UNIVERSIDADE DA BEIRA INTERIOR
Faculdade de Engenharia

Enhancing E-learning Platforms with Social Networks Mining

Jorge Emanuel Frazão Costa

Dissertation submitted in candidature for the degree of Master of Science in
Informatics Engineering
(2nd Cycle Degree)

Supervisor by Prof. Doutor Joel José Puga Coelho Rodrigues

Covilhã, Outubro 2012

Acknowledgements

First of all, I would like to express my appreciation to my supervisor Professor Joel José Puga Coelho Rodrigues for his support, supervision, and encouragement.

My special thanks goes to Orlando Ricardo Esteves Pereira, for his support and exchange of ideas.

I am most grateful to the University of Beira Interior, *Instituto de Telecomunicações*, Covilhã Delegation, Next Generation Networks and Applications Group (NetGNA), Portugal, and the Project E-Learning 3.0 from PT-QREN I&ID/5625/2009, for many kinds of support that was given to me.

Also, my thanks to co-worker Tiago Simões for the entire support provided.

Last, my eternal gratitude to my parents, my mother Anabela Nunes Frazão Costa and my father Sérgio Moreira Costa for all the support, strength, and constant encouragements throughout these years of hard work and dedication.

My thanks to all the other friends not mentioned here.

Resumo

As redes sociais surgiram como um serviço Web com funcionalidades de criação de perfil, criação e interação de amigos. Estas redes evoluíram rapidamente e ganharam uma determinada importância na vida das pessoas. Agora, todos os dias, as pessoas usam as redes sociais para partilhar notícias, interesses e discutir temas que de alguma forma são importantes para elas.

Juntamente com as redes sociais, as plataformas de aprendizagem baseadas em tecnologias, conhecidas como plataformas E-learning têm evoluído muito nos últimos anos. Ambas as plataformas e tecnologias (redes sociais e E-learning) fornecem acesso a informações específicas e são capazes de redirecionar determinado conteúdo para um ou vários indivíduos (personalização).

O tema desta dissertação é motivado pela mineração do conteúdo das redes sociais em plataformas E-learning. Neste sentido, foram selecionadas quatro redes sociais, Facebook, Twitter, Google Plus, e Delicious para servir de estudo de caso à solução proposta. A fim de adquirir, analisar e concretizar uma aplicação correta e precisa dos dados, duas abordagens diferentes foram seguidas: enriquecimento de uma plataforma E-learning atual e melhoria dos motores de busca. A primeira abordagem propõe e elaboração de uma ferramenta de recomendação de documentos Web usando, como principal critério, a informação social para apoiar um sistema de gestão de aprendizagem (LMS). Desta forma, foram construídas três aplicações distintas, designadas por *Crawler*, *SocialRank* e *Recommender*. As informações extraídas serão incorporadas num sistema E-learning, tendo sido escolhida a PLEBOX (*Personal Learning Environment Box*). A PLEBOX é uma plataforma personalizada baseada numa interface inspirada nos sistemas operativos, fornecendo um conjunto de ferramentas (os conhecidos SDK - software development kit), para a criação e gestão de módulos. Dez

unidades curriculares foram avaliadas e os resultados do sistema de recomendação são apresentados.

A segunda abordagem apresenta uma proposta para melhorar um motor de busca com base no conteúdo das redes sociais. Subsequentemente, uma análise profunda é apresentada, justificando os procedimentos de avaliação, afim de criar o ranking de resultados (o *SocialRank*). Por último, os resultados são apresentados e validados em conjunto com um motor de busca. Assim, foi proposta, construída, demonstrada e avaliada uma solução para integrar e oferecer uma melhoria na ordenação de conteúdos Web dentro de um motor de busca. A solução está pronta para ser utilizada.

Abstract

Social Networks appeared as an Internet application that offers several tools to create a personal virtual profile, add other users as friends, and interact with them through messages. These networks quickly evolved and won particular importance in people lives. Now, everyday, people use social networks to share news, interests, and discuss topics that in some way are important to them.

Together with social networks, e-learning platforms and related technologies have evolved in the recent years. Both platforms and technologies (social networks and e-learning) enable access to specific information and are able to redirect specific content to an individual person.

This dissertation is motivated on social networks data mining over e-learning platforms. It considers the following four social networks: Facebook, Twitter, Google Plus, and Delicious. In order to acquire, analyze, and make a correct and precise implementation of data, two different approaches were followed: enhancement of a current e-learning platform and improvement of search engines. The first approach proposes and elaborates a recommendation tool for Web documents using, as main criterion, social information to support a custom Learning Management System (LMS). In order to create the proposed system, three distinct applications (the Crawler, the SocialRank, and the Recommender) were proposed. Such data will be then incorporated into an LMS system, such as the Personal Learning Environment Box (PLEBOX). PLEBOX is a custom platform based on operating systems layout, and also, provides a software development kit (SDK), a group of tools, to create and manage modules. The results of recommendation tool about ten course units are presented.

The second part presents an approach to improve a search engine based on social networks content. Subsequently, a depth analysis to justify the

abovementioned procedures in order to create the SocialRank is presented. Finally, the results are presented and validated together with a custom search engine. Then, a solution to integrate and offer an order improvement of Web contents in a search engine was proposed, created, demonstrated, and validated, and it is ready for use.

Keywords

Contents

Chapter 4

List of Figures

List of Tables

xx

Acronyms

LMS	Learning Management System
SN	Social Networks
API	Application Programming Interface
JSON	JavaScript Object Notation
REST	Representational State Transfer
E-Learning	Electronic Learning
RS	Recommendation Systems
CF	Collaborative Filtering
PLEBOX	Personal Learning Environment Box
PAMS	Personalized Annotation Management System
SR	SocialRank
WS	Web Services
WWW	World Wide Web
URL	Uniform Resource Locator
CU	Course Units
G+	Google Plus
DCR	Dynamic Competitive Recommendation
U.S.	United States
HTTP	Hypertext Transfer Protocol
XML	Extensive Markup Language
RSS	Really Simple Syndication
ATOM	Advanced Tool for Option Modeling
ID	Identifier
SEO	Search Engine Optimization
SQL	Structured Query Language
SDK	Software Development Kit

Chapter 1

Introduction

Social networks offer Web places where people connect to each other. Inside these networks, there are different relations types, such as friendship, co-working, or information exchange [1]. People just need to have a direct or indirect interest in order to create a relation. In general, exploring these networks during a few minutes is possible to find several different groups about our interests, taking people to participate in informal learning and creation of digital literacy [2]. Over there, the use of social networks brings a new way to access and receive information of any interest, replacing several services as RSS. Unifying in one place the personal and professional life. So, custom Web pages visited everyday like newspapers, blogs, or groups are represented on social networks in a single page customized by each user. Newspapers social pages, blogs social pages shared articles and what they wrote to disseminate information and knowledge, and like these pages, people do the same, turning social networks interesting and useful. Currently, several studies have been done showing the growth of social networks in people lives, as is presented in Figure 1 of the Chapter 3.

E-learning technologies as grow and several authors study the combination of these two widely used technologies [3, 4]. The student learning process has modified, and its concept as no longer confined in classrooms with lectures and teachers, or even on libraries. With the contribution of the vast technologies, the method of learn was increased, and expanded, being also an important role in distance education [4]. E-learning through learning management system (LMS) platforms have been supported for learning management. Moodle and Blackboard are great platforms examples for the creation of education objects, distribution of study content, implementation

of learning methods, communication between participants of the educational process, and management of studies [5].

Recommendation tools are being implemented in all kind of applications and until now, are considered one great way to efficiently filter out the overload information, or redirect the user to one product or objective. Online companies have been the main users of these recommenders, leading the user to products that may be interested [6]. However, in these cases the recommendations are generic. In contrast, personalized recommender systems try to achieve the gold standard of recommendations, takes the individual tastes and preferences of users into account [7]. In this case, our approach takes the user to Web documents that may be important to improve their learning, taking to account your personal and curricular information on the LMS account. The vision of this dissertation includes the exploration and bringing social networks information to enhance e-learning platforms.

1   Motivation

The main interest for this research work is the fact that is included in an area of huge interest to the scientific and learning community. Technologies such as social networks and E-Learning became a part of our daily life, whether for personal or professional use.

Beyond, social networks can be part of the key on evaluation of online resources discussed over the years. As instance of that is the vision of Jon Kleinberg, "The quality of a search method necessarily requires human evaluation, due to the subjectivity inherent in notions such as relevance" [8]. The human evaluation can tell to machines what they can consider important and relevant in order to display the results according by the users needs, recommending.

2

The E-Learning research needs constant improvement and new tools for students to increase their knowledge, collaboration, and share. More, with the massive data available people spends much time what they really want to learn.

Last, it does not exist an implementation of such application as described in this dissertation in the literature, so this becomes a major motivation.

2 Problem Definition and Objectives

This dissertation tries to demonstrate how data mining of social networks can be used on E-Learning technologies to produce a basic but useful tool in support of students' knowledge. Currently, several people spend lots of time on social networks and create, share content about their personal and professional life. It changed and created a new model of evaluation through a new and major range of people. People are still only willing to look at the first few ten results [9, 10]. Then, it is necessary to adjust the precision and the relevance criterion to the actual reality. The main goal is to improve the student's support using an LMS.

To reach this main objective the following intermediate objectives were identified:
- Review of the related literature
- Study about social networks and selection the applications for the case study
  - Analysis and exploration of social networks
  - Analysis of shared content on social networks
- Construction of a custom software platform
  - Requirements analysis
  - Application design
  - Construction of a crawler to data mining;
  - Creation and deployment of a Recommender system;

- o Construction and deployment of a several Web Services;
- o Construction of a criterion based on social networks information to establish a Rank;
- o Improve a collaborative tag system helping the documents cataloging.
- Demonstration and validation of the proposed solution

This work of research and engineering work is expected to produce the desired Recommender, but also to provide the opportunity to disseminate the knowledge and software through, al least, a journal paper.

3   Main Contributions

This section is devoted to the main scientific contributions of this dissertation to the social networks data mining, recommendation on e-learning platforms and improving search quality.

The first contribution of this dissertation is a recommendation tool based on social networks content to incorporate on a learning management system (LMS) and which and how the social information can be used. It analyses all the collected data and return the best suggestions to the user based on her course units and hobbies. Further, the system and its layers were demonstrated and evaluated. The contribution was submitted to an international journal.

The second contribution presents a study over four social networks (Facebook, Twitter, Google Plus, and Delicious), about what is possible to obtain through their services. It proposes and explains a specific algorithm, called SocialRank, and a new criterion of relevance for Web resources on the World Wide Web. In order to evaluate and validate the proposed model, a study was performed and results were analyzed. The contribution was submitted to an international journal.

# 4	Dissertation Organization

This dissertation is organized in four chapters that are organized as follows. This chapter, the first, presents the context of the dissertation, focusing the topics under study, the motivation, the problem definition of the main objectives, its main contributions, and the dissertation organization.

The chapter 2, recommendation tool based on social networks content, focuses on available social networks content to help everyday the students life and their experiences on e-learning field, through a recommender.

The chapter 3, improve hypertext results exploring social networks, presents a deep study over the social network information and how this information can influence the best documents on the Web. Next, it proposes and explains a social algorithm, called SocialRank, also used on the above-mentioned Recommender, in order to improve the search results. The results are analyzed and compared with Google search engine results.

The chapter 4, Conclusions and future work, elaborates a final summary and few more remarks about the dissertation and suggests further research works.

5 References

[1]	W. Xin, Q. Jamaliding, and T. Okamoto, "Discovering Social Network to Improve Recommender System for Group Learning Support," in *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on*, 2009, pp. 1-4.

[2]	I. Hitrec, I. Pogarcic, and S. Suman, "eLearning: A social software in higher education learning," in *MIPRO, 2011 Proceedings of the 34th International Convention*, 2011, pp. 1207-1212.

[3]   W. Treepuech, "The application of using social networking Sites with available online tools for teaching and learning management," in *IT in Medicine and Education (ITME), 2011 International Symposium on*, 2011, pp. 326-330.

[4]   L. Min and L. Zhengjie, "The Role of Online Social Networks in Students' E-Learning Experiences," in *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on*, 2009, pp. 1-4.

[5]   J. Kapounova and K. Kostolanyova, "Specific tool for evaluation of eLearning," in *Interactive Collaborative Learning (ICL), 2011 14th International Conference on*, 2011, pp. 423-427.

[6]   A. Klašnja-Milićević, B. Vesin, M. Ivanović, and Z. Budimac, "E-Learning personalization based on hybrid recommendation strategy and learning style identification," *Computers & Education*, vol. 56, pp. 885-899, 2011.

[7]   #252, r. Buder, and C. Schwind, "Learning with personalized recommender systems: A psychological view," *Comput. Hum. Behav.*, vol. 28, pp. 207-216, 2012.

[8]   J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, pp. 604-632, 1999.

[9]   A. Spink, B. J. Jansen, C. Blakely, and S. Koshman, "A study of results overlap and uniqueness among major Web search engines," *Information Processing & Management*, vol. 42, pp. 1379-1391, 2006.

[10]   N. Höchstötter and D. Lewandowski, "What users see – Structures in search engine results pages," *Information Sciences*, vol. 179, pp. 1796-1812, 2009.

Chapter 2

Recommendation Tool for Learning Management System Based on Social Networks Content

This chapter consists of the following article:

Recommendation Tool for Learning Management Systems Based on Social Networks Content

Jorge E. F. Costa, Joel J. P. C. Rodrigues, Orlando R. E. Pereira, Tiago M. C. Simões

Recommendation Tool for Learning Management Systems Based on Social Networks Content

Jorge E. F. Costa · Joel J. P. C. Rodrigues · Orlando R. E. Pereira · Tiago M.C. Simões

Abstract Social networks are constantly growing in terms of number of users, contents, and different applications. Everyday, several hundreds of new users join them and its Internet share increase significantly. Data mining of social networks is useful since new and important content is added in a daily basis. This work focuses on the aggregation of information retrieved from social networks and provides accurate results to a specific user request given his/her characteristics, curricular units or hobbies. Then, the most useful data will be used to support a custom learning management system. This paper proposes a recommendation tool based on several social networks content and explains which and how the social information can be used. It analyses all the collected data and return the best suggestions to the user. The system and its layers were experimented and evaluated, the results are presented and analyzed, and it is ready for use.

Jorge E. F. Costa
University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
E-mail: jorge.costa@it.ubi.pt

Joel J. P. C. Rodrigues
Instituto de Telecomunicações, University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
Tel.: +351-275-242081
Fax: +351-275-319899
E-mail: joeljr@ieee.org

Orlando R. E. Pereira
University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
E-mail: opereira@ubi.pt

Tiago M.C. Simões
University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
E-mail: tiago.simoes@it.ubi.pt

1 Introduction

The electronic learning (e-Learning) has evolved in the recent years. Thus, its concept and application plays an important role in distance education. E-Learning offers an educational environment which uses information and communication technologies to achieve the education goal, including the creation of educational objects, content distribution, implementation of learning methods, communication among participants in the educational process, and studies management [10]. In order to support both communication and information, several learning management systems (LMS) platforms are used. Platforms such as Moodle and Blackboard have a significant role in distance learning since they provide resources for course management and catalog, creation of quizzes and questions, user management and exercises management [10, 18].

Currently, e-learning technologies enable users to continuously capture, share, and manage knowledge and skills on organizations, schools, and universities. Such technologies enable access to specific information and redirect it to an individual person when needed. A good example of information exchange and information redirect is the social networks. Due to the fact of its increased utilization, its contents also increase in diversity and more contents arrive to distinct areas, such as, music, business, photography, or research issues. Social networks are also a great tool to disseminate Websites and Blogs. Currently, the most popular websites such

as Flickr, YouTube, and Facebook are totally connected to social networks [23].

Together with social networks, recommendation systems (RS) are being created for all kind of applications and are being considered the best way to filter out the overloaded information efficiently. Several Web-based companies, like Amazon and eBay, are using RSs to redirect users attention to certain products, making a pre-selection of contents that user might be interested [11]. However, so far, collaborative filtering (CF), is the most successful technique to receive recommendations where a user may receive specific information taking into account its personal information and the information of other users with similar characteristics [20]. Then, this paper proposes an innovate tool to retrieve custom learning data through social networks. Such data will than be incorporated into an LMS system, like the Personal Learning Environment Box (PLEBOX). The PLEBOX solution is a custom platform similar to available operating systems, based on personal learning environments and rich Internet applications technologies that provide a better learning environment for learners. PLEBOX programmers have a group of tools to create learning and management modules that can be installed on the platform, thus, enabling the possibility to include the proposed tool [19]. The main contributions of this paper are the following:

- A recommendation tool based on social networks content to incorporate on a LMS and which and how the social information can be used. It analyses all the collected data and return the best suggestions to the user based on her course units and hobbies.
- Proposes and explains a specific algorithm called SocialRank, a new criterion of relevance for Web resources on the World Wide Web

1.1 Motivation

Nowadays, social networks represent a wide online community offering social connections between users, which could share common interest. Given the social networks growth and characteristics, more and more students communicate with each other through Facebook or Twitter [9]. However, social networks are no longer simple social spaces for making friends, and are evolving towards a new type of information network [27]. Nielsen presented a study about Social Media (Q3 2011) and shows that social networks and Blogs occupy the global surfing Internet time of Americans. This time is more than the double amount of time spent on Internet with other interests like online games.

Social networks offer virtual places where students can gain social and communication skills, while they participate in informal learning, creativity development, and digital literacy. Digital literacy has growing and everyday, lots of content (news, posts, comments, etc.) may be seen in users social wall, performed by others users on the network. This means that all the published contents were evaluated by the poster and, lately, possibly by friends, subscribers or/and followers. Through that evaluation (number of likes/shares/saves/positive words/etc.) is possible assign the content as interest, usefulness, or the reverse. Then, the main motivation for this work comes from the need to retrieve and identify the useful content available in several social networks for users (learners and teachers) of a given eLearning platform.

1.2 Social Networks

In social sciences, a social network comprises a set of people or groups of people (actors) and their interactions (ties) [16]. The representation of a social network is usually given in the mathematical form of a graph $G = (V, E)$, where the set of nodes V means the set of actors and the set of edges $E \subseteq V \times V$ contains the relations between them [7]. Usually, these sites offer services like list of friends, messages, events, management and media uploads. The future of social networks in the Web is promising [2]. There are many available hundreds social networks. Then, for this work was selected the social networks listed bellow given their popularity in terms of number of users (http://www.nielsen.com) and interest for a learning community.

Facebook is the most popular social network since millions of persons using it every day. Its simplicity coupled with its wide range of technologies supported (newspapers, blog posts, YouTube videos, Web pages) contributed for its popularity. One of the Facebook features is the ability to group persons, topics, interests, or even schools, and thus receive notifications only about a specific topic with a chance to contribute.

Google+ or Google Plus is one of the most recent social networks. It is not much different from Facebook in terms of usability. This social network emphasizes more on sharing content with friends circles while Facebook emphasizes on sharing content with everybody.

Twitter is a minimalist social network that relies on small and focused messages (with a maximum of 140 characters) about a specific subject. Twitter is unique due to its characteristics: short descriptions, short hyperlinks, and hash tags. Through hash tags is possible to enumerate a specific topic or person.

Delicious is a different social network since it relies on the users bookmarks. It can be described as a social bookmarking service. Users of Delicious can share bookmarks, create custom tags, and share them among the network users. Thus, Delicious is a tool to save the important links found on the Internet and have been saved by others users [22].

The rest of paper is organized as follows. Section 2 elaborates on the state-of-the-art while Section 3 proposes the recommendation system for social networks, addressing its main characteristics and implementation issues. Section 4 presents the result analysis and explanations of the proposed system. Finally, main conclusions and suggestions for further works are considered in Section 5.

2 Related Work

In recent years, organizations and sciences researches have explored social networks [25]. The information and services are used of diverse and different ways, as generate gains, entertaining, games, or to better know their users. Through that, these organizations can improve the user experience in several systems and applications, which are included areas as electronic learning (e-learning) or recommendation systems. This chapter addresses the actual state of where and how social networks information is used, related with the main characteristics of the work proposed.

2.1 Recommendation with social networks information

Getting good recommendations becomes an important issue when the number of viable options is too large to be read by an individual person. Internet servers provide access to vast amount of information, and consequently, offering recommendations is one of the most pressing challenges. It can be said that search engines provide recommendations, as a list of search results is ordered through link analysis algorithms [3]. Several projects are proposed in recent years with the purpose of data gathering and mining on social media content in order to recommends. So, this section will review some available projects on this topic.

In Guy, et al. [6] is proposed a custom software that recommends three types of information: bookmarked Web pages, blog entries, and communities. This recommendation system is based on the page-hit number and the total time that resource is being used. It calculates the resource specifications and characteristics and provides to users additional similar links. An approach

to increase recommendation effectiveness incorporating social network information on collaborative filtering was proposed by Liua and Leeb [14]. They collected data about users preferences ratings and their social network relationships and evaluated the performance of collaborative filtering on diverse neighbor groups. In Wasim, Shahzadi, Ahmad, and Mahmood [24], an ontology model that relies on Twitter to cluster users with similar interests is proposed. The model was built using concepts from the Wikipedia and Wordnet Webpages. Seok Jong Yu purposed an interesting recommendation algorithm using social network services; It relies on Twitter data and compare the recommendation performance of existing algorithms (PageRank). The study concluded that there is no single optimal recommendation algorithm that satisfies the diversities of users and this can be a limitation of a single algorithm approach. The author evaluated the impact of the dynamic combination of multiple algorithms and confirmed that Dynamic competitive recommendation (DCR) algorithm is more stable and consistent than traditional methods [27].

2.2 Collaborative tagging

Marking content with descriptive terms, also called keywords or tags, is a common way of organizing content for future navigation, filtering or search. Collaborative tagging is the practice of allowing anyone, especially consumer, to freely attach keywords or tags to content [5]. Collaborative tagging systems have become valuable tools for sharing and exploring content, where tag-item associations can be aggregated over thousands or even millions of users. In [21] is designed and accomplished a study of a personalized annotation management system (PAMS 2.0) for managing, sharing, and reusing individual and collaborative annotations. Through PAMS 2.0 the authors investigate the effects of different annotation sharing scenarios on quantity of annotation and its influence on learning achievements. The results show that annotation process on learning achievements becomes effective and the sharing mechanism is positive for the majority of students. In [5] the authors elaborated an study of the collaborative tagging system Delicious, where was able to discover regularities in user activity, tag frequencies, kings of tags used and bursts of popularity in bookmarking. After, the authors present a dynamic model of collaborative tagging that predicts these stable patterns and relates them to imitation and shared knowledge. Noll and Meinel [17] proposed a new approach to personalized Web searches based on collaboration and information sharing about Web documents. The proposal use social

bookmarking and tagging to re-rank the Web search results.

2.3 Social Networks mining

In literature, are found several studies and systems mining the social networks information. In these studies several approaches are presented, as instance TunkRank [26], a tool for measuring the influence of user on Twitter network, which calculates how much attention the followers actually give to their following person. In [8] is analyzed the social tags through the social network Delicious to show how social bookmarking services can enhance Web searches. In [12], the authors did the first quantitative study inside a big collection of information over Twitter. They begin with the network analysis and studied the distributions of followers and followings, the relation between followers and tweets. Next they rank users by the number of followers, PageRank, and the number of retweets, present a quantitative comparison among them. In [15], the authors propose a new social network extraction system called POLYPHONET. It is capable of to extract relations of persons, to detect groups of persons, and to obtain keywords for a person. Furthermore, they developed several new algorithms for social network mining such as those to classify relations into categories, and to make extraction scalable. Finally, a novel architecture called Iterative Social Network Mining is proposed.

The solution proposed on this paper concatenates and improves the different abovementioned features. It provides reliable, robust, dynamic, and quick data mining over the social networks considered in the study. It creates an "importance" for every Web resource, recommend the best resource (to a given user specification), organize the results by a given tag and include rich interfaces for social tagging storing and cataloging.

3 System architecture

The proposed system can be described as a tag-aware collaborative recommendation system for eLearning environments. It focuses on collaborative filtering techniques using tags to model user interests and contents. The system architecture is based on three main layers. Each one has distinct specifications, cooperating among them for a unique goal. It includes the Crawler, the SocialRank, and the Recommendation layer. The Crawler is responsible for data querying and maintaining the documents found in social networks updated. SocialRank is the second layer and is responsible for assigning an importance to all the documents on the

repository. The third layer, the Recommender, is responsible for study, evaluate, recommend, and present the social contents to the user. Such recommendation is based on the user profile and it uses a Web interface to present them.

Figure 1 presents the different paths to achieve results. There are two paths where the information can pass before arriving to the user: an ordinary search: or by recommendation, the focus of this paper. However, the selection result process had an identical behavior. The system was thought to being a support tool for students to gather relevant information for them. In order to achieve this goal, the architecture was proposed taking into account the content recommendation based on the current learning process of the user. Then, a list of keywords per course unit or hobbies is created to help the filter between the user and the repository. Furthermore, the tags establish the connection on recommendation and they are used when the user wants save a document to himself/herself. At this point, a tool based on social and collaborative tagging systems were improved. Social tagging is an innovative and powerful mechanism introduced with Web 2.0 technologies that allow users to freely associate tags to resources, forming the so-called folksonomies [4]. Folksonomy can be seen as a structure that establishes a relation between a set of users, a set of resources, and a set of tags [13]. Every tag is a term chosen by a user with some personal meaning about the resource. This way, tags contain very useful information and can be used to summarize the content of the resource, identifying the type of document, expressing opinions, emotions, qualitative judges, or even to associate people to it [4]. However, social tagging are really interesting because enable users to receive suggestions or recommended tags from other users that save determined content, making the process of tagging more fast, accurate, and constant, such as the Delicious approach. In this system, users can cooperate among them without knowing. For that, the Delicious tags used by users and the meta-tags found on documents are saved, and a top ten ranking to retrieve the most used tags is created. Next, the three layers of the proposed system are described in detail.

3.1 Crawler layer

Crawlers (also known as robots or spiders) are computer programs to browse the World Wide Web (WWW or, simply, Web). Particularly search engines used them and their task can focus on specific subjects or just browsing off and download Web documents. Usually, to start a crawler, a set of Web pages (seed pages) is given
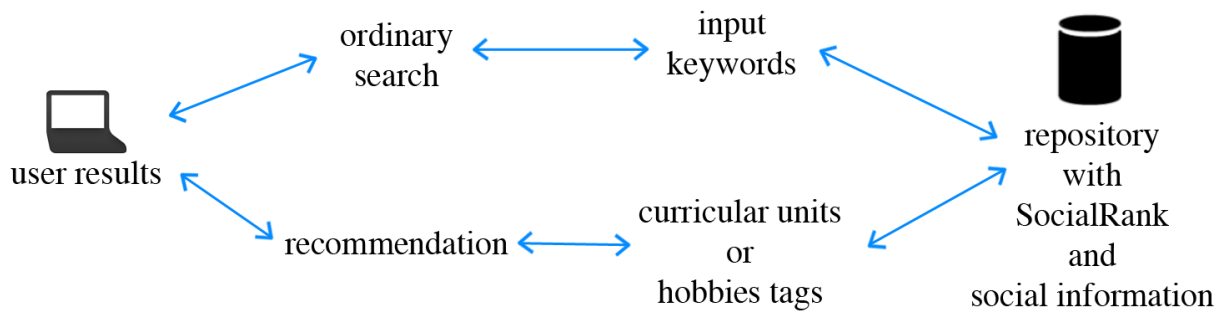
Fig. 1 Distinct paths to obtain the user results

as input. After, with the extracted links, it is determined what links should be visit next based on certain criteria (e.g. not was considered spam, points to one page about the same topic, etc.). Web pages pointed by these links are downloaded and those satisfying certain relevance criteria are stored in a local repository. Moreover, crawlers make an exhausted search and, typically, they try to keep the retrieved information up-to-date [1]. The current deployment the social crawler was created in the Java programming language and uses MySQL to manage all the relevant data. Unlike most of the available crawlers, the aforementioned crawler is fed through a set of Web services (WS) from the considered social networks, searching a specific set of tags each time that service is called.

The use of WS brings many advantages and almost all the social networks offer services to provide means to manage and interact with their content. Several advantages of WS are to make the platform and technology independent, allowing different applications to communicate to each other, and sharing data and services among themselves. They use standardized protocols for communication and low cost of communications.

In order to save resources, the crawler makes queries using a particular method. This method can merge teachers tags, students tags and/or learning tags generating different combinations. (Learning tags could be tags like learning, learn, practice, example, how, among others). Through this method, the evaluation becomes more effective, and the result more filtered. The social crawler architecture is shown in Figure 2.

As may be seen in Figure 2, the process of crawling start in the Queuer. It communicates with the Repository that contains the course unit keywords to be used on the requests to Social Network Servers. If the response contains any data, it will be filtered in the Download and Parser module and only the posts with hyperlinks will be saved in repository. Together with, it will be added to the Scheduler to make a resume about the

hyperlink on all the considered social networks (Facebook, Google+, Twitter, and Delicious). Updater module is not included in the main cycle but it is also important. It will maintain the information up-to-date, and keeps the SocialRank most precisely. Next, this process is described more detail.

The process of context extraction and evaluation leads to a deep social knowledge about a specific resource, including the following:

- Who shared;
- How many people shared;
- How many people liked;
- What was the description;
- Comments.

Distinct of the general social networks, Delicious is a collaborative tagging service. Through it, bookmarks with user-tags and distinct descriptions about one resource can be found. However, not all-important information is social information. The process of Download and Parser will analyze the custom information saved by search engines and goes to the repository as title, description, and meta-tags. As above-mentioned, the Scheduler forces the crawler to look for social information about the domain of every URL as pages, logos, and the above-mentioned data. The Updater has the function to maintain the data updated.

3.2 SocialRank layer

The act of sharing or saving a document means that it has some importance to the user. Plus, between the both actions, the share action can be seen with more significance because the person who shared it thinks that the resource is useful to his/her friends or followers. In order to perceive and rank the shared data in social networks a robust, fast, and reliable algorithm was created. After a deep analyzes, authors concluded that Web documents tend to be uniformly distributed on
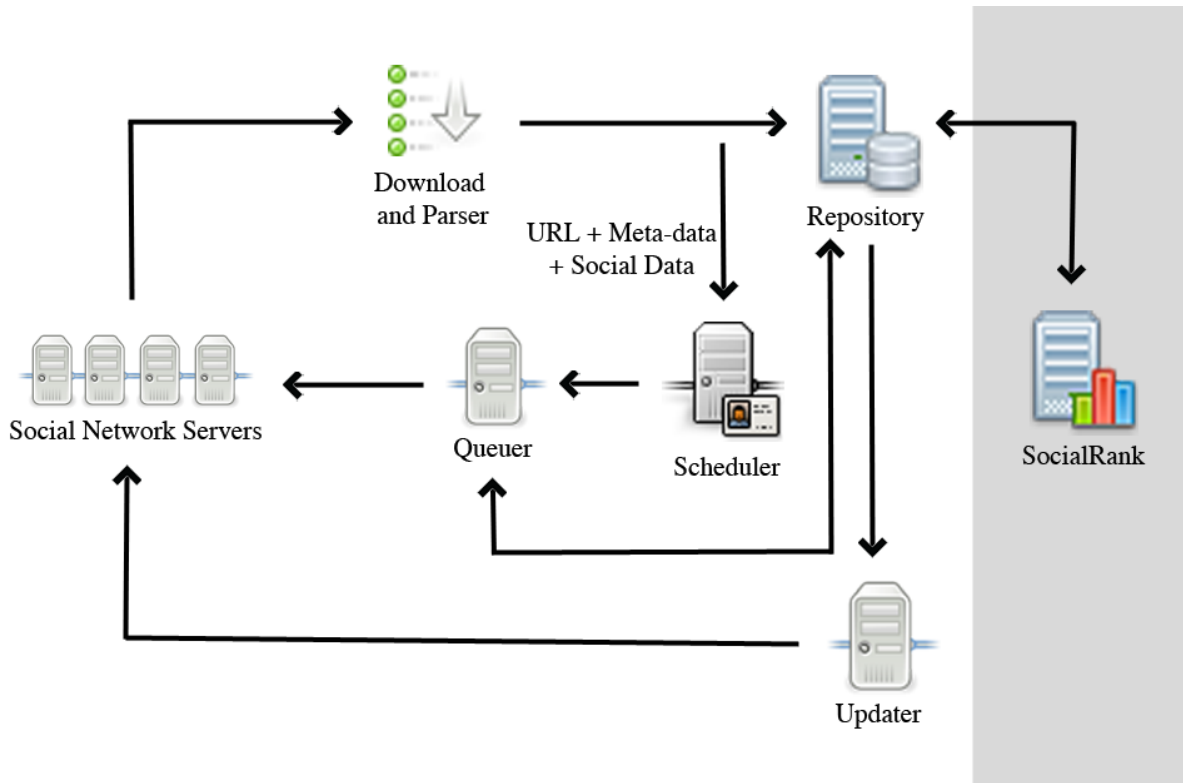
Fig. 2 Data Flow of Social Crawler

social networks. Although, there are several exception scenarios where this does not occur. Several hyperlinks have an exceeded number of social references in one social network (typically on Twitter) when compared with other social networks. Usually, this phenomenon occurs when people need accessing to a certain resource but, in order to access it, they must share it on a specific social network. It is common since instead of the payment being made with cash, it is done with a simple share action.

In order to create an accurate and realistic algorithm the authors created a sub-algorithm for data standardization and unification. The standard deviation (represented by σ) is calculated in order to unify the final results. It expresses the existent variation or dispersion in the average value. So, a low standard deviation indicates that social references tend to be very close to the mean. Equation 1 calculates the SocialRank (SR) of a given resource (x).

$$SR(x) = \frac{\sum r}{t} . \frac{10^{N(log10(\sigma.2))}}{\sigma.2} \qquad (1)$$

Where,
$\sum \mathbf{r}$: Sum of all the references encountered in the considered social networks.
t: Number of social networks considered on the system; currently it includes four SN (Facebook, Google+,

Twitter, and Delicious)
$N(log10(\sigma.2))$: Calculates the integer part (N) of figures on standard derivation multiplied by two.
σ : Standard derivation

3.3 Recommender layer

The recommender is the third layer of the proposed system. It is built as a Web application and follows the development guidelines of the three-tier architecture from Microsoft. Important aspects, such as system robustness, user-friendliness, and clean presentation were considered on its development. In terms of user-friendliness it was achieved by using a simple but intuitive users interface with appropriate icons and few text. Icons can summarize very easily information replacing text in simple and small images.

The recommender layer is quite simple and is primarily based on SocialRank. Currently, it searches on the properties described below, and returns the documents that matches the criteria sorted by top ranking given by the equation 2.

Figure 3 shows a screenshot of the recommender layout. In 1 a filter and its results is shown. During a search, it is possible to filter by social network (if
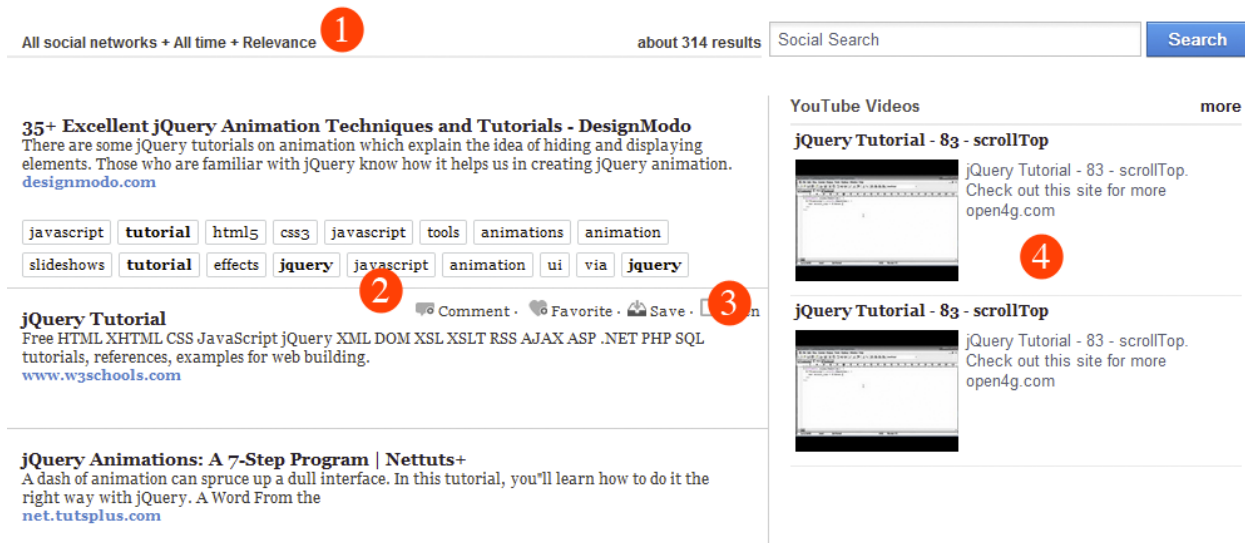
Fig. 3 Recommender results screenshot

the content is available), by statistics, relevance, view count, among others. At 2 some documents found are presented. For each document several information is displayed, such as title, host, description, and social tags. The zone marked with 3 shows the options to user, allowing mark as favorite, comment, or save again and rewrite new tags to the document. Finally, at 4 some videos from YouTube are shown.

Next, is depicted the way in which Social Rank works and its data workflow. Taking in account a simple scenario where the user has a set of Course Units (CU) and the teacher describes it through keywords. These keywords are directly related to the CU and contain information about what will be taught. The Recommender layer will connect the curricular information tags with the hyperlinks found on the social networks implemented and retrieves its information. Then it uses the document importance received from SocialRank.

Some properties used by Recommender:

– SocialRank;
– Number of keywords on a Web document (Title, URL, Description, Meta Tags, Tags on Delicious);
– Positive Words (Found on Posts, Twits or Comments);
– Learning Words;

Positive words, as above-mentioned, is a new concept that grow with social networks on the web, where through them people express emotions, evaluations, or opinions and words like "excellent", "love", "nice", are common in social resources.

Actually, the recommender implements the next formula. They give 50 percent weight to properties like number of keywords, positive words, learning words and the 50 percent left to SocialRank. We normalized the values of all the attributes to real numbers between 0 and 1 (inclusive). For each attribute, we found the maximum (maxPr) and minimum (minPr) values for each property (number of keywords, positive words, and learning words) and then normalized all the properties values.

Each property will have a weight based on he himself and other resources resulted by the query search using the CU keywords. The social content weight will be the sum of all the properties weights with the following equation:

$$R(x) = 0.5 \sum \frac{Pr(x) - minPr}{maxPr - minPr} + 0.5 \frac{SR(x) - min(SR)}{max(SR) - min(SR)} (2)$$

3.4 Recommender in Practice

In practice, the recommender shows popular Web pages that people like a lot, saving in their bookmarks (Delicious) or otherwise sharing on their social wall in Facebook, Twitter or Google Plus (G+).

Also, recommender allows, students/teachers to receive the suggestions or explores the best/recent contents, with the possibility of sharing the own library with anyone, being a fantastic tool to improve and encourage gain of knowledge.

4 Performance evaluation and analysis

To display the amount of resources shared on Social Media, we have chosen 10 course units of different areas and during 5 days we collected data through specific

keywords related with the CU and during these days, the repository recorded 163.933 distinct Web pages.

The follow list present the course units considered:

1. Internet Technologies
2. Programming I
3. Math
4. Chemistry and Physical
5. Geography
6. Mans History
7. Natural Sciences
8. Databases
9. Mobile Programming
10. Oriented Object Programming

Figure 4 presents a chart describing the amount of results per course unit and the percentage of data found on the different social networks.

Through the chart above, we can analyze that G+ and YouTube are the social networks where is found most documents, however, the results can be risky. As described before, the crawler searched during 5 days, but it has faced several restrictions from several social networks such as Facebook, Twitter, and Delicious. These networks do not provide full access to all publications about a specific topic search. Such social networks have a restricted time line of the last 3 weeks and all older results are not presented through Web Services. As expected not every course units have a lot of results, and in the social networks exist several topics more referenced that others.

Figure 5 presents the distribution of retrieved content of the aforementioned social networks. As may be seen, the small size of the Facebook slice may be surprising, being currently the most used social network. However, as above referenced, Facebook limits their services. It possibly catches much information, but unlike Google services, Facebook services need to be looked faster and everyday about what we need recommend.

Is show below a real scenario when the keyword jquery is used in the SocialRank. Jquery is a JavaScript Library that simplifies HTML document traversing, event handling, animating, and Ajax interactions for rapid web development. As abovementioned our repository is small (less than two hundred thousand). Choosing one of the tags already searched by Crawler, we obtained the results presented on Chart 3 (Figure 6). The YY axes present the number of social references of the Web document for each social network. Category XX axes contains the hyperlink identifier order by SocialRank. The last chart bar of each identifier is the SocialRank result. This identifier liaises with the listing following the Chart 3 (Figure 6).

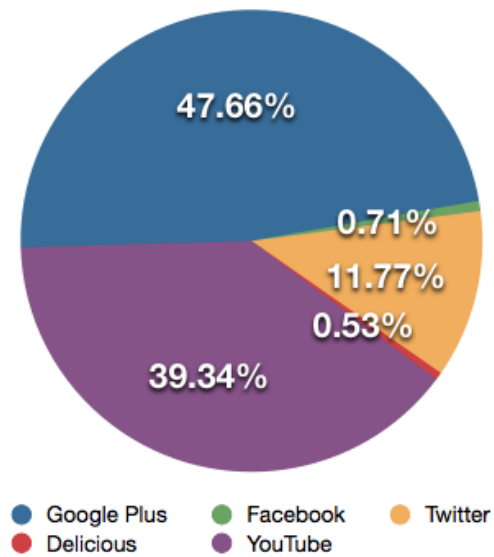Taking into account the jquery keyword, the top ten SocialRank (SR) links are the following:

Fig. 5 Data percentage on repository by Social Network

1. jQuery: The Write Less, Do More, JavaScript Library http://jquery.com/ - SR: 5846
2. Create a new Fiddle jsFiddle http://jsfiddle.net/ - SR: 5735
3. jQuery Mobile — jQuery Mobile http://jquerymobile.com/ - SR: 927
4. jQuery UI Home http://jqueryui.com/ - SR: 875
5. PHP Scripts, WordPress Plugins, HTML5, jQuery, and CSS — CodeCanyon http://codecanyon.net/ - SR: 768
6. JQuery Cycle Plugin http://jquery.malsup.com/cycle/ - SR: 695
7. Nivo Slider - The Worlds Most Popular jQuery - WordPress Image Slider - http://nivo.dev7studios.com/ - SR: 675
8. Isotope - http://isotope.metafizzy.co/ - SR: 670
9. Turn.js - The page flip effect for HTML5 - http://www.turnjs.com/ - SR: 668
10. jQuery google api and other google hosted javascript libraries. - ScriptSrc.net - http://scriptsrc.net/ - SR: 662

Note that using the Google search engine to query the same keyword the results obtained were quite similar to the ones offered by SocialRank. These samples have a very important meaning since they follow the same achieved by the most used search engine and prove that SocialRank can be effectively used to rank specific topics under several Social Networks.

5 Conclusions and Future Work

The recommendation system provides tools for the user receive specific data about his / her interests. It can be configured to access one or several social networks thereby to enlarge the quantity and quality of the results. Plus, the recommendation system does not only
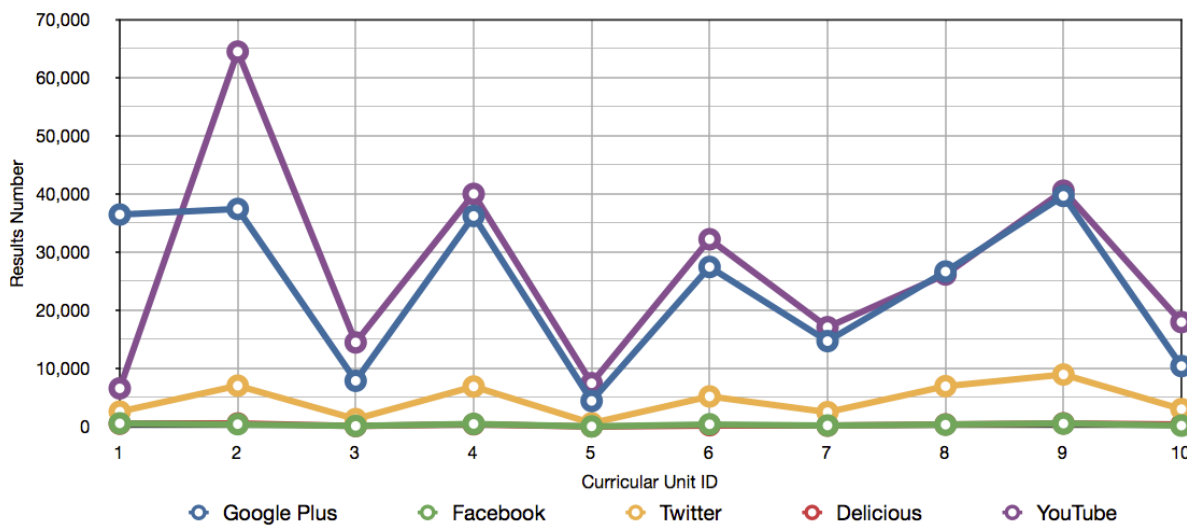
Fig. 4 Number of results on each social network (YY) by Course Unit (XX)
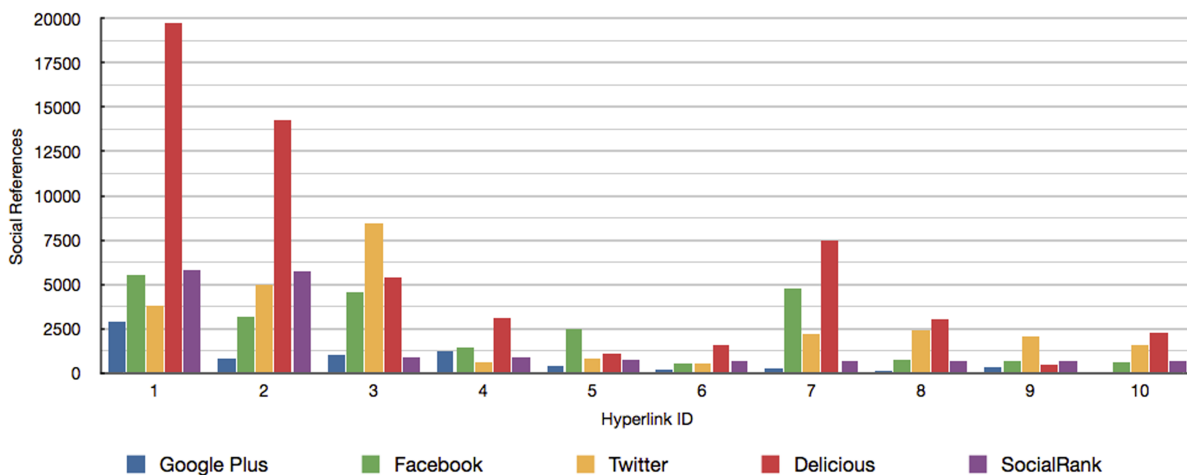
Fig. 6 Distribution of references on each social network (Facebook, Google+, Twitter, and Delicious) and SocialRank of first ten results jquery.

provide results but it assigns a rank based on the quality of the result on several social networks. Through this recommendation system and combined with LMS technologies the users has access to a set of tools that provide a quick and reliable way to get information. Further, all the information retrieved is the one that is most used on the social networks, and thus the most reliable. In the near future, the steady grow in the use of social networks can bring a lot more contents to the system/users and Web pages such as, Wikipedia or Vimeo could improve more options and results to the users. In terms of future work the authors pretend to increase the supported social networks and several Web sites of excellence as soon as they are available through web ser-

vices. Finally, the implementation of a semantic module can offer more functionality to the system.

References

1. Batsakis S, Petrakis EGM, Milios E (2009) Improving the performance of focused web crawlers. Data & Knowledge Engineering 68(10):1001–1013

2. Breslin J, Decker S (2007) The Future of Social Networks on the Internet: The Need for Semantics. IEEE Internet Computing 11(6):86–90

3. Buder J, Schwind C (2012) Learning with personalized recommender systems: A psychological view. Computers in Human Behavior 28(1):207–216

4. Dattolo A, Ferrara F, Tasso C (2010) The role of tags for recommendation: A survey. In: Human System Interactions (HSI), 2010 3rd Conference on, pp 548–555

5. Golder SA, Huberman BA (2006) Usage patterns of collaborative tagging systems. J Inf Sci 32(2):198–208

6. Guy I, Zwerdling N, Carmel D, Ronen I, Uziel E, Yogev S, Ofek-Koifman S (2009) Personalized recommendation of social software items based on social relations. In: Proceedings of the third ACM conference on Recommender systems, ACM, New York, New York, USA, pp 53–60

7. Hanneman RA, Riddle M (2005) Introduction to social network methods

8. Heymann P, Ramage D, Garcia-Molina H (2008) Social tag prediction. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ACM, Singapore, Singapore, pp 531–538

9. Hitrec I, Pogarcic I, Suman S (2011) eLearning: A social software in higher education learning. In: MIPRO, 2011 Proceedings of the 34th International Convention, pp 1207–1212

10. Kapounova J, Kostolanyova K (2011) Specific tool for evaluation of eLearning. In: Interactive Collaborative Learning (ICL), Piestany, Slovakia, pp 423–427

11. Klašnja-Milićević A, Vesin B, Ivanović M, Budimac Z (2010) E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education 56(3):885–899

12. Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, ACM, Raleigh, North Carolina, USA, pp 591–600

13. Liu K, Fang B (2010) Integrating Social Relations into Personalized Tag Recommendation. In: International Conference on Intelligent Human-Machine Systems and Cybernetics, Nanjing, China, pp 292–295

14. Liua F, Leeb HJ (2010) Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications pp 4772–4778

15. Matsuo Y, Mori J, Hamasaki M, Ishida K, Nishimura T, Takeda H, Hasida K, Ishizuka M (2006) POLYPHONET: an advanced social network extraction system from the web. In: Proceedings of the 15th international conference on World Wide Web, ACM, Edinburgh, Scotland, pp 397–406

16. Newman M (2003) The Structure and Function of Complex Networks. SIAM Review 45(2):167–256

17. Noll MG, Meinel C (2007) Web search personalization via social bookmarking and tagging. In: Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference, Springer-Verlag Berlin, Heidelberg ©2007, Busan, Korea, pp 387–380

18. Sabino FMR, Zhou L, Rodrigues JJPC (2011) Enhancing e-learning experience with online social networks. IET Communications 5(8):1147–1154

19. Simões TMC, Rodrigues, Joel J P C, de la Torre I (2011) Personal Learning Environment Box (PLE-BOX): A new approach to E-learning platforms. Computer Applications in Engineering Education

20. Song Y, Zhang L, Giles CL (2011) Automatic Tag Recommendation Algorithms for Social Recommender Systems. ACM Transactions on the Web (TWEB) 5

21. Su AYS, Yang SJH, Hwang WY, Zhang J (2010) A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments. Computers & Education 55(2):752–766

22. Tatkovic N, Bocev M, Putinja T (2010) Implementation of Web 2.0 applications in education. In: MIPRO, Opatija, Croatia, pp 1092–1095

23. Ting IH (2008) Web mining techniques for on-line social networks analysis. In: Service Systems and Service Management, Melbourne, Australia, pp 1–5

24. Wasim M, Shahzadi I, Ahmad Q, Mahmood W (2011) Extracting and modeling user interests based on social media. In: Multitopic Conference (INMIC), Karachi, Pakistan, pp 284–289

25. Weaver AC, Morrison BB (2008) Social Networking. Computer 41(2):97–100

26. Weng J, Lim EP, Jiang J, He Q (2010) Twitter-Rank: finding topic-sensitive influential twitterers. In: Proceedings of the third ACM international conference on Web search and data mining, ACM, New York, New York, USA, pp 261–270

27. Yu SJ (2011) The dynamic competitive recommendation algorithm in social network services. Information Sciences 187:1–14

Chapter 3

Improve Hypertext Results Exploring Social Networks

This chapter consists of the following article:

Improve Hypertext Results Exploring Social Networks

Jorge E. F. Costa, Joel J. P. C. Rodrigues, Orlando R. E. Pereira, Tiago M. C. Simões

Improve Hypertext Results Exploring Social Networks

Jorge E. F. Costa · **Joel J. P. C. Rodrigues** · **Orlando R. E. Pereira** ·
Tiago M.C. Simões

Abstract Internet technologies are constantly evolving as well as the way people use it. Search engines help their users to find higher and better relevant results to theirs searches. Each search engine has several distinct modules in order to retrieve the results expected by users using specific keywords. Social networks are a reliable Web technology that can directly support a content search. Nowadays, social networks are one indispensable tool to communicate through the Internet, and several hundreds of new users join them everyday. They offer more than a simple information resource since people share and save resources that had a certain relevance to them. Several studies have been performed showing the growth of social networks in people lives. This work focuses on exploring the public information available on social networks and how it can be used to improve searching through the Web. This paper presents and analyses, in detail, several social networks services, available contents, and information extraction.

Jorge E. F. Costa
University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
E-mail: jorge.costa@it.ubi.pt

Joel J. P. C. Rodrigues
Instituto de Telecomunicações, University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
Tel.: +351-275-242081
Fax: +351-275-319899
E-mail: joeljr@ieee.org

Orlando R. E. Pereira
University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
E-mail: opereira@ubi.pt

Tiago M.C. Simões
University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
E-mail: tiago.simoes@it.ubi.pt

In order to collect relevant data from social networks, a social crawler is proposed. Moreover, it proposes, describes, demonstrates, and validates a new algorithm to rank Web documents, called SocialRank. The system was evaluated and the results are discussed, being ready for use.

Keywords World Wide Web · Social Networks · Information Retrieval · SocialRank · Data Mining

1 Introduction

Over the last few years, online social networks changed the way how the information is spread over the Internet. Through them, people can know what is going on, not only with their friends but also all over the world. However, search engines still remain the leading tool for finding information efficiently from the massive Internet data [27, 16]. They return a list of pages matching with a given query; such answer is normally based on two fundamental indices. The first is the visibility of a document in the world wide web (WWW, or simply Web) regardless the query issued. The second is related to the degree of similarity between the incoming query and the page [11].

The emergence of Web 2.0 created an explosion of information available on the Web. Everyday, several hundreds of new documents and other information formats are updated and uploaded so is it extremely important to maintain and catalog all the information in order to have it quickly and easily accessible to everyone. Larry Page and Sergey Brin [19] have been able to answer requests from Web users, with PageRank algorithm powered by the well known Google search engine. Currently, several studies have been carried on showing the growth of social networks in the people lives, as may

be seen in Figure 1. Social networks could be an integrative part over the searches on the Web by the vision of Jon Kleinberg (author of the HITS algorithm [16]), he argues The quality of a search method necessarily requires human evaluation, due to the subjectivity inherent in notions such as relevance [16].
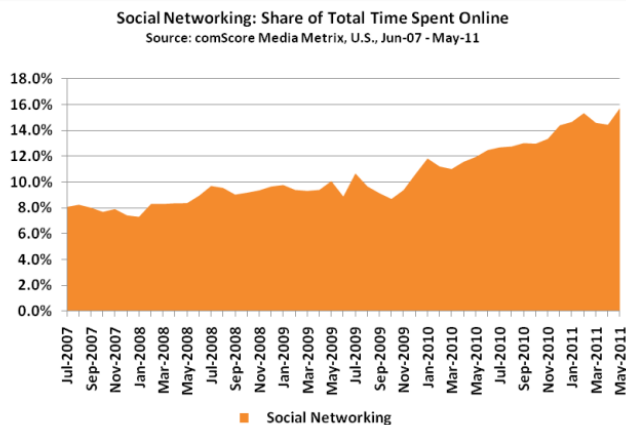

Fig. 1 Time Spent Online on Social Networks between Jul 2007 to May 2011 [5]

Social networks offer Web places where persons connect, interact, and share resources between each other. Inside these networks, there are different relations types, such as friendship, co-working, or information exchange [26]. In social networks, people make references of what is going on, what is relevant, and issues about them in their social wall or his/her friends wall. This new approach turns the Web richer and more social. Social networks improve this information exchange among people and, in few seconds, a simple Web document can be one of most important document available on the Web due to its high visibility. In this study four social networks (Facebook, Twitter, Google Plus, and Delicious) are studied. This work proposes and describes a social rank solution to improve and better sort the Web documents retrieval. The performance evaluation and validation of this proposal is also considered. Then, the main contributions of the paper are the following:

- Extensive review of the related literature focusing on social networks data mining and information retrieval;
- Analysis of social networks services and corresponding contents;
- Proposal and construction of a criterion based on social networks information to establish a rank, called SocialRank;
- Performance evaluation and validation of the proposed algorithm in comparison with the top Internet search engine, the well-known Google.

The rest of the paper is organized as follows. The related work about the topic is elaborated in Section 2 while Section 3 describes the social networks considered on this work. Each mechanism and objects provided by Web Services of theses networks are also explored. Section 4 explains the behavior of the SocialRank Algorithm. A performance evaluation study of this proposal and results are discussed on Section 5. Section 6 concludes the paper and point further research directions.

2 Related Work

A Web search engine is a piece of software that is designed to search for specific information on the Web. The results are gathered using a specific algorithm or a conjunction of multiple algorithms. These mechanisms may also search data from databases, open directories, social networks, or other online information data resources.

2.1 Ranking Algorithms for Search Engines

After years of competition, a small number of search engines with advanced algorithms dominate information seeking on the Web [15]. The need to present the best result is huge, and several search engines have worked to give the best results as quicker as possible to their users. Silverstein, Henzinger, Marais, and Moricz [22] examined query logs from AltaVista, the bigger Search Engine in the nineties. The logs included more than one billion queries and 285 million user sessions. In December 2010, Optify, Inc. made a study over Search Engine Results Page (SERPs). The two studies shown that users usually do not look at more than the first ten results. This means that Search Engines need to use efficient information retrieval techniques to find and organize the desired information [21]. In this section, several algorithms that improved the Web information retrieval are presented.

2.1.1 In-Degree

In the early days of Web search, several search engines (Altavista, HotBoy, etc) had applied the simple heuristic. This simple heuristic that can be viewed as the predecessor of all links analysis algorithms ranks the pages according to their popularity. The number of pages that point it measures the popularity of a page on the web. Although in one of Kleinberg studies, he makes a convincing argument that this algorithm is not sophisticated enough to capture the authoritativeness of a node; if a search engine apply this simple ranking

scheme it should be very easy for a Web master to influence authority; they could simply create thousands of linked pages that point to the authoritative page resulting on a high ranked page but with low real interesting information.

2.1.2 PageRank

Brin and Page [19] extended the idea of In-Degree by observing that Web links do not have the same importance and relevance. PageRank extends this idea and do not count links from all pages equally. They are normalized by the amount of links on a page [3, 4]. Equation 1 is used to achieve a specific Page Rank number.

$$PR(A) = \frac{1-d}{N} + d \sum_{Ti \in M(A)} \frac{PR(Ti)}{C(Ti))} \qquad (1)$$

where PR(A) is the PageRank of page A, PR(Ti) is the PageRank of pages Ti which link to page A, C(Ti) is the number of outbound links on the page Ti and D is a damping factor, which can be set between 0 and 1. Usually, 0.85 is used by default.

PageRank [19] is based on the Random Surfer model and is the mainstay of the highly popular Google search engine. The Random Surfer model assumes that a user randomly keeps on clicking the links on a page and if he/she get bored of a page then switches to another page randomly.

2.1.3 HITS (Hypertext Induced Topic Selection)

Hypertext induced topic selection (HITS) is an iterative algorithm based on the linkage of the documents on the Web. Also known as hubs and authorities, HITS addresses the abundance problem where, during a search process, too many pages are available, which are not relevant to the query. To solve this issue, the algorithm uses the link structure of the Web to discover pages that can be considered as authoritative on a broad search topic [12]. The authority of a Web page is percept as the relevance and importance of that page for a specific topic in the Web community. This algorithm considers a page such as an authority on a topic if many pages relevant to that topic include a reference to that same page. Pages that point to many related authorities are called hubs. This concept can be easily understood by the illustration available at Figure 2.

2.1.4 Focused Rank

Focused rank [18] is a link-based ranking. The algorithm searches points of interest in a particular topic
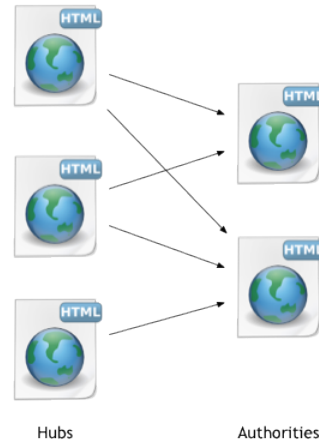

Fig. 2 Illustration of Hubs and Authorities based on links between Web pages

between pages with similar content. If there is a hyperlink between a page u and a page v, they must contain at least one common topic. So the algorithm obtains a set of probabilities on a group of documents and a list of topics. Then, the topical overlap between two documents is less when they have several topics in common.

2.2 Social Network Data Mining

In recent years, organizations and researchers have explored the social networks. The information and services are used in diverse and different ways such as generate gains, entertaining, games, or to understand their users. Through those explorations, these organizations can improve the user experience in several systems and applications, which are included on several areas as electronic learning (e-learning) or recommendation systems. In the available literature, several studies and systems that use data mining of social networks information are found. In these studies several approaches are presented, such as TunkRank [24], a tool for measuring the influence of user on the Twitter network, which calculates how much attention the followers actually give to their following person. In [13], the social tags (keyword annotations) through the social network Delicious show how social bookmarking services can enhance Web searches is analyzed. In [17], the authors did the first quantitative study inside a big collection of information over Twitter. They begin with the network analysis and studied the distributions of followers and followings, and the relation between followers and tweets. Next, they rank users by the number of followers, PageRank [19], and the number of *retweets*, present a quantitative comparison among them. Seok Jong Yu [28] proposed the dynamic competitive recommendation (DCR) algorithm based on the competition

of multiple component algorithms. He concluded that social networks are no longer simple social spaces for making friends, and are evolving toward a new type of information network. In terms of performance evaluation, they could confirm that DCR algorithm presents a more stable and consistent recommendation. Outside research domain, two Web applications that crawl social networks and let people focus their searches for social results, called DuckDuckGo (http://duckduckgo.com) and Topsy (http://topsy.com), were found. DuckDuckGo describes themselves as providers and allows users to search content published on Twitter and on the Web, sorted by relevance or date. However, no data is given about the services or specifications that system uses to access the information. Topsy, in particular, provides an interesting tool called Topsy Social Analytics. Through this tool it is possible to type several keywords and Topsy analytics will chart the number of mentions found to each keyword.

2.2.1 Social Network Ranking approach in Hypertext Systems

Nowadays, social networks are deeply involved with people life, and as abovementioned, social networks are no longer simple social spaces for making friends. Nielsen Company (http://www.nielsen.com), a global leader in measurement and information, provided several interesting information about the state of media document made in the 2rd quarter 2011. Based on its document, in the United States (US), social networks and blogs reach nearly about 80 percent of active US Internet users, representing the majority of Americans time online. Furthermore, Google and Microsoft Bing had introduced and improved their own search with social networks. Google started with the addition of Twitter. However, currently, Google supports only Google Plus. Microsoft introduced the Bing Social (http://bing.com/social), where their crawl Facebook and Twitter posts. It is one first signal from search engines since they need to adapt to the current Web.

3 Social Networks Presentation

The universe of social networks is huge (being currently available more than 200 applications), however, not all of them have the same audience. It was been an important tiebreaker factor on the selection of the social networks for this work. In this section, the selected social networks are introduced and their focus, number of known users, communication application programming interface (APIs), and other important features are described. The considered social networks were selected

taking into account the number of known users and Web relevance. A particular focus on the services that a specific social network offers is given.

3.1 Online Social Networks

1. Facebook

   Facebook was launched in 2004 and its mission passes to make the world more open and connected. Mainly, Facebook allows people staying connected with friends and family, sharing comments with them, and discover what is going on in the world. Actually, this social network had 1 billion active users monthly [29]. To researchers and developers, Facebook provide a powerful tool called Social Graph API. It is the Facebooks core, and it presents a simple and consistent view of Facebook social graph. It provides the objects representation in a graph (e.g. people, photos, events, and pages) and the connections between them. Every object in the social graph has a single ID, and allows the properties of an object access by different URLs. Some of these objects are users, pages, events, groups, applications, status messages, photos, photo albums, profile pictures, and videos, among others. All the objects are connected to each other via relationships, also available through the Graph API. Every time the API is called it answers in the form of JavaScript Object Notation (JSON) objects. The Graph API allows easy access to all the public information about an object. Then, if additional information about a specific user is needed, getting his/her permission first is required. As above described, the Facebook Graph API offers access to several objects. However, it is possible to access through the Web Service search engine to all the public objects in the social graph, and refine the expected results using several parameters.

2. Twitter

   Twitter is an information network and communication mechanism that produces more than 340 millions tweets a day. On Twitter, anyone can read, write, and share messages of up to 140 characters. These messages, also called tweets, are public and available to anyone interested in them. Currently, Twitter has about 140 millions of active users [8]. So, Twitter is a quick, easy to read, and public social network providing a powerful real-time communication platform. Similar to Facebook, Twitter also has its own API. Twitter has two types of API,

the REST API and the Streaming API. Connecting to the streaming API requires keeping a persistent open HTTP connection, differently of the REST API interaction, which needs a request to receive any fresh data. Given the nature of the current work, the REST API approach was selected. Twitter provides access to several resources, such as timelines, tweets, search, streaming, direct messages, friends and followers, users, suggested users, places, and trends, among others. The results can be acquired in different formats, including JSON, eXtensible Markup Language (XML), Really Simple Syndication (RSS), or ATOM feeds.

3. Delicious

Delicious was founded in 2003 as a social bookmarking service. Currently, Delicious is used to save and show users bookmarks. Unlike the most social networks, Delicious follows one different philosophy. Many social networks offer several plugins or services to include in Web pages. Delicious created a plugin to install on the bookmarks bar of the most available Web browsers. Delicious provides two different ways to acquire new contents or specific contents, through feeds or through an API. To obtain new contents feeds can be used. These feeds come in several formats, RSS or JSON, and display what is going on at Delicious. It is really useful to read and access bookmarks on a reader, blog, or even on a third-party application. All the feeds are available by a single URL prefix. However, the results will change depending on several parameters. Username, Tag, or URL, can be different parameters that can change the result set. Thus, Delicious allows the reception of several and diverse feeds, such as recent bookmarks, recent bookmarks by tag, bookmarks for a specific user, public summary information about a user, bookmarks from a users subscriptions, recent bookmarks for a URL, summary information about a URL, etc. Following the most part of social networks, Delicious also offers an API. This API allows a user to add, update, get, or delete information from its own account.

4. Google Plus

Google Plus (G+) was launched in June 2011. Typically, all Google services grow exponentially and Google plus is no exception. Currently, it has about 170 million of active users [10]. Differently from its direct rival Facebook, Google introduced new services as Circles, Hangouts, and Sparks. Google fol-

lows a similar philosophy to Facebook or Twitter, using the like concept through the plus one (1+) button. Despite a joint effort of Google developers to provide a full collection of services, is important to note that Google Plus API still in its infancy. This API can be accessed by REST in order to obtains and explore Google Plus content. This API is organized by resource type and each resource type has one or more data representations and one or more methods. These methods allow getting, searching or list, people, activities, or comments resources.

3.2 Social networks services review

Every social network provides distinct services to access information in their repositories. Companies use these services everyday and they created a new business on the Web with social games, publicity, etc. The authors perspective is different. The main goal includes the information and knowledge searching improvement in hypertext systems. Currently, millions of documents are placed and Search Engine Optimization (SEO) companies, like Google or Microsoft (with Bing) improve their algorithms in order to return best results. In order to evaluate the proposal of a robust method for information retrieval in social networks the considered four social networks are studies in detail. With the multitude of data created at each minute, social networks cannot turn available all the data through a single service. However, a similar service to achieve the expected data (the search service provided by all the social networks) was explored. This service allows searching given several keywords. It will retrieve posts, people, events, or other social information about a specific topic. The followed approach includes the catch of posts with hyperlinks and, thereafter, also other information, such as likes, favorites, shares, *retweets*, people, and comments. Figure 3 summarizes a social networks study over their services. It is important to perceive the social data in common between themselves, in order to use a common criterion for their evaluation. As may be seen on this figure, an information summary returned by the services of each social network is presented. It is possible to observe that four common properties available in every social network are date, name, text, and hyperlinks (if some of them are include on it). The public information that is possible to obtain about a specific user is summarized in Figure 4. As abovementioned at the related work section, information about users could influence the results significantly. Twitter friends and followers are a good example. If a user shows more followers, it means their shares may have more importance. Gender

also can be an important factor. Imagine the application tion knows the user gender, the retrieved information should be related with high sharing rates on this genre.
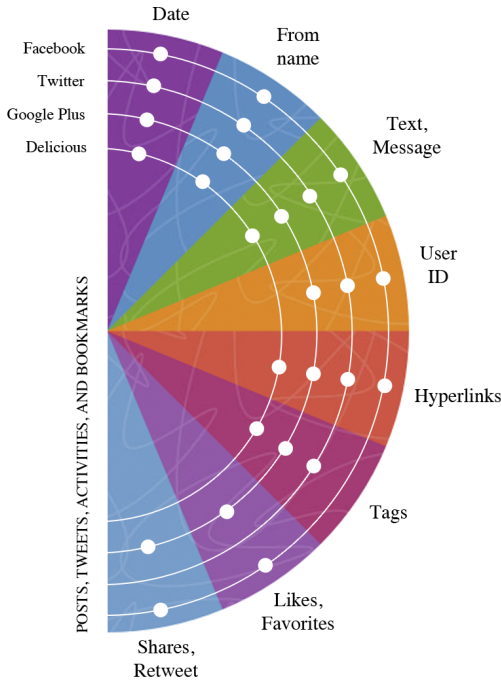

Fig. 3 Information available of Posts, Tweets, Activities, and Bookmarks in each social network

Fig. 4 Information available about users in each social network

Similarly to Figure 3, Figure 4 follows the same approach. In Figure 4 the only information shared by the four social networks is the name and the hyperlink of the users page. However, as already mentioned, the gender and followers, even incomplete, can be useful.

In Figure 5, the chart is focused on hyperlinks information available on the social networks, as URL, Title, Description, Total number of likes, Total number of shares, etc.

In the Figure 5, not all the information about the hyperlinks is given, as Title and Description, however, this issue can be fixed catching the meta-tags of the Web site. These three figures offer a deeper illustration about the information can be obtained in order to improve searching in a hypertext system. As above referred, not all the information is complete. For instance, total likes and owner of the shared information are two features that are not available. Thus, about social information, the total number of shares and saves are the complete data about every resource.
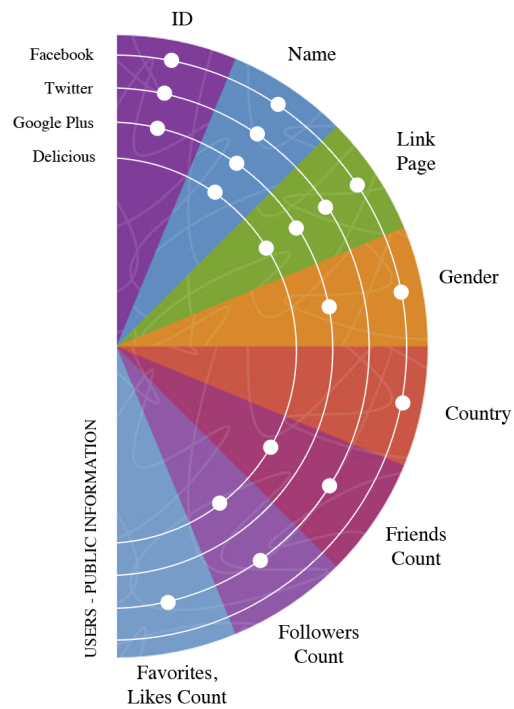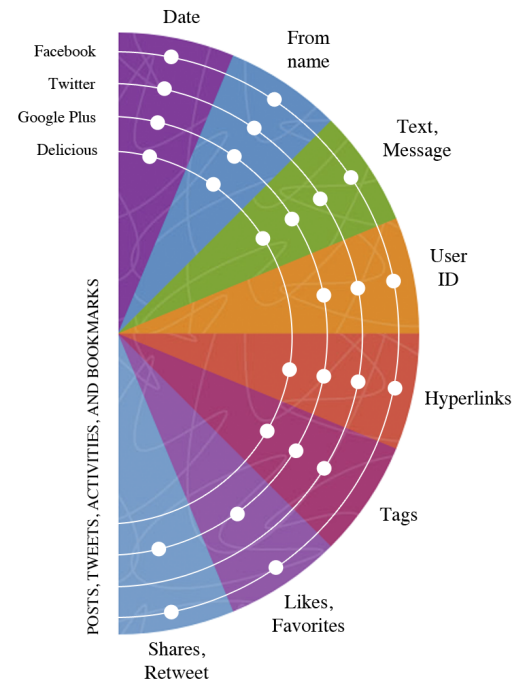

Fig. 5 Information available of hyperlinks in each social network

4 SocialRank Algorithm

After a preliminary study about the resources and services provided by the four considered social networks,

a mathematical model to rank each available resource, called SocialRank, was proposed. To reach the presented results an application with two layers was created. It considers a Crawler and a SocialRank. The Social Crawler is responsible for data querying and keeping the documents found in social networks updated. Finally, the second layer is composed by the SocialRank algorithm, responsible for assigning an importance to all the documents on the repository. Next, the two layers of the proposed model are described in detail.

4.1 Crawlers

Crawlers (also known as Robots or Spiders) are computer programs or suite of programs capable of iteratively and automatically download and explore the Web. Particularly, search engines use them and their task can consider specific subjects or just browsing and download Web documents. Usually, to start a crawler, a set of Web pages (seed pages) on their input is given. After the outgoing links extraction is performed, it determines the links to visit hereafter, based on certain criteria, and so on. Web pages pointed by these links are downloaded, and those satisfying certain relevance criteria are stored in a local repository. Moreover, crawlers perform an exhausted search and typically they try to keep the information up-to-date [2].

4.2 Social Crawler

The current version of the Social Crawler was developed in Java programming language and uses MySQL to manage the social data. Unlike the most general crawlers, the Social Crawler is fed through a set of Web Services provided by the above-mentioned social networks, searching a specific set of keywords each time that service is called. The use of Web Services offers many advantages, since they turn the platform and used technology independent, allowing different applications to communicate with each other and share data and services among them. It also uses a standardized industry protocol for the communication and low cost of communication. Furthermore, almost every social network uses services to provide ways to manage and interact with their content. Through this method, the performance evaluation of the proposal becomes more effective taking into account the common used technologies. Social Crawler architecture is shown in Figure 6.

Following Figure 6, the process of crawling start in the *Queuer*. It communicates with the *Repository* that contains the initial keywords and searched user keywords to be used on the requests to the *Social Network*

Servers. If the response contains any data, it will be filtered in the *Download and Parser* module and the posts with hyperlinks will be saved in a repository, together with social information as comments, likes, and descriptions. Following, every hyperlink will be added to the *Scheduler* to make a resume about them on all the considered social networks (Facebook, G+, Twitter, and Delicious). *Updater* module is not included in the main cycle but it is also very important, since it keeps the information up-to-date. Assigning and using the update frequency, presented on Section 5, keeps the *SocialRank* more precise.

The download process gets social information of a specific resource, such as the following:

- Who shared;
- How many people shared;
- How many people liked;
- What was the description;
- Comments;
- Etc.

Distinct to general social networks, Delicious is a collaborative tagging service. Through it, bookmarks with user-tags and distinct descriptions about one resource can be found. In this social network users are allowing to make their personal collection of favorite Web resources. The content of this collection can attract users with similar minds. Through the simplicity of these services, users can easily create bookmarks and add annotations [1] in form of tags to the bookmarks [20, 9]. Delicious is a social bookmarking service and enjoys a vast popularity. However, not all-important information is social information, and the process of Download and Parser will analyze the custom information always saved by search engines that goes to the repository, such as title, description, and meta-tags. As previously mentioned, the Scheduler forces the crawler looking for social information about the domain of every URL as pages, logos, and the above-mentioned data. The Updater has the function to maintain the data updated. The SocialRank algorithm is explained below.

4.3 Adjusting Request Frequency

It is known that several specific keywords are not mentioned on social networks by millions of people. However, there are other keywords that other millions of users mention everyday (e.g. Obama, music, jquery, etc.). In order to improve one effective crawling and save resources, a method to assigned a crawl frequency to keywords was deployed [25]. The main idea is about tying one timeframe (one day, one week, and two weeks) to
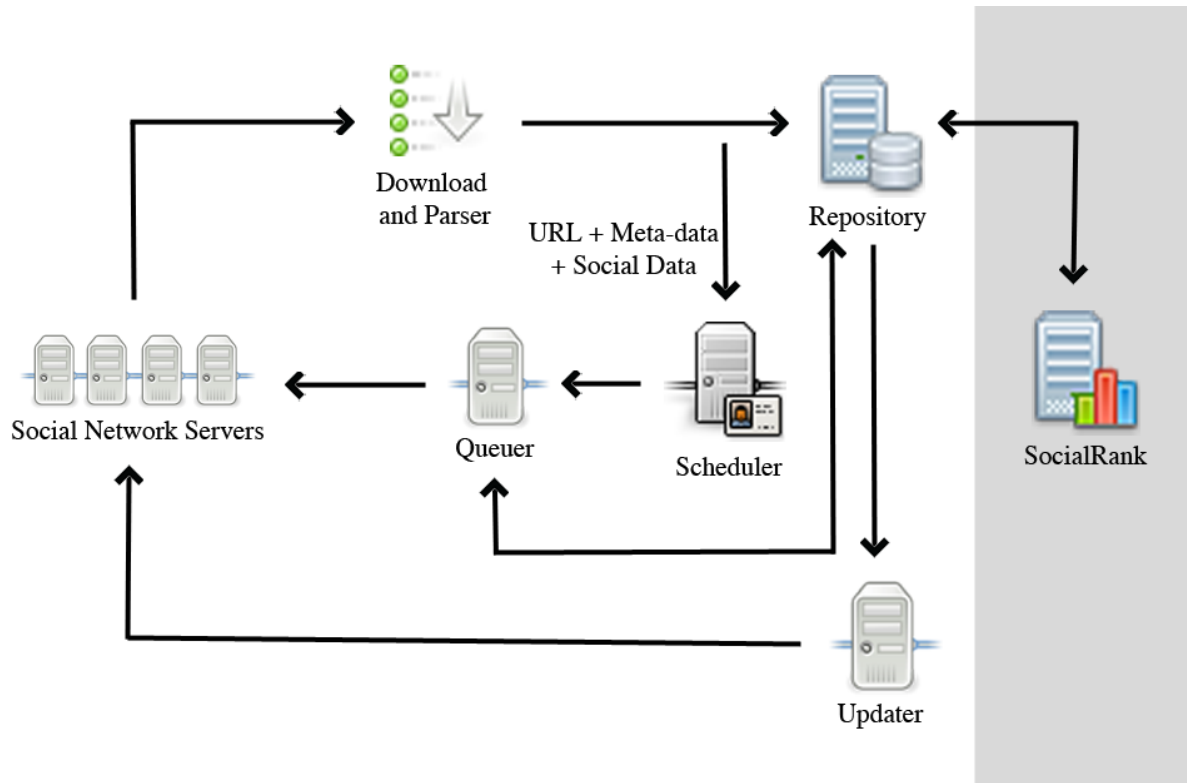
Fig. 6 Illustration of the Social Crawler Data Flow

each keyword. Taking into account the related literature, a crawler may request frequently and its frequency may be adjusted in the future according to the category of the Web site [7]. The followed approach is based on the last twenty requests and it will change depending on the results and the number of results. It is defined in such a way because one term, like euro football, will be mentioned lots of times during the tournament and in the next four years will have a low referral. Taking into account number of days that a specific topic was mentioned, a simplistic method to assign the frequency to keywords was created. The method assigns one of three different results: daily, weekly, and fortnightly influenced by the references of the last 20 days. The method is represented by the following pseudo-code.

> **for** last 20 days **do**
>   **if** $dayReference \leq 0$ **then**
> $counter \leftarrow counter + 1$
>   **end if**
> **end for**
> **if** $counter \leq 4$ **then**
> $result \leftarrow Fortnightly$
> **end if**
> **if** $counter > 4 < 12$ **then**
> $result \leftarrow Weekly$
> **end if**
> **if** $counter \geq 12$ **then**

> $result \leftarrow Daily$
> **end if**

4.4 SocialRank

The act of sharing or saving (such as in Delicious) some document means the document has some importance to the user. Plus, between the both actions, the share action may be seen with more significance because the person who shared it think that resource is useful to his/her friends or followers. In order to perceive and rank the shared data in social networks a robust, fast, and reliable algorithm was created. After a deep analyzes it was concluded that Web documents tend to be uniformly distributed on social networks. Although, there are several scenarios where this behavior does not occur. Several hyperlinks have an exceeded number of social references in a given social network (typically on Twitter) when compared with the other social networks. Usually, this phenomenon happens when people need to access to a certain resource but to access it, they must share it on a specific social network. Instead of a payment be made with cash, it is done with a simple share action. In order to create an accurate and realistic solution the proposed algorithm was refined in order to give some standardization and unification. Then, the

standard deviation (represented by σ) is calculated in order to unify the final results. It expresses the available variation or dispersion in the average value. In Table I it is possible to observe when low standard deviation indicates that social references tend to be very close to the average.

Instance	1	2	3
Facebook	200	200	20
Twitter	1150	230	2400
Google +	80	150	30
Delicious	50	180	5
Standard Deviation	524.02	33.67	1190.88

Table I Instances of the Standard Deviation by Social References

The proposed equation (Equation 2) calculates the SocialRank (SR) of a given resource (x). It will be used after the Scheduler (module) inside the Crawler layer. The equation will sum all the references and divide them by the number of supported social networks (t) on the system, actually four (Facebook, Twitter, Google Plus, and Delicious). In order to smooth the results and achieve more accurate results, the standard derivation (σ) was added.

$$SR(x) = \frac{\sum r}{t} . \frac{10^{N(log10(\sigma.2))}}{\sigma.2} \qquad (2)$$

5 Performance Evaluation and Results

In order to evaluate and validate the proposed model, series of experiments over the SocialRank were performed. The main objective of this evaluation was the validation of results. The retrieved results are quite similar to a custom search engines. So, for comparison purposes the most used search engine, Google, was chosen (Figure 7). Figure 7 presents the top five search engines with corresponding usage quote (in percentage). Furthermore, most of the Web search engines present similar results to Google.

The results are presented with a comparison, considering Google *versus* SocialRank, the algorithm proposed in this work. The SocialRank importance number, is depicted on the third column of each table below and determines the order of results. High SocialRank indicates that page was mentioned a lot, as above described. SocialRank database is smaller when compared with the Google (45 Billions webpages). Until now, this work focus on experiments and prove this approach can give a different view of the hyperlinks space on the World Wide Web.
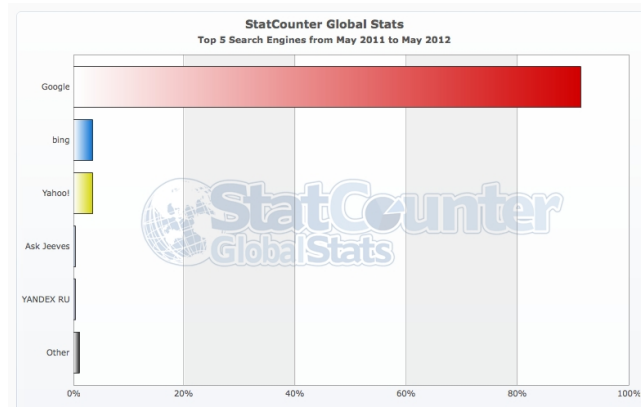

Fig. 7 Search Engines GlobalStats May 2011 to May 2012 [23]

5.1 Example Queries

This section presents the queries used in order to accomplish the experiments. The results are available in the Tables II, III, and IV. Each table presents the first ten results of a given keyword on the two considered systems. Finally, the results are analyzed in detail.

5.1.1 Linux

Linux is an Operating System (OS) and successful open-source software to development and distribution [14]. "Linux" keyword query is the request that presents the most different results (one similar result). The Social-Rank results achieved are most directed to versions and how to install, unlike Google that are directed to linux distributions and definitions. In this sense, it is possible understand that Web users are more concerned on use this Operating System to obtain advantages and benefits. The results are presented on Table II.

5.1.2 Weather

"Weather" keyword is the most similar query of considered three examples. It was found four similar hyperlinks and three of these Web documents are the major weather sites, The Weather Channel, The Weather Underground, and The Weather Network. The results are presented on Table III.

5.1.3 jQuery

jQuery is an Open Source solution and one of the most popular frameworks used to built complex Web pages [6]. As "Weather", "jquery" keyword results also include the most popular and known hyperlinks. Together with Google results, SocialRank results are directed to

#	Proposed System Results	Google Results	SocialRank
1	*www.backtrack − linux.org/*	*en.wikipedia.org/wiki/Linux*	4074.71
2	http://linuxmint.com/	www.ubuntu.com/	3670.51
3	http://www.kernel.org	www.linux.org/	1031.97
4	*www.makelinux.net/kernel_map*	*www.linux.com/*	778.76
5	https://one.ubuntu.com/	www.redhat.com/	660.42
6	http://www.linuxjournal.com/	www.debian.org/	620.69
7	http://www.linuxalt.com/	distrowatch.com/	583.01
8	http://www.linuxliveusb.com/	www.kernel.org/	556.90
9	http://www.gnupg.org/	www.linuxfoundation.org/	552.43
10	http://www.pendrivelinux.com/	www.mandriva.com/	539.59

Table II SocialRank Query Results to the **linux** keyword.

#	Proposed System Results	Google Results	SocialRank
1	*www.weather.com*	*www.weather.com/*	45848.54
2	spaceweather.com	www.bbc.co.uk/weather/	32842.06
3	www.livewxradar.com/	www.theweathernetwork.com/	4885.57
4	*www.wunderground.com/*	*www.accuweather.com/*	4763.39
5	www.weather.gov	www.wunderground.com/	3756.46
6	weatherspark.com/	www.metservice.com/	752.61
7	accuweather.com/	www.weathersa.co.za/	594.62
8	www.myweather.com/	www.metoffice.gov.uk/weather/	592.44
9	weather.yahoo.com/	weather.gov/	497.41
10	www.theweathernetwork.com	weather.cnn.com/	480.65

Table III SocialRank Query Results to the **Weather** keyword.

#	Proposed System Results	Google Results	SocialRank
1	*jquery.com/*	*jquery.com/*	5845.75
2	jsfiddle.net/	www.w3schools.com/jquery/	5734.51
3	jquerymobile.com/	jqueryui.com/	951.05
4	*jqueryui.com/*	*en.wikipedia.org/wiki/JQuery*	875.43
5	codecanyon.net/	jquery.org/	768.19
6	jquery.malsup.com/cycle/	jquery.org/license/	715.83
7	nivo.dev7studios.com/	jquerymobile.com/	674.79
8	isotope.metafizzy.co/	www.noupe.com/jquery/	670.01
9	www.turnjs.com/	jquerytools.org/	667.92
10	scriptsrc.net/	twitter.com/jquery	662.48

Table IV SocialRank Query Results to the **jQuery** keyword.

learn and apply this technology. The results are presented on Table IV.

6 Conclusions and Future Work

This paper conducted a study on social networks contents and services. A study addressing how and what information can be retrieved through the selected social networks Web services (Facebook, Twitter, Google Plus, and Delicious) was performed. The main contribution of this paper includes the proposal and validation of an algorithm based on references distributed by social networks for establishing a new criterion based on social networks information, called SocialRank. In order to validate the algorithm, several experiments against the top Internet search engine, Google, were conducted. The results were quite interesting and promising. It was observed that proposed solution performed very well for

the four considered social networks. Based on all the performed experiments, it can be concluded that social networks data mining will be imperative on information retrieval. However, a large-scale Web search engine is a complex system and much remains to be done. The immediate goals include developments to improve search efficiency and integration in a real Web search engine system. Another area that requires more research is the request frequency. Also, to add simples features supported by most search engines like boolean operators, negation, and stemming. Simple experiments indicate that SocialRank can be personalized by increasing the evaluation and experiments around friends and followers number, which can be seen as a factor of importance. Gender also can be an important factor. Redirecting the results with a specific gender, high sharing rates to a male or female user can be found. These suggestions are addressed for further works.

References

1. Bao S, Xue G, Wu X, Yu Y, Fei B, Su Z (2007) Optimizing web search using social annotations. In: Proceedings of the 16th international conference on World Wide Web, ACM, Banff, Alberta, Canada, pp 501–510
2. Batsakis S, Petrakis EGM, Milios E (2009) Improving the performance of focused web crawlers. Data Knowl Eng 68(10):1001–1013
3. Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. In: Proceedings of the seventh international conference on World Wide Web 7, Elsevier Science Publishers B. V., Brisbane, Australia, pp 107–117
4. Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. Comput Netw ISDN Syst 30(1-7):107–117
5. comScore (2012) Media Metrix Core Reports - comScore, Inc. URL http://www.comscore.com/
6. David M (2011) Developing Websites with jQuery Mobile. Focal Press
7. Dixit A, Sharma AK (2010) A mathematical model for crawler revisit frequency. In: Advance Computing Conference (IACC), 2010 IEEE 2nd International, pp 316–319
8. Dorsey J (2012) Twitter for Business. URL https://business.twitter.com/basics/what-is-twitter/
9. Golder SA, Huberman BA (2006) Usage patterns of collaborative tagging systems. J Inf Sci 32(2):198–208
10. Google Inc (2012) Now with 170 million users google get a complete redesign. URL http://thenextweb.com/google/2012/04/11/now-with-170-million-users-google-gets-a-complete-redesign/
11. Gori M, Numerico T (2003) Social networks and web minorities. Cognitive Systems Research 4(4):355–364
12. Gupta A, Jindal R (2008) An Overview of Ranking Algorithms for Search Engines. In: Proceedings of the 2nd National Conference; INDIACom-2008, New Delhi
13. Heymann P, Ramage D, Garcia-Molina H (2008) Social tag prediction. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ACM, Singapore, Singapore, pp 531–538
14. Israeli A, Feitelson DG (2010) The Linux kernel as a case study in software evolution. Journal of Systems and Software 83(3):485–501
15. Jason Morrison P (2008) Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web. Information Processing & Management 44(4):1562–1579
16. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM 46(5):604–632
17. Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, ACM, Raleigh, North Carolina, USA, pp 591–600
18. Miao TAA, Das O'Brien T, Zhen WGZ (2007) A link-based ranking scheme for focused search. In: Proceedings of the 16th international conference on World Wide Web, Canada
19. Page L, Brin S, Motwani R, Winograd T (1999) The PageRank Citation Ranking: Bringing Order to the Web. Tech. rep.
20. Ram S, Wei W (2010) How social is social bookmarking? In: Business Applications of Social Network Analysis (BASNA), 2010 IEEE International Workshop on, pp 1–8
21. Selvan MP, Sekar Ac, Dharshini AP (2012) Survey on Web Page Ranking Algorithms. International Journal of Computer Applications 41:1–7
22. Silverstein C, Marais H, Henzinger M, Moricz M (1999) Analysis of a very large web search engine

query log. SIGIR Forum 33(1):6–12

23. StatCounter (2012) StatCounter Global Stats - Browser, OS, Search Engine including Mobile Market Share. URL http://gs.statcounter.com/

24. Weng J, Lim EP, Jiang J, He Q (2010) Twitter-Rank: finding topic-sensitive influential twitterers. In: Proceedings of the third ACM international conference on Web search and data mining, ACM, New York, New York, USA, pp 261–270

25. Wolf JL, Squillante MS, Yu PS, Sethuraman J, Ozsen L (2002) Optimal crawling strategies for web search engines. In: Proceedings of the 11th international conference on World Wide Web, ACM, Honolulu, Hawaii, USA, pp 136–147

26. Xin W, Jamaliding Q, Okamoto T (2009) Discovering Social Network to Improve Recommender System for Group Learning Support. In: Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on, pp 1–4

27. Yan L, Gui Z, Du W, Guo Q (2011) An Improved PageRank Method based on Genetic Algorithm for Web Search. Procedia Engineering 15(0):2983–2987

28. Yu SJ (2012) The dynamic competitive recommendation algorithm in social network services. Inf Sci 187:1–14

29. Zuckerberg M (2004) Key Facts - Facebook Newsroom. URL http://newsroom.fb.com/content

32

Chapter 4

Conclusions and Future Work

This chapter presents a synthesis of the main achievements and point to several directions for future work. The main objective of this dissertation was the proposal and construction of a recommender to incorporate inside a custom LMS over social networks information and course units for each student.

Chapter 1 introduced and delimited the topic of the dissertation, fixed the objectives, and presented the main contributions of this work. Chapter 2 described the anatomy of the proposed recommendation tool. It began with an introduction about e-Learning and recommendation system, the best way to filter out the overloaded information efficiently. After, it focused on several social networks, which could provide better learning contents. The state of the art about recommendation systems over social networks information was also presented in this chapter. The performance evaluation analysis of the proposed system was also discussed.

Based on good results achieved by SocialRank, Chapter 3 followed a new approach generalizing the proposed model to improve search quality on Web search engines. Several studies about social networks and their services were also studied. The performance assessment and results are analyzed in comparison with Google search engine results.

To achieve the above described contributions, several objectives were proposed. The study began with the review of the related literature. Further, a study about social networks and related topics, including their services, was conducted. With this study, available and useful information about social networks services was identified. A prototype approach to

demonstrate, evaluate, and validate the proposed solution was created and described in detail. The proposed ranking algorithm for classifying social networks information, called SocialRank, was included on the prototype and its results are very promising. Then, the entire dissertation objectives were successful accomplished.

To conclude this dissertation, there are some suggestions for further research works:

1. Improve the number of social networks, such as StumbleUpon, Wikipédia, or Tumblr (images); this could give a strong support and new results to users.
2. Apply and deep the SocialRank, since it seems very close to the current Web.

Based on all the performed experiments, it can be concluded that social networks data mining will be authoritative on information retrieval. Simple experiments indicate that SocialRank can be personalized by increasing the evaluation and experiments around friends and followers number, which can be seen as a factor of importance.

Through this document the author really expect that work can contribute to increase support for students learning process. Furthermore, following the second proposed approach, better search engine results may be obtained from social networks.