

**AGENT-BASED DYNAMICS MODELS FOR OPINION
SPREADING AND COMMUNITY DETECTION IN
LARGE-SCALE SOCIAL NETWORKS**

By

Jierui Xie

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:

Boleslaw K. Szymanski, Thesis Adviser
Gyorgy Korniss, Thesis Adviser
Mark K. Goldberg, Member
Mohammed J. Zaki, Member

Rensselaer Polytechnic Institute
Troy, New York

May 2012
(For Graduation August 2012)

ABSTRACT

Human behavior is profoundly affected by the interaction of individuals and the social networks that link them together. In this thesis, two topics in the context of social network analysis (SNA) are studied. One is the opinion dynamics, and the other is the community structure discovery.

Opinion dynamics. In order to provide useful insights into understanding the evolution of opinions, ideologies or attitudes, this thesis explores a simple, abstract opinion dynamics model, called *binary agreement model*, where there are two *competing* opinions. The contribution of this thesis is to quantify the effect of *committed minorities* who hold unshakable opinions. In particular, such effect is investigated in two scenarios.

The first scenario is one committed group competing with the opposing uncommitted majority. The study shows how the prevailing majority opinion in a population can be rapidly reversed by a small fraction p of randomly distributed committed agents. When the committed fraction grows beyond a critical value p_c , there is a dramatic decrease in the time, T_c , taken for the entire population to adopt the committed opinion. For complete graphs, when $p < p_c$, $T_c \sim \exp(\alpha(p)N)$, where α is a function of p and N is the network size, while for $p > p_c$, $T_c \sim \ln N$. Simulation results for sparse networks (e.g., Erdős-Rényi (ER) random graphs, Barabasi-Albert (BA) scale-free networks and real-world social networks) show qualitatively similar behavior.

The second scenario is the more general case where two groups committed to distinct opinions A and B , and constituting fractions p_A and p_B , coexist. The study shows using mean-field theory that the phase diagram in parameter space (p_A, p_B) , consists of two regions, one where two opinions coexist, and the remaining where one opinion always dominates the other. The scaling exponent associated with the exponential growth of switching times is found to be a function of the distance from the second-order transition point. Lastly, the nature (i.e., discontinuous and continuous) of transitions on sparse networks is explored by decomposing the system

into linear trajectories and deriving appropriate order parameters. The simulation results show that sparse networks are also characterized by the same qualitative phase diagram as the fully connected networks. As a comparison, the influence of global social media that serves as a kind of committed opinion is briefly discussed.

Community detection. Mining communities that allow multiple memberships is challenging especially in large-scale networks. This thesis presents a fast algorithm, called Speaker-listener Label Propagation Algorithm (SLPA), for *overlapping* community detection. SLPA is an extension of Label Propagation Algorithm (LPA). It spreads labels according to dynamic interaction rules and maintains label distributions in nodes' memories. Experiments in both synthetic and real-world networks show that SLPA has an excellent performance in identifying both node and community level overlapping structures. The performance is remarkably stable under different quality measures including normalized mutual information (NMI), Omega Index and F-score. SLPA can be applied to various structures, including weighted, unweighted, directed and undirected networks. With time complexity that scales linearly with the number of edges in the network, SLPA is successfully applied to networks with million nodes.

Detecting and tracking communities in a dynamic network where changes arrive as a stream is another challenging issue in real-world applications. Instead of computing communities on each snapshot independently, algorithms that *incrementally* update communities are very useful in the case of real time monitoring of huge data streams such as the Internet traffic or online social interactions. This thesis proposes *LabelRankT*, a decentralized online algorithm, to detect evolving communities in large-scale dynamic networks. LabelRankT by its own is based on a stabilized label propagation algorithm proposed in this thesis. It maintains the previous partitioning and dynamically updates only nodes involved in changes. As compared to other static algorithms including MCL and Infomap, LabelRankT achieves similar performance but with lower computational cost. Furthermore, it significantly outperforms and is more than 100 times faster than dynamic detection algorithms such as facetNet and iLCD. Importantly, LabelRankT is highly parallelizable allowing the computation to be distributed to each individual node. Such property will be

particularly useful for applications like wireless sensor networks and mobile ad hoc networks, where each node in the network corresponds to a physical platform.