732A98 Visualization
Division of Statistics and Machine Learning
Department of Computer and Information Science

Laboratory work 6

Instructions

- Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
- **Include all your codes as an appendix into your report; you are also recommended to show parts of the codes in the flowing text of the report.**
- A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes.
- Create a report to the lab solutions in RMarkdown. Make sure that it is can be compiled to HTML and that all paths in RMD file are relative to the current directory where the RMD file is located. **Reports that can not be compiled are returned without revision.**
- Put the RMD file and all supporting files into one ZIP archive when you submit it to LISAM.
- The lab report should be submitted via LISAM before the deadline.

Assignment 1. Network visualization of terrorist connections

Files **trainData.dat** and **trainMeta.dat** contain information about a network of the individuals involved in the bombing of commuter trains in Madrid on March 11, 2004. The names included were of those people suspected of having participated and their relatives.

File **trainMeta.dat** contains the names of individuals (first column) and Bombing group (second column) which shows "1" if person participated in placing the explosives and "0" otherwise. According to the order in this file, persons were enumerated 1-70.

File **trainData.dat** contains information about connections between the individuals (first two columns) and strength of ties linking (from one to four):

1. Trust--friendship (contact, kinship, links in the telephone center).
2. Ties to Al Qaeda and to Osama Bin Laden.
3. Co-participation in training camps and/or wars.
4. Co-participation in previous terrorist Attacks (Sept 11, Casablanca).

1. Use visNetwork package to plot the graph in which
    a. you use strength of links variable
    b. nodes are colored by Bombing Group.
    c. size of nodes is proportional to the number of connections (function strength() from IGRAPH might be useful here)
    d. you use a layout that optimizes repulsion forces (visPhysics(solver="repulsion")).
    e. all nodes that are connected to a currently selected node by a path of length one are highlighted

    Analyse the obtained network, in particular describe which clusters you see in the network.

2. Add a functionality to the plot in step 1 that highlights all nodes that are connected to the selected node by a path of length one or two. Check some amount of the largest nodes and comment which individual has the best opportunity to spread the information in the network. Read some information about this person in Google and present your findings.
3. Compute clusters by optimizing edge betweenness and visualize the resulting network. Comment whether the clusters you identified manually in step 1 were also discovered by this clustering method.
4. Use adjacency matrix representation to perform a permutation by Hierarchical Clustering (HC) seriation method and visualize the graph as a heatmap. Find the most pronounced cluster and comment whether this cluster was discovered in steps 1 or 3.

Assignment 2. Animations of time series data.

The data file ***Oilcoal.csv*** provides time series about the consumption of oil (million tonnes) and coal (million tonnes oil equivalents) in China, India, Japan, US, Brazil, UK, Germany and France. Marker size shows how large a country is (1 for China and the US, 0.5 for all other countries).

1. Visualize data in Plotly as an animated bubble chart of Coal versus Oil in which the bubble size corresponds to the country size. List several noteworthy features of the investigated animation.
2. Find two countries that had similar motion patterns and create a motion chart including these countries only. Try to find historical facts that could explain some of the sudden changes in the animation behavior.
3. Compute a new column that shows the proportion of fuel consumption related to oil: $Oil_p = \frac{Oil}{Oil+Coal} * 100$. One could think of visualizing the proportions Oil_p by means of animated bar charts; however smooth transitions between bars are not yet implemented in Plotly. Thus, use the following workaround:
    a. Create a new data frame that for each year and country contains two rows: one that shows Oil_p and another row containing 0 in Oil_p column
    b. Make an animated line plot of Oil_p versus Country where you group lines by Country and make them thicker

    Perform an analysis of this animation. What are the advantages of visualizing data in this way compared to the animated bubble chart? What are the disadvantages?
4. Repeat the previous step but use "elastic" transition (easing). Which advantages and disadvantages can you see with this animation? Use information in https://easings.net/ to support your arguments.
5. Use Plotly to create a guided 2D-tour visualizing Coal consumption in which the index function is given by Central Mass index and in which observations are years and variables are different countries. Find a projection with the most compact and well-separated clusters. Do clusters correspond to different Year ranges? Which variable has the largest contribution to this projection? How can this be interpreted? (Hint: make a time series plot for the Coal consumption of this country)

732A98 Visualization
Division of Statistics and Machine Learning
Department of Computer and Information Science

Submission procedure

Assume that X is the current lab number, Y is your group number.

If you are neither speaker nor opponent for this lab,

- Make sure that you or your group comrade submits the group report using *Lab X* item in the *Submissions* folder before the deadline

If you are a speaker for this lab,

- Make sure that you or your group comrade does the following before the deadline:
  - submits the group report using *Lab X* item in the *Submissions* folder before the deadline
  - Goes to Study room *Group Y→Documents* and opens file *Password X.txt*. Then the student should put your group report into ZIP file *Lab X_Group Y.zip* and protect it with a password you found in Password X.txt
  - Uploads the file to *Collaborative workspace →Lab X* folder

If you are opponent for this lab,

- Make sure that you or your group comrade submits the group report using *Lab X* item in the *Submissions* folder before the deadline
- After the deadline for the lab has passed, go to Collaborative workspace→*Lab X* folder and download the appropriate ZIP file. Open the RMD file in this ZIP file by using the password available in *Course Documents →Password X.txt,* compile it, read it carefully and prepare (in cooperation with your group comrade) **at least three questions/comments/improvement suggestions per lab assignment** in order to put them at the seminar.