Google Summer of Code 2009
Gephi proposal #2

Network algorithms and statistics

Difficulty: Easy
Required skills: Basic graph theory, Java

This proposal focuses on network analysis, through statistics and various algorithms. Though Gephi focus on powerful and easy-to-use exploration features, it is not enough to understand and classify all networks. Classical results like graph density, clustering coefficient or the number of weakly connected components help to recognize pattern and in general find features in network structure and data. Hence we propose to implement some of these algorithms in Gephi statistical module:

- Watts Strogatz Clustering Coefficient
- Pagerank
- HITS
- Network Diameter
- Node Betweenness Centrality
- Average Shortest Path

Implementing these algorithms will convey good overview on classical graph algorithms and some knowledge about fascinating results (like Watts-Strogatz Small Worlds model). Execution and results visualization tests will use native Gephi features.

A flexible architecture is able to host statistics algorithms and coding into needs only little programming experience. Moreover we will offer mentorship to get into existing interfaces. Existing documentation, resources and source code already exist on Internet for all of these algorithms. That's why this proposal is an easy yet motivating idea.

Resources:

- Detailed algorithms specification (https://nwb.slis.indiana.edu/community/?n=Algorithms.HomePage)
- JUNG (http://jung.sourceforge.net)

Algorithms' publications:

- **HITS :** Kleinberg JM (1999). "Authoritative sources in a hyperlinked environment." Journal of the ACM, 46(5), 604–632. URL http://citeseer.ist.psu.edu/kleinberg99authoritative.html

- **Betweenness Centrality:** Brandes U (2001). “A Faster Algorithm for Betweenness Centrality.” Journal of Mathematical Sociology, 25(2), 163–177.
- **Pagerank:** Page L, Brin S, Motwani R, Winograd T (1998). “The PageRank Citation Ranking: Bringing Order to the Web.” Technical report, Stanford Digital Library Technologies Project. http://citeseer.ist.psu.edu/page98pagerank.html
- **Clustering Coefficient:** Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.