Part I

Proposal

1 The discourse of newspapers - A network analysis

The focus of this project is to analyse the discourse of a news outlet via a graph construction. The connections between individual topics and in particular how they change over time will be investigated.

1.1 Data

The New York Times is one of the most popular newspapers in the world and publishes thousands of articles every month. We use the meta data of these articles (specific keywords or tags) as a source of consistent data and subsequently construct the discourse network.

1.2 Network

The aim is to construct a simple undirected, weighted and homogeneous graph where a ..

- .. Vertex represents a given keyword.
- .. Link (and its weight) represent the likelihood that two keywords appear in the same article.

The precise implementation details for calculating link probabilities are not yet certain. One simple method would be to simply take the percentage in which two keywords were used in the same article as their respective weight, however better alternatives may be identified during later testing.

1.3 Analysis

Such a network enables a wide range of possible graph analysis methods which can be compared over multiple time steps to provide clearer insight into the collected data. Some points of interest that will be targeted include:

- The most discussed topics (Comparing metrics such as: Centrality, Betweenness, Closeness and PageRank)
- The relationships between different topics (Comparing results of different clustering methods)

The precise details regarding the use and implementations of above methods may shift slightly with the aim of extracting meaningful/interesting results.