

Preface

Introduction

Most of the contributions in the present book contain recent studies on community detection and/or evolution and represent extended versions of a selected collection of articles presented at the 2013 IEEE/ACM international Conference on Advances in Social Network Analysis and Mining (ASONAM), which took place in Niagara Falls in Canada between August 25 and 28, 2013. The topics covered by this book can be categorized into two groups: community detection and evolution in the first seven chapters, and two other related topics, namely link prediction and influence/information propagation or maximization, in the last four chapters.

Community Detection and Evolution

The discovery of cohesive groups, cliques, and communities inside a network is one of the most studied topics in social network analysis. It has attracted many researchers in sociology, biology, computer science, physics, criminology, and so on. Community detection aims at finding clusters as subgraphs within a given network. A community is then a cluster where many edges link nodes of the same group and few edges link nodes of different clusters.

A general approach to community detection consists in considering the network as a static view in which all the nodes and links in the network are kept unchanged throughout the study. Recent studies focus also on community evolution since most social networks tend to evolve over time through the addition and deletion of nodes and links. As a consequence, groups inside a network may expand or shrink and their members can move from one group to another one over time.

Most of the studies on community evolution use topological properties to identify the updated parts of the network and characterize the type of changes such as network shrinking, growing, splitting, and merging. However, recent work has

focused on community evolution/detection by relying entirely on the behavior of group members in terms of the activities that occur in the network rather than exclusively considering links and network density.

Another interesting feature of social networks is the cohesiveness of a group and how it varies over time. In fact, the cohesiveness of a group is a social factor that assesses how members of a group are close to each other, and may help predict a possible community splitting or disaggregation. Chapters “[Entanglement in Multiplex Networks: Understanding Group Cohesion in Homophily Networks](#)”–“[The Power of Consensus: Random Graphs Have No Communities](#)” are proposed to portray trends towards cohesiveness evaluation.

Chapter about “[The Emergence of Communities and Their Leaders on Twitter Following an Extreme Event](#)” by Yulia Tyshchuk, Hao Li, Heng Ji, and William A. Wallace, combines natural language processing together with social network analysis to explore Twitter messages in order to identify actionable ones, construct an actionable network, identify communities with their central actors, and show the behavior of the community members. The approach has been evaluated on two important real-life events, namely the 2011 Japan Tsunami and the 2012 Hurricane Sandy. The results help understand the behavior of communities as a whole or as individual members of such cohesive groups. Since the two events have different characteristics, the behavior of involved people is dissimilar from one event to the other one. In particular, it was observed that there was a limited participation of Government on Twitter during the 2011 Japan Tsunami compared to an active involvement during the 2012 Hurricane Sandy. Moreover, the leadership roles were stronger in the second than in the first event, while the cohesion in virtual communities on Twitter seems weaker for the Hurricane Sandy.

Chapter titled “[Hierarchical and Matrix Structures in a Large Organizational Email Network: Visualization and Modeling Approaches](#)” by Benjamin H. Sims, Nikolai Sinitsyn, and Stephan Eidenbenzof studies the visualization and modeling aspects of community detection. Indeed, the email network of a large scientific research organization is analyzed in order to visualize and model organizational hierarchies in complex network structures. To that end, formal organizational divisions and levels are integrated with network data to get an insight into the interactions between subdivisions of the organization and other external organizations. In order to manage the complexity of the large email network, the Girvan-Newman algorithm for community detection is applied. Then, a power law model to forecast degree distribution of organizational email traffic is defined based on the hierarchies that hold between managers and employees.

Chapter labeled “[Overlaying Social Networks of Different Perspectives for Inter-network Community Evolution](#)” by Idrissa Sarr, Joseph Ndong, and Rokia Missaoui uses probability and possibility theories as two alternate solutions to discover perspective (temporary) communities and highlight community evolution. Starting from snapshots of the network at different time periods, the underlying social network is analyzed in order to first identify active actors (i.e., actors that participate in at least a predefined number of activities) during a set of time slots, and then delimit the perspective communities they form over time. Beside the fact

that the approach tracks the evolution of the network and identifies the perspective communities, it gives a basic way to identify both active and passive users. The latter group of users can be seen as churners in customer relationship management (CRM) applications. Furthermore, mapping perspective communities to an initial (or important) network adds new links that improve the network accessibility, and hence the information flow circulation.

Chapter titled “[Study of Influential Trends, Communities, and Websites on the Post-election Events of Iranian Presidential Election in Twitter](#)” by Seyed Amin Tabatabaei and Masoud Asadpour analyzes 1,375,510 tweets of Twitter users who were interested in Iranian Presidential election and its post-events. The top URLs that appeared on the tweets indicate that the most influential websites are those related to social networking and social media websites. Important keywords used in the tweets during nine days are extracted and the most popular websites among two distinct groups of users (Persian and English speaking users) are found. These groups represent the core part of the network and help in interacting with abroad to communicate the news, events, and messages. Peripheral users are identified as well as a few subcommunities within the groups. The specification of subcommunities (i.e., the supporters of political groups) is done based on the keywords extracted from the tweets using a customized version of TF-IDF. Another result shows a strong link between the posted tweets and the political events that occurred the same day.

Chapter titled “[Entanglement in Multiplex Networks: Understanding Group Cohesion in Homophily Networks](#)” by Benjamin Renoust, Guy Melançon, and Marie-Luce Viaud deals with group cohesiveness in complex networks, mainly, in bipartite graphs. The authors use the homophily concept to assess similarity between actors and the group homogeneity they have. The key idea is that attributes are exploited while investigating how they interact. In other words, authors focus on measuring the cohesion of a group through the interactions that take place between attributes of actors. Hence, actor behavior is used to measure the intensity of interactions and group cohesiveness. Therefore, it can be stated that interactions between actors are a key element to identify group structure and cohesiveness. Instead of projecting a bipartite network onto a single-type network with entities of a same type, which can lead to a loss of information or hide subtle characteristics of the original data, the authors propose to directly study the multiplex networks. By doing so, they demonstrate the feasibility of detecting community structure within complex networks without the need to compute one-mode projections.

Chapter titled “[An Elite Grouping of Individuals for Expressing a Core Identity Based on the Temporal Dynamicity or the Semantic Richness](#)” by Billel Hamadache, Hassina Seridi-Bouchelaghem, and Nadir Farah is related to group detection and especially to core identification in social networks. The core of a network can be seen as a central part having a high influence on the communication flows that involve the other nodes. Basically, the work can be seen as another contribution to existing studies in group detection by adding the semantic and temporal dimensions. In fact, temporal dynamic behavior or semantic concepts of social entities are an additional input to exploit in order to characterize and strengthen significantly a group structure and highlight its cohesiveness. The key idea of this work is that

actors of a social network are likely to change their interactions over time by adding or removing relations with others. This has an impact on their social position in the network and/or their possible affiliation to one or more social groups. The temporal change is in fact induced by many factors influencing actor behavior. Therefore, using a semantic dimension such as the connection causality, the positive opinion of socializing, and relationship kinds may help gauge the shape of groups and their cohesiveness.

Chapter by Romain Campigotto and Jean-Loup Guillaume on “[The Power of Consensus: Random Graphs Have No Communities](#)” defines the notion of consensual communities and shows that they do not exist within a random graph. The principle exploited by the authors is that the outcome of multiple runs of a non-deterministic community detection algorithm is certainly more significant than the outcome of a single run. Authors define a consensual community as a set of nodes, which are frequently classified in the same community through multiple computations. In other words, a consensual community is a repeatable outcome (set of communities) obtained from a set of community detection algorithm computations. The main reason for using consensual communities rather than classical communities comes from the fact that most techniques used to compute communities can usually provide more than one solution. This may depend on the initial configurations or the order in which nodes are considered. Moreover, consensual communities can provide a deeper insight into the structure of the network since they summarize many partitions and encode more information on the structure such as figuring out the overlapping communities. However, when considering random graphs, authors show that it is quite impossible to find consensual communities. The reason is that all pairs of nodes have the same probability to be connected in random graphs. Furthermore, authors demonstrate through various community detection algorithms the existence of a threshold beyond which a trivial consensual community containing all the nodes is found and below which each node forms a consensual community.

The remainder of the book covers a few use cases of community structures that address other issues in social network analysis, namely link prediction and influence/information propagation and maximization.

Link Prediction

This important topic in social network analysis aims at predicting if two given nodes have a relationship or will form one in the near future. It is exploited in many social media applications such as the ones that need an embedded recommender system to suggest new and relevant ties to the users. Like in community detection, similarity and proximity principles are widely used for link prediction. Moreover, information about network communities can improve the accuracy of similarity-based link prediction methods.

Chapter “[Link Prediction in Heterogeneous Collaboration Networks](#)” written by Xi Wang and Gita Sukthankar concerns link prediction in heterogeneous collaboration networks. It studies both supervised and unsupervised link prediction in networks where nodes may belong to more than one community, procreating different types of collaborations. Links in heterogeneous networks happen for different reasons, and hence cannot be considered in a homogeneous manner. To take into account such a fact, a new supervised link prediction framework, called Link Prediction using Social Features (LPSF), is proposed and integrates a re-weighting scheme of the network by exploiting features of nodes extracted from patterns of salient interactions in the network. It is shown that the proposed re-weighting method in LPSF better reflects the intrinsic ties between nodes and provides a better prediction accuracy for supervised link prediction methods.

Chapter titled “[Characterization of User Online Dating Behavior and Preference on a Large Online Dating Site](#)” by Peng Xia, Kun Tu, Bruno Ribeiro, Hua Jiang, Xiaodong Wang, Cindy Chen, Benyuan Liu, and Don Towsley studies user behavior of an online dating website in order to understand how user attributes can help predict who will date whom. By doing so, the authors try to provide outstanding guidelines to design a recommendation system for online dating website. This means that the present work can be seen as a link prediction issue since the recommendation is done once two users are likely to date based on their profiles. An interesting aspect that this paper points out is that the connections between individuals in the underlying network are not deeply related to simple and traditional mechanisms such as preferential attachment or homophily. Actually, user attributes based on preferential attachment cannot be simply used because user behavior in choosing attributes at a given date may largely be done randomly. Moreover, authors observe that the geographic distance between two users and the photo count of users play an important role in their dating behavior, and therefore it is important to differentiate between the effective preferences of users and the random selection of attributes. The main concerns during the approach validation are: (1) How often does a user send and receive messages and how does these operation change over time? and (2) What is the correlation or link between the sender and receiver behavior based on their own profiles?

Influence/Information Propagation and Maximization

Influence propagation is usually modeled using propagation models such as Linear Threshold Model and Independent Cascade Model. These models assume that a node is influenced based on the opinions of the local network neighborhood. It has been recently shown that it is more simple and realistic to model the propagation of negative influence, which is more contagious, than modeling the positive influence. Moreover, relying on community membership to study influence maximization is a viable alternate solution that researchers have considered recently as described in the last two chapters of this volume.

Chapter titled “[Latent Tunnel Based Information Propagation in Microblog Networks](#)” by Chenyi Zhang, Jianling Sun, and Ke Wang deals with Information propagation without relying entirely on the link structure of social networks. The key novelty of the approach is to mine the published messages within a microblog platform and extract the hidden topics to identify the seed users. The basic assumption is that a target message is more likely to be forwarded or re-tweeted if it is interesting to both the sender and the recipient, and an interested user is more likely to react to a message. Hence, when a topic catches the attention of two actors through previous messages, the authors conclude that both actors will probably react to the messages related to that topic and share a hidden link. They afterward identify the seeds of users that will maximize the propagation by identifying those actors, which, when they publish a message, their recipients are likely to forward it, and so on. To reach their goal, the authors unveil the latent topics associated with social links by relying on a standard topic modeling technique based on Latent Dirichlet Allocation. The modeling approach highlights the topic distribution for each link that explains its nature in information flow. These obtained distributions are used to estimate the propagation probability of a link for the target message.

Chapter by Mahsa Maghami and Gita Sukthankar about “[Scaling Influence Maximization with Network Abstractions](#)” tackles the problem of influence maximization in social networks with an application in the advertising domain. A solution is developed to find the influential nodes in a social network as targets of advertisement based on the network structure, the links among the actors in the network, and the limited advertising budget. The solution is a hierarchical influence maximization approach for product marketing that constructs an abstraction hierarchy to scale and adapt optimization techniques to larger networks. An exact solution is provided on smaller partitions of the network, and a candidate set of influential nodes is selected to be propagated upward to an abstract representation of the original network. The process of abstraction, solution, and propagation is iteratively executed until the resulting abstract network becomes small enough to use an exact optimization solution.

To conclude this preface, we would like to thank all authors for their significant contributions that give a broad spectrum of research work on social network analysis, mainly in community detection and evolution, link prediction, and influence propagation. Our warm thanks go also to the reviewers for their careful evaluation of the submissions and their useful comments and suggestions.

August 2014

Rokia Missaoui
Idrissa Sarr

Social Network Analysis - Community Detection and
Evolution

Missaoui, R.; Sarr, I. (Eds.)

2014, XVIII, 272 p. 108 illus., 102 illus. in color.,

Hardcover

ISBN: 978-3-319-12187-1