Chapter 3
Graph Mining Techniques for Networking Applications: A Review
Charu C. Aggarwal
IBM T. J. Watson Research Center, USA
Eric Yu-En Lu
Cambridge University, UK

ABSTRACT
Computer networks are inherently graphical in structure, since they contain multiple geographically distributed nodes, which are connected by communication links. In many cases, such as large military networks, this also leads to a network-centric view of the data. This naturally leads to an information network representation of the data. In information networks, the nodes represent the different elements of information, and the links represent the logical relationships between them. An even higher layer of logical organization is a social network, in which the nodes correspond to the different entities in the network, and the links correspond to the interactions. The common element among the different kinds of networks which are encountered in practical network-centric representation of the data is that they can all be represented in the form of a structural graph. This structural behavior can be used in order to glean different kinds of practical insights. This chapter will provide an overview of these different graph mining techniques and their applications to the various kinds of networks.

INTRODUCTION
This chapter will study a number of graph mining applications in the context of network-centric applications. Such graph mining applications are relevant to different levels of conceptualization of network-centric applications. We will study three different levels of conceptualization, which are referred to as information, social, and communication networks respectively. The lowest level of conceptualization for network-centric applications are communication networks which carry the flow of information between different nodes. It has recently been noticed that the structural behavior (Chakrabarti et. al. 2008) (Chakrabarti et. al. 2004) of communication networks is very useful in characterizing the behavior of network-centric applications which are built on top of these networks. Higher
levels of logical abstraction are built on top of large communication networks which carry the actual flow of information. Some examples of such abstractions are information networks and social networks.

An information network is a logical network of data, information, and knowledge objects, which may be acquired and extracted from disparate sources such as documents, images, sensors, as well as advanced data analytical processes. Information networks are thus defined as conceptual or physical organizations of information in which nodes represent information from a variety of sources, and the links represent the conceptual and logical relationships between these different pieces of information. Information networks essentially provide a network-centric view of data organization. The links in an information network are distinct from the physical links in a communication network, though the underlying physical links may also affect the conceptual organization represented by the information network links. Furthermore, the underlying links may be very different for different kinds of information networks. For example, in the case of the web, this may correspond to pointers between web pages (or hyperlinks), and in the case of military networks, these links may refer to how the information elements in different entities interact with one another. In general, information networks refer to a network-centric organization of the data and their relationships with one another.

Social networks are closely related to information networks, can be considered a higher layer representation. In social networks, nodes correspond to entities, which may interact with one another and the links between them may constitute a variety of interactions between the different participants. Examples of social networks include Facebook, LinkedIn, and Myspace. We note that many intra-organization and multi-organization applications share a number of characteristics with social networks. For example, the email communication between different organizations may have similar structure to a social network. In such cases, it may be useful to determine the important communities of interaction in the social network, and other structural behavior which provides feedback about the flow of information in the social network.

Since communication, information and social networks have a natural graph structure, it is interesting to explore whether some of the recently proposed techniques for graph mining can be used in order to extract and manage the underlying data in the information networks more effectively. In this chapter, we will examine this relationship. Furthermore, we will also examine how the underlying physical structure of the communication network may affect the network-centric organization. For example, large organizations or military coalitions may have large networks which they rely on for the flow of information between them. Similarly, the entities in a military battlefield, such as tanks, battalions, and other physical assets may form a logical information network, in which the nodes correspond to the information elements in the different objects, and the links correspond to the relationships between them. In many cases, the structural relationships can be used to isolate important regions of the network which may contain information which is relevant to a specific task. This can be used to determine the nature of the information flows in the network, and the importance of the different nodes in such information flows.

This chapter will study the relevance of graph mining methods to a variety of networking applications. Specifically, we will study the relevance of recent graph and structural data mining techniques to these methods. This paper is organized as follows. In the next subsection, we will discuss the concept of information networking and its applications to multi-organization and coalition applications. In section 2, we will discuss different graph mining techniques, and their relevance to different information networking applications. Specifically, we will discuss recent mining techniques such as