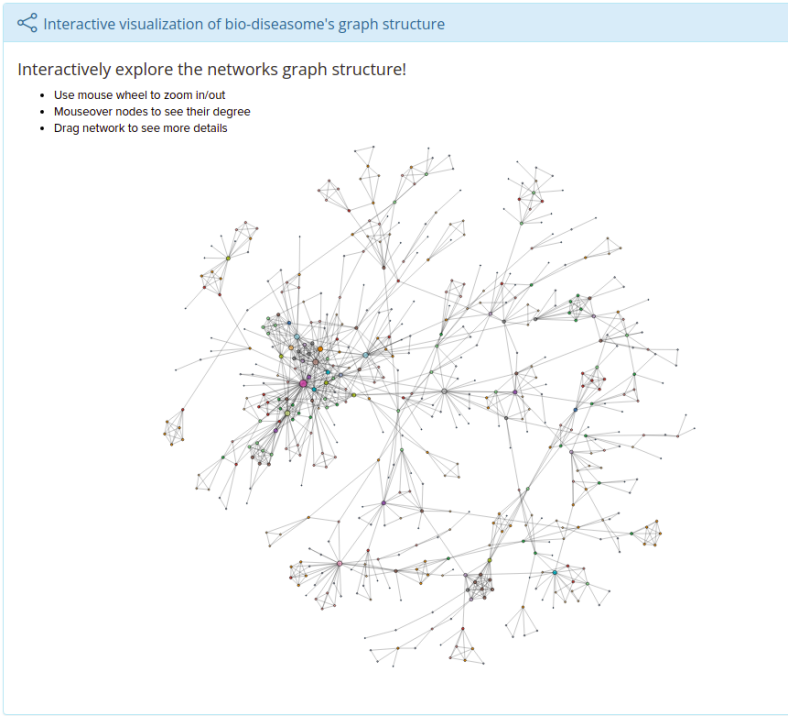

Figure 3: Reference node-link diagram of bio-diseasome (Source: NetworkRepository)

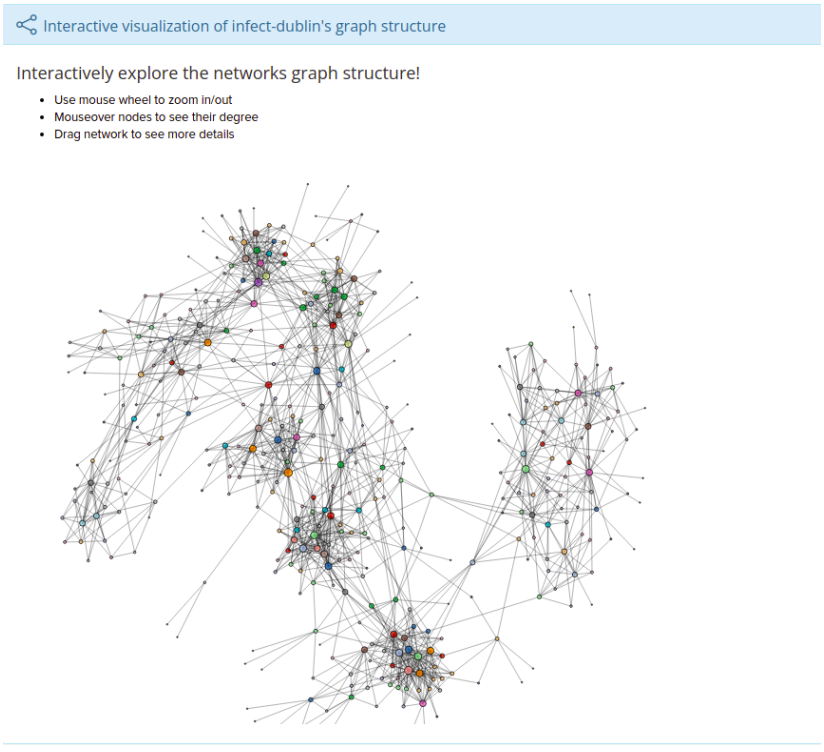

Figure 4: Reference node-link diagram of infect-dublin (Source: NetworkRepository)

- The network datasets are labeled using numerical indices. Hence, these indices may be used as-is for the visualization tasks.
- The tabular data for chronic diseases contain several irrelevant columns. Pick columns based on the semantics of the variable, and known data type (categorical or numerical), as not all the variables are needed to be plotted. The same applies to the tabular data on sickle cell anemia, too.
- For remodeling, other data mining methods such as distance matrix, similarity matrix, ordering for building hierarchy, etc. may be used.
- These datasets are considered small datasets. Hence, use all rows in the multivariate datasets, and all nodes and edges in the network datasets for visualizations, unless any of this is spurious data. This implies that subsampling these datasets for visualizations will not be entertained.
- The visualizations must be rendered on a browser, hence, appropriate libraries must be used to generate HTML pages for these visualizations.
