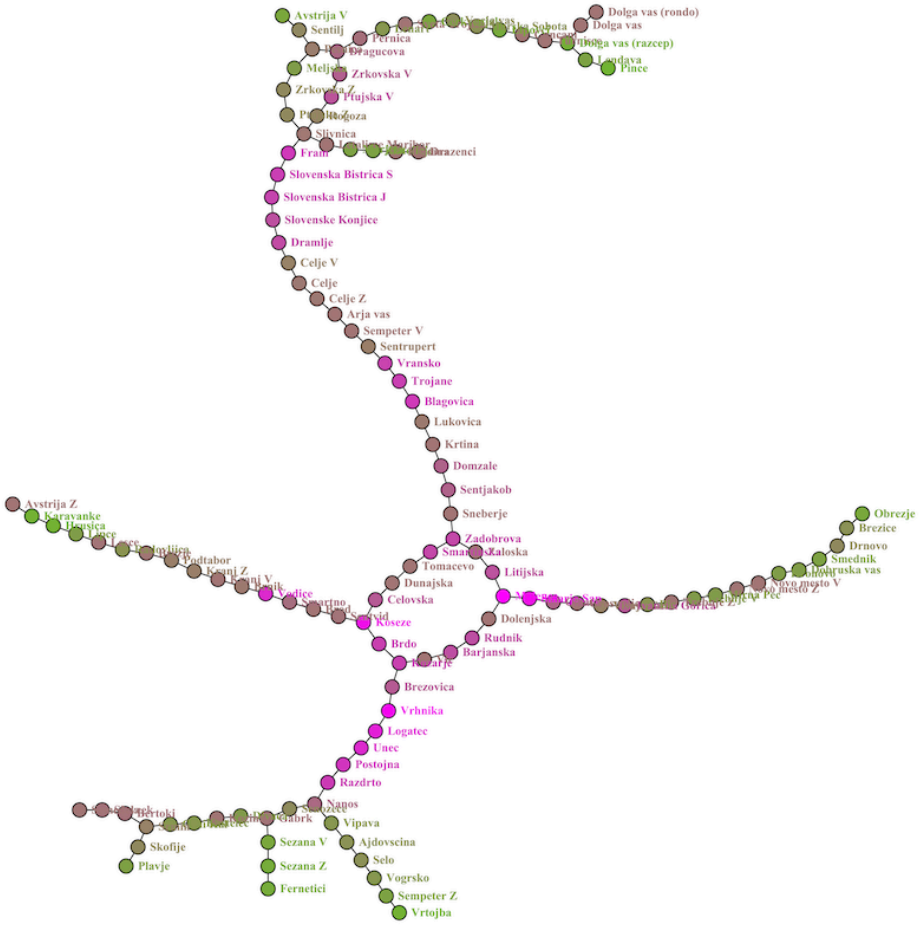
Link betweenness, node similarity, errors & attacks

I. Betweenness in transportation networks

You are given two transportation networks in Pajek format.

- Slovenian highways network from 2010 (highways.net)
- European highways network from Wikipedia (euroroads.net)



1. **(code)** Compute standard statistics of the networks. Are the results expected?

2. **(code)** Find the most important highways according to the link betweenness centrality $\sigma_{ij} = \sum_{st \notin \{i,j\}} \frac{g_{st}^{ij}}{g_{st}}$, where g_{st} is the number of geodesic paths between nodes s and t, and g_{st}^{ij} is the number of such paths through the link between nodes i and j. Which highways have the highest σ_{ij}?

3. **(homework)** For the Slovenian highways network, compute the Pearson and Spearman correlation coefficients between the highways betweenness centrality σ_{ij} and their actual traffic load. (Assume that the traffic load of a highway is the average of the traffic loads of its endpoints.) Is σ_{ij} positively correlated with traffic load?

II. Movie recommendations with PageRank

You are given a small knowledge graph of 1337 movies in Pajek format (movies_graph.net). Nodes represent either individual movies or their different *modes* such as language, country, genres, actors, director etc.

1. **(code)** Compute standard statistics of the network. Are the results expected?

2. **(code)** Find the most important movies according to the PageRank algorithm $p_i = \alpha \sum_j A_{ij} \frac{p_j}{k_j} + \frac{1-\alpha}{n}$, where A is the network adjacency matrix, n is the number of network nodes, k_i is the degree of node i and α is the damping factor set to 0.85. Which movies have the highest PageRank score?

3. **(code)** Consider random walks with restarts $p_i^t = \alpha \sum_j A_{ij} \frac{p_j^t}{k_j} + (1-\alpha)\delta_{it}$, where t is a selected teleport node and δ is the Kronecker delta. How could you use this algorithm to find movies similar to, e.g., *Moana*?

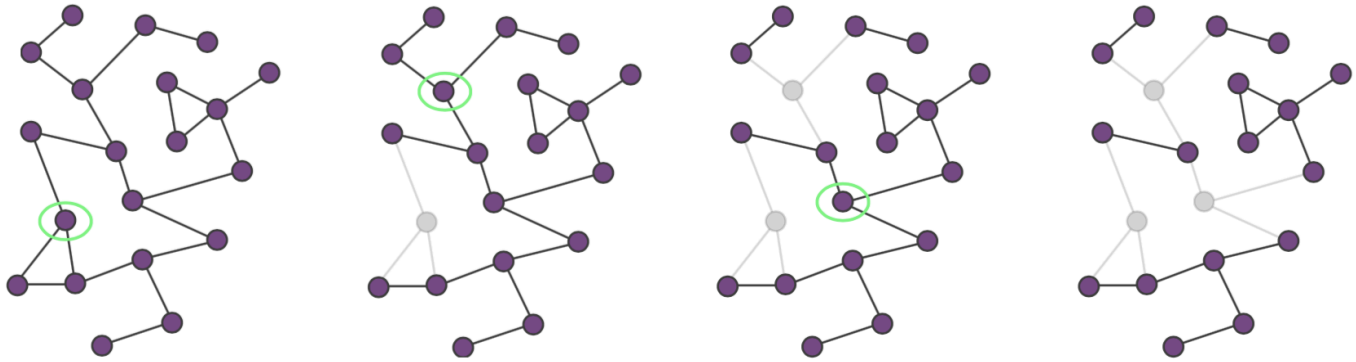
4. **(homework)** Consider the personalized PageRank algorithm $p_i^{[t]} = \alpha \sum_j A_{ij} \frac{p_j^{[t]}}{k_j} + (1-\alpha)[t]_i$, where $[t]$ is a selected personalization vector, $\sum_i [t]_i = 1$. How could you use this algorithm to find movies similar to, e.g., dramas starred by Tom Hanks, action and adventure movies with Johnny Depp or movies co-starred by Brad Pitt and George Clooney?

5. **(discuss)** Examples above include only *positive* queries by measuring similarity between the movies and selected modes. But how could handle also *negative* queries such as, e.g., you do not like romantic movies or a particular actor?

III. Errors and attacks on the Internet

You are given the *nec* Internet overlay map in Pajek format (nec.net). Routers, switches and hubs on the Internet fail from time to time. Your task is to study how this affects the Internet's ability to stay *connected*.

1. **(discuss)** Simulate such failure by *random* removal of nodes and measure the fraction of nodes in the largest connected component. Next, consider also a malicious individual attacking the Internet. Simulate such *attack* by removing highly linked nodes (i.e. hubs) and again measure the fraction of nodes in the largest connected component.



2. **(code)** For each of the two scenarios, plot the fraction of nodes in the largest connected component after removing 0%–50% of the nodes. Is the Internet robust to random failures? IS the Internet robust to targeted attacks?

3. **(homework)** Repeat the experiments also for Erdös-Rényi random graphs with the same number of nodes and links as the Internet overlay map.