

Chapter 7

Graph Mining Approaches to Study Volunteer Relationships in Sourceforge.net

ABSTRACT

The contribution of volunteers in the development of Free and Open Source Software in Sourceforge.net is studied in this paper. Using Social Network analysis, the small set of developers who can maximize the information flow in the network are discovered. The propagation of top developers across past three years are also studied. The four algorithms used to find top influential developers gives almost similar results. The movement of top developers over past years was also consistent. Influential nodes in a network are very important to diffuse influence on the rest of the network. They are most often highly connected within the network. The existing algorithms are efficient to identify them. However, the challenge is in selecting a seed set that can spread the influence instantaneously with least effort. In this paper, a method is defined to spread influence on the entire network by selecting the least number of non-overlapping influential nodes faster than a well known existing algorithm. Further to this, the number of clusters in the network is also determined simultaneously from the seed set of the networks.

INTRODUCTION

Increase in the use of personal computers in late 1980s has encouraged much wider use of Social network analysis (SNA) methods because it has meant increased ability to manage large data sets and to visualize social network data in a wide variety of ways (Pan, 2007). Social network analysis is based on an assumption of the importance of relationships among interacting units. It is also used to study social relations among a set of actors as it concerns with the network structure formulation and solution (Borgatti, Mehra, Brass et al., 2009; Carrington, Scott & Wasserman, 2005). Such structures are usually captured in graphs. At organisations, collaboration in networks is critical to innovation. Paradoxically these networks are taken for granted, frequently invisible and rarely managed (Lim, Quercia & Finkelstein, 2010; Thomas, Valluri & Karlapalem, 2006).

Social network analysis is focused on uncovering the patterning of people's interaction. A rapid growth in work across organisation as well as geographic boundaries with outsourcing, off-shoring, virtual organisations and business process networks combined with trends such as the rise of blogs, online communities and social networking sites such as Friendster and LinkedIn as well as the rapid growth of collaborative software have all contributed to the emergence of SNA from the academic closet. The network's perspective encompasses theories, models, and applications that are expressed in terms of relational concepts or processes. Along with growing interest and increased use of network analysis has come a consensus about the central principles underlying the network perspective. The semantic web is an emerging concept that launches the idea of having data on the web defined and linked in a way that it can be used by people and processed by machines. The semantic web and social network models support each other (Jamali & Abolhassani, 2006).

Graph Data Mining in Social Network Analysis

Graphs have become very important in modelling complicated structures. Many domains produce data that can be intuitively represented by graphs. The process of discovering interesting facts and information about these graphs is difficult and challenging to work with, because real world data is very huge for any sort of raw interpretation. Graph mining which is a data mining approach has to be collaborated with SNA to obtain the results of the data sets so as to provide us with an insight of the frequent sub graphs of interest.

21 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/graph-mining-approaches-to-study-volunteer-relationships-in-sourceforgenet/193461

Related Content

A New Data Mining-Based Framework to Test Case Prioritization Using Software Defect Prediction

Emad Alsukhni, Ahmad A. Saifan and Hanadi Alawneh (2017). *International Journal of Open Source Software and Processes* (pp. 21-41).

www.irma-international.org/article/a-new-data-mining-based-framework-to-test-case-prioritization-using-software-defect-prediction/190482

Corpus Tools and Technology

(2020). *Computer Corpora and Open Source Software for Language Learning: Emerging Research and Opportunities* (pp. 22-43).

www.irma-international.org/chapter/corpus-tools-and-technology/256698

Open for Social: How Open Source Software for E-Learning can take a Turn to the Social

James Laffey, Matthew Schmidt and Christopher Amelung (2010). *International Journal of Open Source Software and Processes* (pp. 49-64).

www.irma-international.org/article/open-social-open-source-software/41953

On Solving the Multi-Objective Software Package Upgradability Problem

Nouredine Aribi and Yahia Lebbah (2018). *International Journal of Open Source Software and Processes* (pp. 18-38).

www.irma-international.org/article/on-solving-the-multi-objective-software-package-upgradability-problem/213932

Evaluating the Potential of Free and Open Source Software in the Developing World

Victor van Reijswoud and Emmanuel Mulo (2012). *International Journal of Open Source Software and Processes* (pp. 38-51).

www.irma-international.org/article/evaluating-the-potential-of-free-and-open-source-software-in-the-developing-world/101205