**Assignment 1**
**Link prediction**
**DUE: Tue. March 10, 2015**

# 1 Question 1

1. For the given network G <V, E> we choose four times: $t_0 < t_0^{'} < t_1 < t_1^{'}$. Then we give to prediction algorithm the access to the network G$[t_0, t_0^{'}]$ and it must output the list of edges not presented in G$[t_0, t_0^{'}]$, that are predicted to appear in network G$[t_1, t_1^{'}]$. It was link prediction problem definition.
   There are different predictors, that uses different heuristics. They are: common neighbours, Jaccard's coefficient, Adamic/Adar, preferential attachment, $Katz_\beta$, rooted $PageRank_\alpha$, $SimRank_\gamma$ and few others.
   The dataset was used for this research consists of 5 subjects from scientific collaboration network. Only nodes with enough amount of publication were discovered.
   The core consists of authors with 3 publications in training and test datasets.
   There are methods, based on neighbourhood. They consider the following assumption: node X and Y will more likely perform a connection in future if the have a lot of common connections.
   Another group of methods is Ensemble of all paths. It refines the notion of shortest part by considering the ensemble of all paths between two nodes.
   Also there are high level approaches that can be combined with groups of methods, mentioned before.

2. Comparison with random link predictor, that chooses random two nodes to predict edge between them has shown that all predictors are better then random one.
   Furthermore there is a problem of small world: there are a lot of common connections in academic society that fails methods, measured neighbourhood.

3. We can consider all the network without restrictions and limitations. But information will be more correct. However the size of the problem will increase. We can also measure relevance of subjects in collaborations and introduce this concepts in our model.

# 2 References

[1] David Liben-Nowell and Jon Kleinberg. “The link-prediction problem for social networks”. In: Journal of the American society for information science and technology 58.7 (2007), pp. 1019–1031.