Social Network Analysis

S

Roberto Marmo
University of Pavia, Italy

INTRODUCTION

This chapter describes what is the social networks analysis to quantitatively analyze individual behaviors and social relationships represented as a graph. This chapter discusses network information visualization to visualize a network, software useful to analyze data set from a large variety of *data sources*, others sections discuss applications such as Web mining techniques, influencer detection, churn prediction, viral marketing. Finally, future challenges and research directions are explained.

BACKGROUND

A social network is a social structure composed of individuals (organizations, company etc.) which are connected by relationships and interactions between individuals, so it is possible to create a rich relational interdependency and content for data mining.

Online social network focuses on building on Internet communities of people who share interests and/or activities, who are interested in exploring the interests and activities of others or who are interested to communicate, interact and share. Some well-known social networking Websites are: Facebook as general network, LinkedIn and Viadeo as business social network, Flickr about photo sharing etc. Thus, online social network is a relevant part of human life (Fu, Chen, Liu, & Wang, 2007; Goth, 2008) and it is truly the reflection of today's society.

MAIN FOCUS

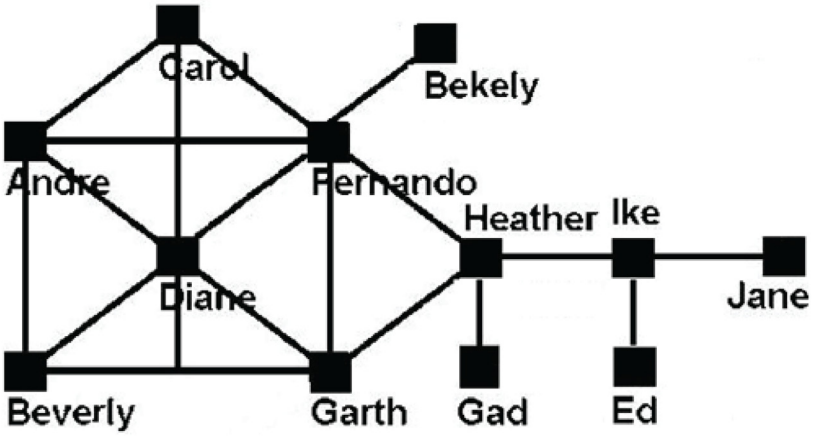
A social network can be modelled as graph-based visualization (Figure 1) composed of individuals (organizations, company etc.) also called nodes, which are connected by links represent relationships and interactions between individuals, a rich relational interdependency and content for mining.

Social network analysis (SNA) is a mathematical technique developed in modern sociology, in order to understand structure and behaviour between members of social systems, to map relationships and social connections between individuals in social network, also to serve up business intelligence on the ties. SNA assumes that individuals are all interdependent, instead other approaches to business problems assume that what individuals do, think, and feel is independent of who they know. SNA is related to network theory and graph theory, so the network topology helps to determine a network's usefulness to its individuals. It is possible to classify objectives in: static to find community structures, dynamic to monitor community structure evolution and to spot abnormal individuals or abnormal timestamps. Some typical problems in SNA include discovering groups of individuals sharing the same properties (Schwartz, & Wood, 1993), evaluating the importance of individuals (Kautz, Selman, & Milewski, 1996). Another goals of SNA regards to explain the observed network, identification of subgroup, inferring real-world connections and discovering, labelling, characterizing communities (Adamic & Adar, 2007).

Evaluation of people location in network, that is the centrality of a node, is relevant to understand networks and their participants. These measures provided by social network analysis give us insight

DOI: 10.4018/978-1-4666-5202-6.ch200

Figure 1. An example of social network: individuals on nodes (black square) and links represent relationships (black line)



into the various roles and groupings: who are the connectors, leaders, bridges, isolates, the clusters existing and who is in them, who is in the kernel of network and who is on the periphery.

SNA is descriptive, because it is built with only a few global parameters, so it is not useful for making prediction of future behaviour of network. This is due to networks availability, few information about each node and lack of data. In this sense, SNA assumes local variables (gender, contacts, etc), network variables depending on business problem, first-order variables (immediate connections behaviours) and second order variables (behaviours of friends of friends). In Web 2.0 age we have very large social networks creating massive quantities of data and we have substantial quantities of information at level of individual nodes suitable to build statistical models of individuals. The relevant difficult regards how to extract social data from a set of very different communication resources (Matsuo, Tomobe, & Nishimura, 2007).

The social connections of users can be a rich source of information and may be used to discover personal information about users. It is necessary to apply a more substantial procedure of sanitization on the graph before its release. Bonchi, Castillo, Gionis, and Jaimes (2011) propose a business process classification framework to put the research topics in a business context and provide an overview of what we consider key problems and techniques in social network analysis and mining from the perspective of business applications. In particular, this paper discusses data acquisition and preparation, trust, expertise, community structure, network dynamics, and information propagation.

SNA can perform predictive analysis. This includes using network phenomena such as a tie to predict individual level outcomes (often called peer influence or contagion modelling), using individual-level phenomena to predict network outcomes such as the formation of a tie/edge or particular type of triad, or using network phenomena to predict other network phenomena, such as using a triad formation at time 0 to predict tie formation at time 1.

Adjacency matrix is a simply way to represent a network by representing which vertices of a graph are adjacent to which other vertices: if person i and j are connected with one direct link we have (i,j)=1 and (i,j)=0 otherwise. Using matrix algebra on adjacency matrix it is possible to evaluate some numerical features useful for SNA, such as computing the intensity of relation between person (Kolaczyk, 2009; Scott, 2000; Wasserman & Faust, 1994) based on: