

Chapter 6

Finding Influential Nodes in Sourceforge.net Using Social Network Analysis

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.

INTRODUCTION

Social Network Analysis (SNA) has been an important tool for analyzing various domains of human behaviour. By bringing human interactions into graph structures they have helped gain new insights. Basic measures like betweenness, centrality, cohesion, reach etc can unveil lot of information about the modelled scenario. Concepts like ‘Small World Phenomenon’ are empirically verified across different cases. With the emergence of new computational tools there has been renewed interest in the application of Social Network Analysis into new domains. FOSS is primarily a global network of volunteers and thus makes for an ideal case to study using SNA

DOI: 10.4018/978-1-5225-3707-6.ch006

Tools. The rich research available in SNA can be helpful in understanding the complex phenomenon which makes FOSS work. The present work focuses on one aspect of SNA namely the influence maximization problem. The main objective here is to find a small set of most influential nodes in a social network so that their aggregated influence in the network is maximized. In the context of FOSS, the most influential nodes correspond to developers who are very well connected in FOSS ecology and therefore have maximum chance to propagate information in the network. Four well known algorithms namely High Degree, LDAG, SPS-CELF++ and SimPath are used to find top 5 influential developers. The propagation of top developers during past three years are also studied.

Early attempts of applying social network analysis to Free and Open Source Software phenomenon was undertaken by studying the properties such as degree distribution, diameter, cluster size and clustering coefficient. The emergence of the small world phenomenon in such networks was also studied (Xu, Christley, & Madey, 2006). The issue of finding the most influential nodes in social networks is treated in the literature as problem of centrality. In one of earliest comprehensive study on the centrality of social networks nine centrality measures were proposed. Among them three were based on degrees of points, next three on betweenness of points and last three on closeness. These different conceptions give three different views of centrality namely control, independence and activity (Freeman, 1979). The first natural greedy strategy solution which performed better than node selection heuristics based on notions of degree centrality and distance centrality continues to influence the studies in this domain (Kempe, Kleinberg & Tardos, 2003). A greedy algorithm for the target set selection problem where an initial set of nodes are selected to maximize the propagation in a social network was also proposed by same authors (Kempe, Kleinberg, & Tardos, 2005).

The later works have used variety of approaches to find top-k nodes in social networks. Finding top influential nodes using bond percolation method is tried (Kimura, Saito & Nakano, 2007). Using the metaphor of 'Outbreak Detection' the problem of information diffusion was discussed in different way. The novel solution was proposed as CELF algorithm which is reported to perform 700 times faster than simple greedy algorithm (Leskovec et al., 2007). Effective heuristics derived from the independent cascade model was comparable or better than greedy algorithms in finding top nodes for influence maximization problem (Chen, Wang & Yang, 2009). Most of the

11 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/finding-influential-nodes-in-sourcefor-genet-using-social-network-analysis/193460

Related Content

An Effective Approach to Test Suite Reduction and Fault Detection Using Data Mining Techniques

B. Subashini and D. Jeya Mala (2017). *International Journal of Open Source Software and Processes* (pp. 1-31).

www.irma-international.org/article/an-effective-approach-to-test-suite-reduction-and-fault-detection-using-data-mining-techniques/203646

Collaboration in Open Source Domains: A Perspective on Usability

Görkem Çetin and Mehmet Göktürk (2009). *International Journal of Open Source Software and Processes* (pp. 17-28).

www.irma-international.org/article/collaboration-open-source-domains/38903

The System for Population Kinetics: Open Source Software for Population Analysis

Paolo Vicini (2009). *International Journal of Open Source Software and Processes* (pp. 29-43).

www.irma-international.org/article/system-population-kinetics/38904

Data Mining User Activity in Free and Open Source Software (FOSS)/ Open Learning Management Systems

Owen McGrath (2010). *International Journal of Open Source Software and Processes* (pp. 65-75).

www.irma-international.org/article/data-mining-user-activity-free/41954

Exploratory Analysis of Free and Open Source Software Ecology

(2018). *Free and Open Source Software in Modern Data Science and Business Intelligence: Emerging Research and Opportunities* (pp. 93-103).

www.irma-international.org/chapter/exploratory-analysis-of-free-and-open-source-software-ecology/193459