Title: Exploring the correlation between a country's history/geography and its road network graph

Motivation:

- Intuitively, countries with different histories, geographies and development timelines should have differently structured roads
- The road graphs of european countries, with thousands of years of gradual development history, should have more in common with each other rather than the ones found, for example, in the US, given their relatively rapid development
- Do measurable graph properties actually reflect this intuition?
- Are real-life road networks distinguishable from random networks?
- Data: the following graphs available from public repositories (snap.stanford.edu and networkrepository.com)
  - road-italy-osm (~7M nodes, 7M edges)
  - road-germany-osm (12M nodes, 12M edges)
  - road-great-britain-osm (8M nodes, 8M edges)
  - road-belgium-osm (1M nodes, 2M edges)
  - road-luxembourg-osm (115k nodes, 115k edges)
  - road-netherlands-osm (2M nodes, 2M edges)
  - road-road-usa (24M nodes, 29M edges)
  - road-roadNet-CA (2M nodes, 3M edges)
  - road-roadNet-PA (1M nodes, 2M edges)
  - road-roadNet-TX (1M nodes, 2M edges)
  - road-asia-osm (12M nodes, 13M edges)

Method:

- Problem: Compute average centrality scores, motifs, clustering coefficient of road graphs and use them as features for clustering of said data. Confront features of real data with random graphs
- Algorithms: We are going to try using the algorithms seen in class that provide an exact solution. If that proves too slow we will employ an approximate algorithm

Intended experiments:

- We will use the python language together with the graph analysis library networkx (networkx.org) which implements many common graph operations
- Machine for experiments: We plan to interchangeably use the personal computers of the members of this group (a linux pc with 8gb RAM, a windows pc with 8gb RAM, a windows pc with 12 gb RAM. None posess a GPU). Should they prove insufficiently powerful for this task we will make use of Google Colab
- Experiments:
  - For each graph: compute average closeness centrality and betweenness centrality, clustering coefficient and number of motifs up to a certain number of nodes depending on computational time
  - Save each computed statistic as a feature in a multi-dimensional space describing each graph
  - Generate some random graphs based on the actual road networks and compute the same statistics for them
  - Employ a clustering algorithm to group together road networks that share similar characteristics
  - Verify if states with similar histories and geographies are grouped together in the same cluster and if they are distinguished from randomly generated networks