

Chapter 16

Web Mining and Social Network Analysis

Roberto Marmo
University of Pavia, Italy

ABSTRACT

Research on social networks has advanced significantly due to wide variety of on-line social websites and very popular Web 2.0 application. Social network analysis views social relationships in terms of network and graph theory about nodes (individual actors within the network) and ties (relationships between the actors). Using web mining techniques and social networks analysis it is possible to process and analyze large amount of social data (such as blogtagging, online game playing, instant messenger etc.) and by this to discover valuable information from data. In this way, we can understand the social structure, social relationships and social behaviors. This new approach is also denoted as social network mining. These algorithms differ from established set of data mining algorithms developed to analyze individual records, because social network datasets are called relational due to centrality of relations among entities. This chapter also sets out a process to apply web mining.

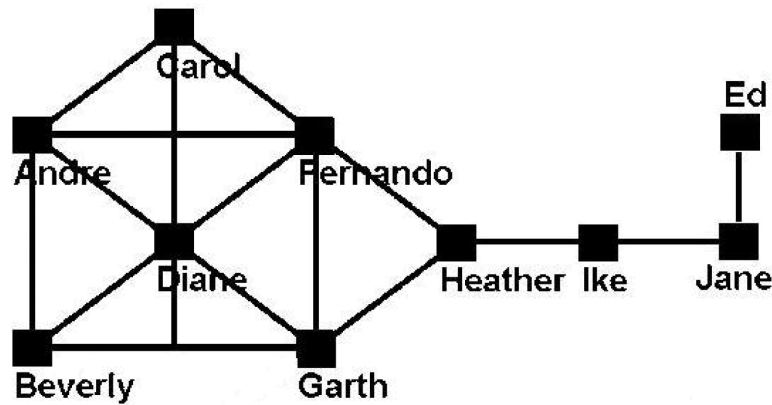
INTRODUCTION

A social network is a social structure made 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. Figure 1 shows an example of social network.

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. So, they are very popular Web 2.0 application. Some well-known social networking websites are: Facebook as general network, LinkedIn and Viadeo as business social network, Flickr about photo sharing etc. Thus, social network is a relevant part of human life (Fu, 2007; Goth, 2008).

DOI: 10.4018/978-1-4666-2455-9.ch016

Figure 1. An example of social network: individuals on nodes and links represent relationships



This chapter describes how to use web mining techniques and online social networks analysis to obtain analysis regarding user profile and behaviour, information suitable for marketing, sociology etc. The analysis are focused on web resources such as user content, network structures, user behaviour on network website and how user creates its network

This chapter is structured as follows: section 1 presents social network analysis, section 2 presents web mining algorithms and techniques can be used for social networks analysis, section 3 presents techniques and process, section 4 discusses some applications. Finally, future challenges and research directions are explained.

SOCIAL NETWORK ANALYSIS

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 between individuals in social network, also to serve up business intelligence on the ties. Social network analysis 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 com-

munity structure evolution and to spot abnormal individuals or abnormal time-stamps.

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 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.

Social network analysis is descriptive rather than predictive, 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 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 (Jin, 2007; Matsuo, 2007).

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 properties useful for

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/web-mining-social-network-analysis/73446

Related Content

Hybrid Query and Data Ordering for Fast and Progressive Range-Aggregate Query Answering

Cyrus Shahabi, Mehrdad Jahangiri and Dimitri Sacharidis (2005). *International Journal of Data Warehousing and Mining* (pp. 49-69).

www.irma-international.org/article/hybrid-query-data-ordering-fast/1751

Vertical Fragmentation in Databases Using Data-Mining Technique

Narasimhaiah Gorla and Pang W.Y. Betty (2008). *International Journal of Data Warehousing and Mining* (pp. 35-53).

www.irma-international.org/article/vertical-fragmentation-databases-using-data/1812

Classifying Very High-Dimensional Data with Random Forests Built from Small Subspaces

Baoxun Xu, Joshua Zhexue Huang, Graham Williams, Qiang Wang and Yunming Ye (2012). *International Journal of Data Warehousing and Mining* (pp. 44-63).

www.irma-international.org/article/classifying-very-high-dimensional-data/65573

From Change Mining to Relevance Feedback: A Unified View on Assessing Rule Interestingness

Mirko Boettcher, Georg Ruß, Detlef Nauck and Rudolf Kruse (2009). *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction* (pp. 12-37).

www.irma-international.org/chapter/change-mining-relevance-feedback/8435

Combining Data Warehousing and Data Mining Techniques for Web Log Analysis

Torben Pedersen, Jesper Thorhauge and Søren Jespersen (2007). *Research and Trends in Data Mining Technologies and Applications* (pp. 1-28).

www.irma-international.org/chapter/combining-data-warehousing-data-mining/28419