4. Social Network Analysis

Task

- Use an existing Movie dataset [https://archive.ics.uci.edu/ml/datasets/Movie]
  - Focus on „casts" - data that provide information about actors in movies
  - You can also use already converted data to csv from https://github.com/cernoch/movies [https://github.com/cernoch/movies]
- Convert „casts" data to a graph
  - Create a node for each actor
  - Create an edge if two actors appeared in the same movie
- Perform Social Network Analysis
  - Compute general statistics
  - Identify key players using centrality measures
  - Identify clusters/communities in graph
  - Compute „Kevin Bacon" number for each actor with selected key player
  - Visualise important aspects of the analysis

Instructions for submitting

In your private namespace on EDUX provide the following information:

- Provide general statistics about the dataset
  - e.g. number of nodes and edges, density, number of components
- Provide list of top key players using different centralities
- Describe top clusters/communities
- Describe „Kevin Bacon" numbers
  - e.g. top actors with the highest/lowest number
  - average number
- Insert visualisations from networkx/Gephi (images/screenshots)
  - for visualisation you can use only subset of actors/movies e.g. movies from specific category, movies with max 5 actors etc.
- Insert a file with the graph exported to a GEXF format
  - All computed values should be present as attributes
- Provide the link to your implementation
  - You can use: https://gitlab.fit.cvut.cz [https://gitlab.fit.cvut.cz] or https://github.com/ [https://github.com/] or https://bitbucket.org [https://bitbucket.org]
- Comment on
  - issues during the design/implementation
  - ideas for extensions/improvements/future work

/mnt/www/courses/MI-DDW.16/data/pages/hw/04/start.txt · Poslední úprava: 2017/03/27 21:47 autor: kuchajar