Information Extraction to Improve Link Prediction in Scientific Social Networks

Victor Ströele, Fernanda Campos, Crystiam K. Pereira
PGCC – Postgraduate Program in Computer Science
UFJF – Federal University of Juiz de Fora
Juiz de Fora, Brazil
victor.stroele@ice.ufjf.br, fernanda.campos@ufjf.edu.br, crystiamkelle@gmail.com

Geraldo Zimbrão, Jano M. Souza
COPPE – Systems Engineering and Computer Program
UFRJ – Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
zimbrao@cos.ufrj.br, jano@cos.ufrj.br

***Abstract*— Link Prediction is a classic social networks analysis problem. Knowing in advance future actions in social network can help, for example, agents decision. Link Prediction techniques are based on metrics that have different approaches. In this paper, we model a multi-relational scientific social network to assess the impact of content extraction on topological metrics. Thus, a metric composed of topological and semantic approach is proposed to solve link prediction problem. The results were compared with those presented by Katz metric.**

Keywords—Information Extraction; Link Prediction; Scientific Social Network

I. INTRODUCTION

Thousands of people create billions of connections through social media every day, but few of them understand how each click builds relationships that, in overall, create a large social network where people share knowledge in a virtual environment [1, 2]. The result of these interactions is a complex network that links people to other people, documents, objects, places, concepts, and so on. This growing number of people virtually interacting has given rise to many social networks types and new tools are being developed in order to analyze these interactions.

A social network reflects a social structure that can be represented by individuals or organizations and their links. In general, links represent one or more types of interdependence (as idea and religion) or more specific relationships (such as knowledge exchange and friendship). So with this social structure, data and information exchange between individuals or organizations can be studied and analyzed at different detail levels.

Studies show that there is a trend of researchers from the same institution, or linked to related institutions, to collaborate more with each other [3]. Propose new relationships can help researchers to create new ties and improve links of their scientific social network.

In order to allow that relationships suggestions have influence of many scientific contact, we model a multirelational scientific social network based on a proposed model in previous works [4-6].

Relationships suggestion can speed up communication process and consequently improve scientific production of educational institutions. In this paper, new relationships suggestion will be made through *Link Prediction* techniques. Link prediction has always had great interest from researchers, because knowing in advance future actions in a social network can help agents decision of this social network, for example.

The main goal of this work is to improve relationship suggestion module developed in previous work, in which we used only information about scientific social network topology [5]. Proposed metric will be improved to incorporate semantic elements through information extraction techniques. Therefore, we will suggest new relationships based on both social network structure and researchers' interests.

Although previously obtained results have been satisfactory, new studies are showing that only structural information of social network are not enough to produce good predictions. In this work, we will extract semantic information to incorporate them into the structural information from social network. With these two information type, we aim to improve relationship prediction.

Paper organization: section 2 presents related works; in section 3 we show the multi-relational scientific social network model; section 4 studies the metrics; in section 5 we do metrics evaluation; in section 6 are the comparative analyses; in section 7 we present the final remarks.

II. RELATED WORKS

Identify changes and analyze future behavior in social network based on past information is called Link Prediction (Relationship Prediction). According to Liben-Nowell & Kleinberg [7] in this problem, given a social network snapshot, we seek to accurately predict the edges that will be added to network in the future.

Metrics definition are in table 1 [8]. The table uses certain symbols that are now defined: # denotes the number of elements in a set. U_n is the set of bi-directed links of the social network at time step n. A bi-directed link between nodes v_i and v_j is noted as $u_{i,j}$ and $u_{j,i}$. $\Gamma(v_i)$ is the set of

978-1-5090-1915-1/16/$31.00 ©2016 IEEE