Chapter 7
Dynamic Social Network Mining: Issues and Prospects

Luca Cagliero
Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy

Alessandro Fiori
IRC@C: Institute for Cancer Research at Candiolo, Italy

ABSTRACT
The past few years have witnessed the rapid proliferation of Web communities such as social networking sites, wikis, blogs, and media sharing communities. The published social content is commonly characterized by a high dynamicity and reflects the most recent trends and common user behaviors.

The Data mining and Knowledge Discovery (KDD) process focuses on discovering and analyzing relevant information hidden in large data collections to support expert decision making. Hence, the application of data mining techniques to data coming from social networks and online communities is definitely an appealing research topic.

This Chapter overviews most recent data mining approaches proposed in the context of social network analysis. In particular, it aims at classifying the proposed approaches based on both the adopted mining strategies and their suitability for supporting knowledge discovery in a dynamic context. To provide a thorough insight into the proposed approaches, main work issues and prospects in dynamic social network analysis are also outlined.

INTRODUCTION
During recent years, the steadfast popularity increase of social networks and online communities has caught the attention of the data mining research community. Since their introduction, social network sites have attracted millions of users, many of whom have integrated these sites into their daily practices. The possibility to easily share multimedia content, thoughts, and electronic documents opens room to a novel form of communication and gives access to a huge amount of shared information.

DOI: 10.4018/978-1-4666-4213-3.ch007
Since social data coming from social networks, blogs, twikis, and bookmarking sites has become a powerful source of knowledge, the in-depth analysis of both the social network structure and the published content may trigger many insightful actions. For instance, the study of common social user behaviors may significantly improve the effectiveness of e-commerce services. Moreover, the analysis of the annotations made by Web users may drive product recommendation as well as ease Web resources’ retrieval.

The application of Data mining and Knowledge Discovery (KDD) techniques to social data has become more and more attractive for research purposes. For instance, Heymann et al. (2008) discovered relevant tag associations to model the metadata associated by Web users to public resources. The discovered tag associations are also used to drive the annotation process of partially annotated resources. In parallel, social tagging systems have also been used to classify videos and photos in order to improve the performance of search engines (Yin et al., 2009). At the same time, Wikipedia content has been used to infer semantics-based models, such as ontologies and taxonomies (Suchanek et al., 2008).

Data mining approaches entail either modeling peculiar characteristics of the analyzed data or predicting the label/value to assign to a given object/variable given a set of labeled data. The large variety of different data mining techniques which may be applicable to social data prompts the need of better understanding (i) the objective of the most recently proposed algorithms, (ii) their suitability for being applied in dynamic contexts, and (iii) the usefulness of the discovered knowledge for in-depth analysis. Peculiar aspects that should be taken into account by data mining analysts are, for instance, the algorithm efficiency in terms of computational time, the average prediction accuracy, and the succinctness and readability of the generated models. Since these properties actually depend on the social context under analysis, choosing the most suitable data mining approach is a challenging task. Furthermore, since social data continuously evolve over time, data mining algorithms are often required to be easily adaptable to dynamic contexts.

Interested readers may look into the large body of survey papers or books regarding social network analysis and mining (Liu, 2007; Scott & Carrington, 2011; Bonchi et al., 2011). They describe the characteristics of real-life social Web systems (Scott & Carrington, 2011), the applicability of intelligent data mining and information retrieval techniques (Liu, 2007), as well as they provide an in-depth insight into the most challenging research directions in social data analysis (Bonchi et al., 2011). However, a systematic classification and overview of the most recent data mining approaches applied to data coming dynamic social networks and online communities are still missing.

This chapter focuses on overviewing recently proposed data mining approaches in the context of social network and online community analysis. Unlike previous overviews (e.g. Liu, 2007; Scott & Carrington, 2011; Bonchi et al., 2011), this work provides a detailed overview of state-of-the-art approaches from a data mining perspective. In particular, it aims at better clarifying the scope of recent data mining research in social network analysis, highlighting the issues encountered by the most renewed data mining systems, and discussing prospects of the data mining research on social networks. Hence, a particular attention has been paid to the mining methods applied to perform social network mining and analysis. Furthermore, a relevant effort has also been devoted to better understanding the ability of data mining research in capturing the evolution of the underlying social networks through the analysis of the most recurrent trends. To achieve this goal, an analysis of the most established data mining applications oriented to dynamic social network analysis is presented.

This chapter is organized as follows. Section “Data mining techniques” outlines fundamentals of five among the most established data mining