Abstract

Online social and information networks, like Facebook and Twitter, exploit the influence of neighbors to achieve effective information sharing and spreading. The process that information is spread via the connected nodes in social and information networks is referred to as diffusion. In the literature, a number of diffusion models have been proposed for different applications like influential user identification and personalized recommendation. However, comprehensive studies to discover the hidden diffusion mechanisms governing the information diffusion using the data-driven paradigm are still lacking. This thesis research aims to design novel diffusion models with the structural and behaviorable dependency of neighboring nodes for representing social networks, and to develop computational algorithms to infer the diffusion models as well as the underlying diffusion mechanisms based on information cascades observed in real social networks.

By incorporating structural dependency and diversity of node neighborhood into a widely used diffusion model called Independent Cascade (IC) Model, we first propose a component-based diffusion model where the influence of parent nodes is exerted via connected components. Instead of estimating the node-based diffusion probabilities as in the IC Model, component-based diffusion probabilities are estimated using an expectation maximization (EM) algorithm derived under a Bayesian framework. Also, a newly derived structural diversity measure namely dynamic effective size is proposed for quantifying the dynamic information redundancy within each parent component.

The component-based diffusion model suggests that node connectivity is a good
proxy to quantify how a node’s activation behavior is affected by its node neighborhood. To model directly the behavioral dependency of node neighborhood, we then propose a co-activation pattern based diffusion model by integrating the latent class model into the IC Model where the co-activation patterns of parent nodes form the latent classes for each node. Both the co-activation patterns and the corresponding pattern-based diffusion probabilities are inferred using a two-level EM algorithm. As compared to the component-based diffusion model, the inferred co-activation patterns can be interpreted as the soft parent components, providing insights on how each node is influenced by its neighbors as reflected by the observed cascade data.

With the motivation to discover a common set of the over-represented temporal activation patterns (motifs) characterizing the overall diffusion in a social network, we further propose a motif-based diffusion model. By considering the temporal ordering of the parent activations and the social roles estimated for each node, each temporal activation motif is represented using a Markov chain with the social roles being its states. Again, a two-level EM algorithm is proposed to infer both the temporal activation motifs and the corresponding diffusion network simultaneously. The inferred activation motifs can be interpreted as the underlying diffusion mechanisms characterizing the diffusion happening in the social network.

Extensive experiments have been carried out to evaluate the performance of all the proposed diffusion models using both synthetic and real data. The results obtained and presented in the thesis demonstrate the effectiveness of the proposed models. In addition, we discuss in detail how to interpret the inferred co-activation patterns and interaction motifs as the diffusion mechanisms under the context of different real social network data sets.

**Keywords:** Social networks, diffusion networks, Independent Cascade Model, structural diversity, co-activation patterns, temporal activation motifs, expectation maximization algorithms
Table of Contents

Declaration i

Abstract ii

Acknowledgements iv

Table of Contents vi

List of Tables ix

List of Figures x

Chapter 1 Introduction 1

1.1 Social Networks and Information Diffusion 1

1.2 Structural and Behavioral Dependency of Neighbors for Diffusion 4

1.3 Our Contributions 6

1.3.1 Modeling Structural Dependency of Node Neighborhood (Chapter 3) 6

1.3.2 Modeling Behavioral Dependency of Node Neighborhood (Chapter 4 and 5) 7

1.4 Organization of the Thesis 8

Chapter 2 Related Work on Diffusion Modeling in Social Networks 9

2.1 Independent Cascade and Linear Threshold Models 9

2.2 Game-theoretic Models 13
2.3 Stochastic Process Models ........................................... 14
2.4 Epidemic Diffusion Models ........................................... 16
2.5 Other Variants of Diffusion Models ................................. 18
2.6 Summary ................................................................. 19

Chapter 3  A Component-based Diffusion Model with Structural Diversity for Social Networks 20
3.1 Introduction ............................................................... 21
3.2 Problem Formulation and Methodology ............................ 22
   3.2.1 Preliminaries ..................................................... 23
   3.2.2 Problem Formulation .......................................... 23
   3.2.3 Learning Algorithm ............................................ 31
3.3 Experiments ............................................................. 39
   3.3.1 Experimental Settings ........................................... 39
   3.3.2 Performance Evaluation ....................................... 41
   3.3.3 Experiments on Synthetic Data ................................. 42
   3.3.4 Experiments on Real Data ...................................... 46
   3.3.5 Run-time ............................................................ 49
   3.3.6 Sensitivity to Component Identification Methods ........... 50
3.4 Summary ................................................................. 52

Chapter 4  Inferring Latent Co-activation Patterns for Information Diffusion 53
4.1 Introduction ............................................................... 54
4.2 Problem Formulation and Methodology ............................ 56
   4.2.1 Preliminaries ..................................................... 56
   4.2.2 Formulation ........................................................ 57
   4.2.3 Learning Algorithm ............................................ 58
4.3 Experiments ............................................................. 61
   4.3.1 Experimental Settings ........................................... 62