Thesis Proposal 3: Prediction of Social Networking Dynamics

Advisor: Dimitar Shterionov
Promoter: Gerda Janssens
Research group: Declarative Languages and Artificial Intelligence (DTAI)

Context:
Currently, many people stay in touch using online social networks as Facebook, Twitter, LinkedIn, Google+, etc. Social networks also describe relations between people sharing common interests. Some social networks have a very large size (e.g., Facebook size grew from 150000000 active users in 2009 to over 400000000 in 2010). Learning from large-size networks, although quite useful, is hard. But we can reason about relations in social networks and their evolution based on smaller-size samples (snapshots). The confidence of the result, however, will be characterized by a probabilistic factor corresponding to sampling ratio, size of the samples, characteristics of sampled users, etc. This thesis is a study on how to model social networks using (probabilistic) logic programming and how to reason about its relations.

Goal:
Study of the dynamics of social networks from the perspective of probabilistic logic programming.

Main Points:
• collect social networks' data – from existing data or from a crawler which runs on regular time periods.
• process the data and create required sets (for training, testing and verifying the modeling methods).
• encode the data as a (probabilistic) logic program.
• write algorithm(s) for analyzing the dynamics of the networks (i.e. the samples) and retrieving probabilistic regularities in the network's evolution.

Towards the student:
For this thesis, you will be required to collect data sets representing social networks, either from existing sources, or by creating a crawler to collect the required information. You should suggest a relational model of the collected data. You will need to have knowledge of scripting language(s), eg., php, to build the crawler. You will have to work with probabilistic logic programming – to encode and reason about the data and also to become familiar with graph theory. The time/work load can be divided into 40% literature and 60% coding.

You will have to build a crawler to gather necessary information. This script should not violate any law or privacy setting. It is advisable not to restrict yourself to one particular network (e.g., Facebook). Also interesting are new and (currently) faster evolving networks, eg., google+.

After retrieving separate snapshots of the network find a logical relationship between them and a probabilistic expression(s) which characterizes it. Investigate the possibility to predict (accurately) the evolution of the network. Namely, what will the social network look like in the next time step. Compare the prediction with real results. Fine-tune and generalize the algorithm for state prediction of arbitrary systems: using the notion of a “Markov chain”, write an algorithm to predict the next stage of an arbitrary Markov chain, which is given as input.

Literature and related work:
• http://arxiv.org/abs/0906.0060
• www.wseas.us/e-library/conferences/2007egypt/papers/568-496.pdf
• Amit Goyal, Francesco Bonchi, Laks V. S. Lakshmanan – Learning Influence Probabilities In Social Networks
• http://www-personal.umich.edu/~mejn/netdata/
• Salvatore A. Catanese, Pasquale De Meo, Emilio Ferrera, Giacomo Fiumara, Alessandro Provetti – Crawling Facebook for Social Network Analysis Purposes
• Alan Mislove, Hema Swetha Koppula, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee, Growth of the Flickr Social Network
• Matthew A. Russel, Mining the Social Web
• etc.