Chapter 43
Social Networks Discovery Based on Information Retrieval Technologies and Bees Swarm Optimization:
Application to DBLP

Yassine Drias
USTHB, LRIA, Algeria

Habiba Drias
USTHB, LRIA, Algeria

ABSTRACT

Unlike the previous works where detecting communities is performed on large graphs, our approach considers textual documents for discovering potential social networks. More precisely, the aim of this paper is to extract social communities from a collection of documents and a query specifying the domain of interest that may link the group. We propose a methodology that develops an information retrieval system capable to generate the documents that are in relationship with any topic. The authors of these documents are linked together to constitute the social community around the given thematic. The search process in the information retrieval system is designed using BSO, the bee swarm optimization method in order to optimize the retrieval time for large amount of documents. Our approach was implemented and tested on CACM and DBLP and the time of building a social network is quasi instant.

INTRODUCTION

Social networks have emerged with the advent of the Internet and are nowadays of unquestionable importance. The main objective of this study is the development of a methodology for discovering social networks from a collection of documents. Information retrieval technologies are exploited to reach the target, which is to build social networks from textual documents. The system we developed generates a group of individuals who have the same center of interest, which is specified

in a query. Such cluster stimulates its members to constitute an effective social network that may provoke interaction between each other.

MOTIVATION

The problematic of searching authors 'who work on the same topic as I do' is one of the recent and wide phenomena that we observe these last years. Indeed researchers often try to build their proper social network by their own. Our contribution helps in that sense and offers a tool that is capable to yield such community in an automatic and efficace way.

METHODOLOGY

Of course, since the idea of dealing with such issue is based on information retrieval technologies, the system basis has to be implemented in an effective way. We know information retrieval distinguishes two main stages, indexing and retrieval. Indexing consists in translating the documents in a format that allows making a quick access to their words. Lex, the lexical analyzer generator most commonly spread, notably in compilers construction, is used for the implementation of the index and the dictionary.

For the design of documents retrieval, we use a robust method that can handle a large collection of documents and which is based on a bio-inspired approach called Bee Swarm Optimization (BSO) (Drias et al. 2005).

Finally, the query specifying for instance a subject of link or cooperation is treated by the system, which yields a set of references. Then a discriminating technique based on references characterictics is launched to analyse the results and keep the authors of the resulted documents that are apt to constitute a social net.

The whole system is implemented and tested on two benchmarks: CACM, which is a collec-

tion of articles published by ACM and DBLP, which is a database that contains bibliographic information on journals conference proceedings in computer science.

RELATED WORKS

Social networks mining has been widely studied this last decade and several works have reached important advancements in the area. Applications to domains like biology, economics and marketing have been also undertaken and the results were promising. Among the main issues that interest the researchers are the mining methods, community identification and modeling social rating networks. Lots of investments such those of (Clauset et al. 2004, Flake et al. 2000, Fortunato 2010, Newman & Girvan 2004, Radicchi et al. 2004) have been devoted to community identification. All of these articles build the social network from a given large graph and differ from each other by the method designed to extract the community structure. In (Flake et al. 2000), the authors focused especially on web communities and in (Domingos & Richardson 2001), the authors considered the marketing application. Other axes that were also investigated concern the analysis of the social network as in (Leskovec et al. 2008) and the scoring and evaluation of the social community as in (Newman & Girvan 2004, Domingos & Richardson 2001). The community scoring function quantifies how 'efficient' is the community.

On the other hand, information retrieval has known extremely interesting developments for more than four decades. The general concepts and techniques are well described in (Christopher et al. 2008, Rijsbergen 1979, Salton 1976).

Finally BSO, the bee swarm optimization approach is introduced in (Drias et al. 2005) and one of its applications to web information retrieval is published in (Drias et al. 2010).

Related Content

Should Investors Trust Equity Analysts?
Xiaomin Guo (2015). *International Journal of Business Analytics (pp. 45-61).*
www.irma-international.org/article/should-investors-trust-equity-analysts/126245

Should Investors Trust Equity Analysts?
Xiaomin Guo (2015). *International Journal of Business Analytics (pp. 45-61).*
www.irma-international.org/article/should-investors-trust-equity-analysts/126245

A Multimodal Based Approach for Face and Unique Mark Based Combination for Confirmation of Human
Prateek Srivastavaand Rohit Srivastava (2019). *International Journal of Business Analytics (pp. 16-28).*
www.irma-international.org/article/a-multimodal-based-approach-for-face-and-unique-mark-based-combination-for-confirmation-of-human/231514

Recognising Influences on Attitudes to Knowledge Sharing in a Research Establishment: An Interpretivist Investigation
Penny Hart (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1038-1058).*
www.irma-international.org/chapter/recognising-influences-on-attitudes-to-knowledge-sharing-in-a-research-establishment/142665

Reliability Analysis for Degrading Systems with 100% Quality Inspection after Burn-In
Hao Pengand Qianmei Feng (2014). *International Journal of Business Analytics (pp. 34-47).*
www.irma-international.org/article/reliability-analysis-for-degrading-systems-with-100-quality-inspection-after-burn-in/115519