Computer Communications 73 (2016) 200–210

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Endorsement deduction and ranking in social networks

Hebert Pérez-Rosés[a,b,*], Francesc Sebé[a], Josep Maria Ribó[c,†]

[a] Department of Mathematics, Universitat de Lleida, Spain
[b] University of Newcastle, Australia
[c] Department of Computer Science and Industrial Engineering, Universitat de Lleida, Spain

ARTICLE INFO

Article history:
Available online 8 September 2015

Keywords:
Expertise retrieval
Social networks
LinkedIn
ResearchGate
PageRank

ABSTRACT

Some social networks, such as LinkedIn and ResearchGate, allow user endorsements for specific skills. In this way, for each skill we get a directed graph where the nodes correspond to users' profiles and the arcs represent endorsement relations. From the number and quality of the endorsements received, an authority score can be assigned to each profile. In this paper we propose an authority score computation method that takes into account the relations existing among different skills. Our method is based on enriching the information contained in the digraph of endorsements corresponding to a specific skill, and then applying a ranking method admitting weighted digraphs, such as PageRank. We describe the method, and test it on a synthetic network of 1493 nodes, fitted with endorsements.

1. Introduction

Directed graphs (digraphs) are an appropriate tool for modelling social networks with asymmetric binary relations. For instance, the blogosphere is a social network composed of blogs/bloggers and the directed 'recommendation' or 'follower' relations among them. Other examples include 'trust' statements in recommendation systems (some user states that he/she trusts the recommendations given by some other user) and 'endorsements' in professional social networks. Additionally, weighted arcs appear in situations where such relations can accommodate some degree of confidence ('trust' or 'endorsement' statements could be partial).

LinkedIn and ResearchGate are two prominent examples of professional social networks implementing the *endorsement* feature. LinkedIn[1] is a wide-scope professional network launched in 2003. More than a decade later it boasts a membership of over 364 million, and it has become an essential tool in professional networking. The LinkedIn endorsement feature, introduced about three years ago,[2] allows a user to endorse other users for specific skills.

On the other hand, ResearchGate[3] is a smaller network catering to scientists and academics. It was launched in 2008, and it reached five million members in August, 2014. ResearchGate also introduced an endorsement feature recently.[4] From the endorsements shown in an applicant's profile, a potential employer can assess the applicant's skills with a higher level of confidence than say, by just looking at his/her CV.

The two endorsement systems described above are very similar: for each particular skill, the endorsements make up the arcs of a directed graph, whose vertices are the members' profiles. In principle, these endorsement digraphs could be used to compute an authority ranking of the members with respect to each particular skill. This authority ranking may provide a better assessment of a person's profile, and it could become the basis for several social network applications.

For instance, this authority ranking could be the core element of an eventual tool for finding people who are proficient in a certain skill, very much like a web search engine. It could also find important applications in profile personalization. For example, if a certain user is an expert in some field, say 'Operations Research', the system can display ads, job openings, and conference announcements related to that field in the user's profile. Finally, we can envisage a world where people could vote on certain decisions via social networks. For example, a community of web developers could decide on the adoption of some particular web standard. In that scenario, we might think about a *weighted voting scheme*, where the weight of each vote is proportional to the person's expertise in that area.

Now, people usually have more than one skill, with some of those skills being related. For example, the skill 'Java' is a particular case of the skill 'Programming', which in turn is strongly related with the

* Corresponding author at: Department of Mathematics, Universitat de Lleida, Spain. Tel.: +34 973 702 781.
E-mail address: hebert.perez@matematica.udl.cat, Hebert.Perez@gmail.com (H. Pérez-Rosés).
† Deceased.

[1] http://www.linkedin.com.
[2] More precisely, on September 24, 2012.
[3] http://www.researchgate.net.
[4] On February 7, 2013.

http://dx.doi.org/10.1016/j.comcom.2015.08.018
0140-3664/

skill 'Algorithms'. It may well happen that a person is not endorsed for the skill 'Programming', but he/she is endorsed for the skills 'Java' and 'Algorithms'. From those endorsements it can be deduced with a fair degree of confidence that the person also possesses the skill 'Programming'. In other words, a person's ranking with respect to the skills 'Java' and 'Algorithms' affects his/her ranking with respect to the skill 'Programming'.

If the members of a social network were consistent while endorsing their peers, this 'endorsement with deduction' would not add anything to simple (i.e. ordinary) endorsement. In this ideal world, if Anna endorses Ben for the skill 'Java', she would be careful to endorse him for the skill 'Programming' as well.[5] In practice, however,

1. People are not systematic. That is, people do not usually go over all their contacts methodically to endorse, for each contact and alleged skill, all those contacts which, according to their opinion, deserve such endorsement. This may be the source of important omissions in members' profiles.
2. People are not consistent, for consistency, like method, would require a great effort. In an analysis of a small LinkedIn community consisting of 3250 members we have detected several inconsistencies. For example, there are several users who have been endorsed for some specific programming language, or a combination of programming languages, but have not been endorsed for the skill 'Programming'. Deciding whether there is an inconsistency entails some degree of subjectivism, for inconsistencies ultimately depend on the semantics of the skill names. Nevertheless, we can safely assert that practically 100% of the profiles sampled by us contained some evident inconsistency or omission. The Appendix lists some of the more significant inconsistencies and omissions encountered, together with a more comprehensive discussion about LinkedIn's endorsement mechanism.
3. Skills lack standardization. In most of these social networks, a set of standard, allowed skills has not been defined. As a result, many related skills (in many cases, almost synonyms) may come up in different profiles of the social network. Consider, for example, skills such as 'recruiting', 'recrutments', 'IT recruiting', 'internet recruiting', 'college recruiting', 'student recruiting', 'graduate recruiting', etc. which are, all of them, common in LinkedIn profiles. It may well happen that an expert in 'recruiting' has not even assigned to him/herself that specific skill, but a related one such as 'recruitments', which would hide him/her as an expert in the 'recruiting' skill.

Endorsement with deduction may help address those problems, and thus provide a better assessment of a person's skills. More precisely, we propose an algorithm that enriches the digraph of endorsements associated to a particular skill with new weighted arcs, taking into account the correlations between that 'target' skill and the other ones. Once this has been done, it is possible to apply different ranking algorithms to this enhanced digraph with the purpose of obtaining a ranking of the social network members concerning that specific skill.

1.1. Related work

This research can be inscribed into the discipline of *expertise retrieval*, a sub-field of information retrieval [1]. There are two main problems in expertise retrieval:

1. Expert finding: attempts to answer the question "*Who are the experts on topic X?*". In our approach, this question is answered by taking all the network members who are within a certain percentile of the ranking for topic *X*.
2. Expert profiling: addresses the question "*Which skills does person Y possess?*". We could answer this question by computing the rankings with respect to all the skills claimed by person *Y*, and taking those skills for which *Y* has fallen within the pre-defined percentile mentioned above.

Traditionally, these problems above have been solved via document mining, i.e. by looking for the papers on topic *X* written by person *Y*, combined with centrality or bibliographic measures, such as the H-index and the G-index, in order to assess the expert's relative influence (e.g. [29]). This is also the approach followed by ArnetMiner,[6] a popular web-based platform for expertise retrieval [45].

Despite their unquestionable usefulness, systems based on document mining, like ArnetMiner, face formidable challenges that limit their effectiveness. In addition to the specific challenges mentioned by Hashemi et al. [20], we could add several problems common to all data mining applications (e.g. name disambiguation). As a small experiment, we have searched for some known names in ArnetMiner, and we get several profiles corresponding to the same person, one for each different spelling.

That is one of the reasons why other expertise retrieval models resort to the power of PageRank in certain social networks, such as in the perused scientific citation and scientific collaboration networks (e.g. [10,20]). Another interesting example related to PageRank and social networks is TwitterRank [48], an extension of PageRank that measures the relative influence of Twitter users in a certain topic. Like our own PageRank extension, TwitterRank is topic-specific: the random surfer jumps from one user to an acquaintance following topic-dependent probabilities. However, TwitterRank does not consider any relationships among the different topics.

To the best of our knowledge, there are no precedents for the use of endorsements in social networks, nor for the use of known relationships among different skills, in the context of expertise retrieval. The closest approach might be perhaps the one in [41], which uses the ACM classification system as an ontology that guides the mining process and expert profiling. Another (very recent) model that uses semantic relationships to increase the effectiveness and efficiency of the search is given in [27].

Another related field which has attained a growing interest in the last few years is that of reputation systems, that is, systems intended to rank the agents of a domain based on others' agents reports. Strategies for ranking agents in a reputation system range from a direct ranking by agents (as used in eBay) to more sophisticated approaches (see [30] for a survey). One particularly important family of reputation system strategies is that of PageRank-based algorithms. There are many of such approaches. For instance, [8] provides an algorithm based on the so-called Dirichlet PageRank, which addresses problems such as: (1) some links in the network may indicate distrust rather than trust, and (2) how to infer a ranking for a node based on the ranking stated for a well-known subnetwork.

Another example of reputation system (again, based on PageRank) is one explained in [40]. In this case, a modification of the PageRank algorithm is used to create a reputation ranking among the members of an academic community. One remarkable issue of this approach is that the network does not exist explicitly, but it is created ad-hoc from the information harvested from the personal web pages of the members (e.g. a couple of members are connected if they have authored a research article together).

A thorough study of reputation systems is clearly beyond the scope of this article, but in any case, all these scenarios above differ significantly from our application for expertise retrieval with deduction of new endorsements, based on existing endorsements of related skills, and information about the correlation between skills.

[5] Some people may argue that knowledge of a programming language does not automatically imply programming skills, but this semantic discussion is out of the scope of this paper.

[6] http://www.arnetminer.org.