MODELING DYNAMICS OF SOCIAL AND COMMUNICATION NETWORKS

By

Konstantin Mertsalov

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: COMPUTER SCIENCE

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Mark Goldberg, Thesis Adviser
James Hendler, Member
Malik Magdon-Ismail, Member
William A. Wallace, Member

Rensselaer Polytechnic Institute
Troy, New York

December 2009
(For Graduation December 2009)

ABSTRACT

Communication and social networks have been shown to change over time. We present a novel approach to modeling the dynamics of large social and communication networks. Our model is based on the locality principle; it postulates that every node of the network is associated with a small local area, comprised of the nodes "close" to the center-node. The nodes of the locality are those that are most likely to receive communications from the center-node. Different implementations of the locality idea were considered and tested on the real-life network of blogs managed by the blog-provider LiveJournal and on the network of Twitter messages. It turned out that the locality areas generated by short random walks with probabilistic restarts produce the best approximation for both LiveJournal and Twitter networks.

Our method for measuring the network dynamics and validation of the model includes a set of parameters traditionally used for this purpose, such as average diameter of the graph and the clustering coefficient, and the vertex in-degree and out-degree distribution. While out-degree distribution, used as the input, represents the individual properties of the network nodes, the vertex in-degree distribution represents the global property of the network, and it is used for testing and the validation of the model.

The communication network, which can be measured, is viewed as a sampling of the invisible social relationship network. Traceable communication data was used for reasoning about a social network and modeling its dynamics. We used two datasets for varification - Twitter message network and LiveJournal comments graphs. We developed a machinery for collecting the communication data from the public section of LiveJournal. The data we collected covers more than two years of activity and contains information related to 1.8 million bloggers, 79 million posts and 223 million comments. Using permutation methods for the statistical analysis of the observed network, we measured and characterized the network dynamics. It turns out that individuals in the Blogosphere maintain surprisingly stable relationships i.e. if a pair communicated in the past, they are likely to continue communicating

as long as both stay active. On the other hand, most bloggers were not persistently active for a long time. We also observed that bloggers choose to communicate with individuals who were very close to them (via shortest distance) in the historical communication network. These observations allowed us to fit the model for accurate simulation of LiveJournal dynamics.