Chapter VI
Social Network Models for Enhancing Reference-Based Search Engine Rankings

Nikolaos Korfiatis
Copenhagen Business School, Denmark

Miguel-Ángel Sicilia
University of Alcalá, Spain

Claudia Hess
University of Bamberg, Germany

Klaus Stein
University of Bamberg, Germany

Christoph Schlieder
University of Bamberg, Germany

ABSTRACT

In this chapter we discuss the integration of information retrieval information from two sources—a social network and a document reference network—for enhancing reference-based search engine rankings. In particular, current models of information retrieval are blind to the social context that surrounds information resources, thus they do not consider the trustworthiness of their authors when they present the query results to the users. Following this point we elaborate on the basic intuitions that highlight the contribution of the social context—as can be mined from social network positions for instance—into the improvement of the rankings provided in reference-based search engines. A review on ranking models in Web search engine retrieval along with social network metrics of importance such as prestige and centrality are provided as background. Then a presentation of recent research models that utilize both contexts is provided, along with a case study in the Internet-based encyclopedia Wikipedia, based on the social network metrics.
INTRODUCTION

Since the introduction of information technology, information retrieval (IR) has been an important branch of computer and information science, mainly due to the ability to reduce the time required by a user to gather contextualized information and knowledge (Baeza-Yates & Ribeiro-Neto, 1999). With the introduction of hypertext (Conklin, 1987), information retrieval methods and technologies have been able to increase their accuracy because of the high amount of meta-information available for the IR system to exploit. That is not only information about the documents per se, but information about their context and popularity. However the development of the World Wide Web has introduced another dimension to the IR domain by exposing the social aspect of information (Brown & Duguid, 2002) produced and consumed by humans in this information space.

Current ranking methods in information retrieval—which are used in Web search engines as well—exploit the references between information resources such as the hypertextual (hyperlinked) context of Web pages in order to determine the rank of a search result (Dhyani, Keong, & Bhowmick, 2002; Faloutsos, 1985). The well-known PageRank algorithm (Page, Brin, Motwani, & Winograd, 1998; Brin & Page, 1998) has proved to be a very effective paradigm for ranking the results of Web search algorithms. In the original PageRank algorithm, a single PageRank vector is computed, using the link structure of the Web, to capture the relative “importance” of Web pages, independent of any particular search query. Nonetheless, the assumptions of the original PageRank are biased towards measuring external characteristics. In fact, Page et al. (1998) conclude their article with the sentence: "The intuition behind PageRank is that it uses information which is external to the Web pages themselves—their backlinks, which provide a kind of peer review."

That is to say that backlinks (i.e., incoming links) are considered as a positive evaluation of the respective Web site. The PageRank therefore does not distinguish whether the user setting the link agrees with the content of the other Web page or whether she or he disagrees. This fact underlines that current reference-based ranking algorithms often do not take into consideration that an information resource is a result of cognitive and social processes. In addition to its surrounding hyperlinks, a social context (Brown & Duguid, 2002) underlies the referencing of those resources.

This suggests that a critical point in improving link-based metrics would be that of augmenting or weighting the pure backlink or reference model with social information, provided that linking is in many cases influenced by social ties, and trustworthiness critically depends on the social relevance or consideration of the authors of the pages. Recently, several independent researchers have provided different models for this kind of social network analysis as applied to ranking or assessing the quality of Web pages (Sicilia & Garcia, 2005; Hess, Stein, & Schlieder, 2006; Stein & Hess, 2006) or activity spaces such as Usenet and wikis (Korfiatis & Naeve, 2005). Other approaches such as those presented by Borner, Maru, and Goldstone (2004) have also dealt with integrating different networks by analyzing the simultaneous growth of coauthor and citation networks in time.

The main objective of this chapter is to provide a survey for recently proposed measures on a document reference network that integrate information from a second source: a social network. Therefore, we provide a bridge to the areas of social network analysis (SNA) and ranking methods applied in information retrieval.

To this end this chapter is organized as follows: The next section describes the basic intuition behind the concepts and the definitions provided by this chapter. We then provide an insight to the classic models of citation and link-based ranking