Homework 2. Social network analysis

Lecturer: Semen Sorokin

September - December 2020

In this assignment you will work with real data of social networks. Assuming that you know basic instruments for social network analysis and you can perform necessary data wrangling to define patterns of human behavior or social interaction inside a small community. All operations can be done using python (networkx/igraph) or/and gephi.

1 Create data - 0.5 point

You can use your own vk-profile and this application (https://vk.com/app3861133) to unload your friendship network. There are two options:

- Only your buddies (depth = 1)
- Your buddies and their buddies (depth = 2) (be careful - someone can have more than thousand links and it will take plenty of time to unload)

Include your node in graph or not - your decision, but do not forget about this during analysis. You can use other way to get data, but please describe it.

2 Clear data - 0.5 point

Check if data for some nodes is missing, clear (deleted users) or fill it in. **Attention**: if more than 50 slots are empty, do not fill it manually, try to find automatic ways to do that or define research limitations.

3 Basic statistics - 1 point

Number of nodes, edges, components. Density. If you would like to use only one component of social network, explain why and also describe reasons of many component appearance. Decide if graph should be oriented or not.

4 Descriptive analysis - 1 point

- Degree distribution (+ log-log scale)
- Power low distribution fitting

Explain result.

5 Centrality - 1 point

Use degree, betweenness, closeness, eigenvector centrality. Visualize node with different node color and node size. Compare these patterns with common sense. Which metrics better describe real situation?

6 Communities - 1 point

Try different algorithms for communities unfolding. Visualize and explain result.

7 Communities comparing - 1 point

Select "best" communities and compare them between each other using sub-graph structure and basic statistics

8 Conclusion - 1.5 points

Describe found patterns and checked hypotheses from different point of view.

9 Additional points - 2.5 points

- Usage of different instruments and their combinations. **Example**: data clearing (pandas) ⇒ basics and visualization (gephi) ⇒ advanced communities unfolding (networkx)
- Advanced filtering: k-core decomposition for sub-graph selecting, external attribute information (gender segregation)
- Applying different configuration parameters for network creating and filtering, analysis of more than one social network graph, appeared from the same data source

As a result I expect you to create folder on github with a notebook (pdf) file containing source code (in python) and all necessary data (input data, gephi visualization results and etc.) Good luck.

Deadline: 19 December