

# MultiRank: Reputation Ranking for Generic Semantic Social Networks

Xixi Luo  
Tetherless World Constellation  
Rensselaer Polytechnic Institute  
Troy, New York 12180  
luox5@rpi.edu

Joshua Shinavier  
Tetherless World Constellation  
Rensselaer Polytechnic Institute  
Troy, New York 12180  
shinaj@rpi.edu

## ABSTRACT

This paper presents a technique for calculating “reputation” or influence of users and artifacts in semantic social networks: in particular, as an incentive mechanism to encourage reuse of complex resources such as ontologies. Adapting the PageRank algorithm to the relational schemas of typical social network applications, this technique allows the programmer first to define via minimal rules the ways in which reputations of users and artifacts are likely to influence one another, then to obtain a mechanical, global ranking which reflects those rules in combination with the graph structure of the network. The mapping of multi-way relations such as usage and annotation to the binary-relational domain of PageRank is illustrated using the Actor-Concept-Instance model of ontologies. A lightweight software implementation,<sup>1</sup> currently under development, will provide a convenient way to add reputation-based functionality to Java-based community applications.

## Keywords

Semantic Web, social network, reputation, incentive, PageRank, relational model

## 1. INTRODUCTION

Collaborative tagging systems have achieved tremendous popularity in the form of online media-sharing communities such as Delicious,<sup>2</sup> Flickr,<sup>3</sup> and CiteULike.<sup>4</sup> This is true in spite of the well-known shortcomings of tagging, including ambiguities of natural language such as variations in spelling, pluralization and part of speech [5]. Some of these shortcomings can certainly be addressed by Semantic Web technologies: for instance, by substituting controlled vocabularies for folksonomies. However, the obvious success of tagging systems indicates that their advantages outweigh their lack

<sup>1</sup><http://multirank.googlecode.com/>

<sup>2</sup><http://delicious.com/>

<sup>3</sup><http://flickr.com/>

<sup>4</sup><http://citeulike.org/>

of clear semantics in many cases. On the other hand, the benefit of ontology-based annotation comes at the cost of significantly higher complexity. Emerging “Web 3.0” community applications such as Freebase<sup>5</sup> and various semantic wikis bridge this divide to some extent by providing a little more semantics than tagging systems, but a little more flexibility than typical ontology management tools. In such an environment, there is a need for effective incentive mechanisms to facilitate the complex task of building high-quality knowledge structures “from the bottom up”. One such mechanism is the implicit “reputation” of ranking systems, which suggests to users the “best”, most important, or most popular resources to use. The rest of this paper will focus on a specific ranking system, called MultiRank, which is based on an adaptation of the PageRank[3] algorithm to so-called semantic social networks.

### 1.1 The Actor-Concept-Instance model

As a minimal framework for semantic social networks, we will use a tripartite model of *actors* (human users or robots), *concepts* (tags, keywords, or possibly classes drawn from a controlled vocabulary) and *instances* (shared objects such as multimedia files, often contributed by actors themselves). All three elements are essential to the model: in a *semantic* social network, some notion of semantic annotation of instances with concepts is implied, whether this takes the form of simple folksonomy tagging or the sophisticated type system of a formal ontology. Furthermore, as the meaning of these artifacts is very much dependent on the context in which they are created and used [9], any measure of “reputation” should also take actors – authors, contributors – into account. This adds a social dimension to the otherwise bipartite model of traditional semantic networks (for example, of RDF graphs).

### 1.2 Multi-way relationships

We will use this tripartite model to illustrate the notion of *multi-way* relationships among actors, concepts and instances. Such a relationship may involve any number of elements, in contrast to the binary relationships of simple graphs. For example, the annotation of an instance – with a class – by an actor involves three distinct elements, and so cannot be completely represented by a simple binary edge. Multi-way relationships are common in database applications but complicate the otherwise simple recursive defini-

<sup>5</sup><http://www.freebase.com>

tion of PageRank,<sup>6</sup> and we claim that there is no single *right* way to map them into the binary-relational domain. Instead, we provide a generic framework with which an application programmer can define such mappings, in terms of patterns of ranking propagation between various pairs of elements in a relation.

## 2. REPUTATION FROM RELATIONSHIPS

Intuitively, *reputation* is a collective measure of trustworthiness in the estimation of the community [7]. A user’s reputation has both prescriptive and descriptive value: it is *prescriptive* in that it defines “good behavior” on the part of the user and thereby specifies the way in which users can gain reputation, and *descriptive* in that it provides a way to rank and classify users on the basis of their reputations [1]. The reputation of *resources* identifies the “best” or most important resources and thereby singles them out as candidates for imitation or reuse. For the purpose of this paper, reputation is an *implicit* statement of trustworthiness: much like the original formulation of PageRank, MultiRank is an attempt to measure human interest and attention based on the network of relationships within the community. Such an approach holds the possibility of making minimal demands on the user, while scaling well and delivering subjectively accurate results despite a high degree of heterogeneity in the quality and structure of the network.

### 2.1 Propagation of ranking

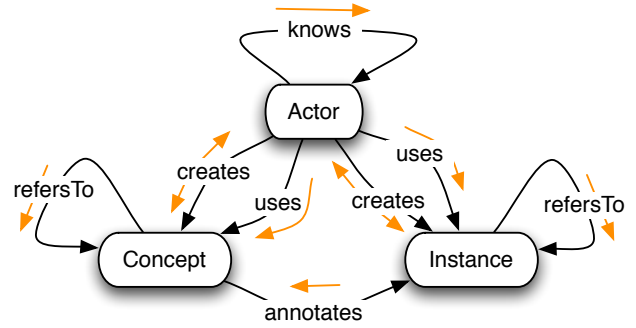
The notion of propagation of ranking (here: reputation) through directed edges is the basis of the PageRank algorithm: if the sum of the ranking of the nodes with edges to a given node is high, then the ranking of the node itself should be high. The contribution of MultiRank is in the construction of a “virtual” binary-relational graph  $G_{prop}$ , on which to run PageRank in order to derive reputation values. The nodes of this graph are the the actors, concepts and instances of the semantic social network, while its edges are chosen so as to propagate ranking from node to node in a way that reflects the intended or expected flow of reputation within the network. In general, the reputation of an actor or artifact tends to increase the reputation of another item with which it associates (for instance, by “creating”, “using”, “knowing”, or otherwise drawing attention to that item).

In the following, we list several informal and intuitive “rules” for the flow of reputation ranking in a hypothetical semantic social network (see Figure 1), with which we motivate the idea of *propagation patterns* defined in the next section:

1. If actor  $a_1$  knows actor  $a_2$ , then  $a_1$ ’s reputation should propagate to  $a_2$ . This rule reflects the fact that an actor benefits from being known by other, high-reputation actors. Note that we’ve chosen to let reputation propagate in only one direction. In this respect, our social network resembles an environment like Twitter,

<sup>6</sup>The PageRank of a node  $p$  in an unweighted, directed graph is generally given by  $PR(p) = \frac{1-d}{N} + \sum_{q \in B(p)} \frac{PR(q)}{N_q}$ , where  $N$  is the total number of nodes,  $B(p)$  is the set of nodes  $q$  with edges to  $p$ ,  $N_q$  is the out-degree of  $q$  and  $d$  is a *decay factor* which helps to dampen the effect of graph cycles. Weighted PageRank is defined similarly. Note that the expression  $\frac{PR(q)}{N_q}$  represents the ranking of the edge  $(p, q)$ .

Figure 1: Flow of reputation in an example network



in which it is possible to “follow” a popular individual without that individual necessarily following you.

2. If actor  $a$  creates artifact  $t$ , then  $a$ ’s reputation should propagate to  $t$ , and vice versa. This reflects the fact that artifacts created by an actor with a high reputation are to some extent *authoritative*: their reputations benefit from association with their creator. Conversely, an actor’s reputation benefits (perhaps to an even greater extent) from association with artifacts she has created which have achieved a high reputation.
3. If actor  $a$  uses artifact (concept or instance)  $t$ , then  $a$ ’s reputation should propagate to  $t$ . This is a somewhat weaker version of rule 2: an artifact should gain some reputation through a high-reputation actor who has associated himself with it. However, the reverse is not true: merely associating oneself with a great resource does not make one great.
4. If concept  $c$  annotates instance  $i$ , then reputation should propagate “backwards” from  $i$  to  $c$ , the direction of *annotates* being unimportant. Although each of these rules is debatable, we imagine the annotated instance to draw attention – and thus reputation – to the annotating concept, but not the reverse.
5. If artifact  $t_1$  refers to or artifact  $t_2$  or , then  $t_1$ ’s reputation should propagate to  $t_2$ . Again, a high-reputation item should increase the reputation of other items with which it associates. If the artifacts happen to be web pages and the links happen to be hyperlinks, then this rule is particularly close to ordinary PageRank.

### 2.2 Mapping to a binary-relational network

Now that we have an intuitive idea of propagation of ranking in the virtual network, let us formally describe the derivation of that network from a collection of multi-way relationships. For the purpose of clarity, we will introduce the notions of terms, variables, and bindings, in analogy to the SPARQL[11] query language and to relational databases. In the following, a *term*  $t \in T$  is any item in the social network (be it an actor, concept or instance), a *variable*  $v \in V$  is an abstraction which may be replaced with a term, and a *binding*  $b \in V \times T$  is a pair which connects a variable to a term. Furthermore, a *relation* is an abstract relationship among variables which carries a particular meaning. For the purpose of calculating reputation, we reduce that meaning to a

**Table 1: Propagation patterns for example relations**

Relation	Variables	Propagation	Rules
actor knows actor	$\{a_1, a_2\}$	$(a_1, a_2, 0.6)$	1
actor creates concept	$\{a, c\}$	$(a, c, 0.4)$ $(c, a, 1.0)$	2
actor creates instance	$\{a, i\}$	$(a, i, 0.4)$ $(i, a, 1.0)$	2
actor uses concept to annotate instance	$\{a, c, i\}$	$(a, c, 0.2)$ $(a, i, 0.2)$ $(i, c, 0.8)$	3, 4
concept refers to concept	$\{c_1, c_2\}$	$(c_1, c_2, 0.6)$	5
instance refers to instance	$\{i_1, i_2\}$	$(i_1, i_2, 0.6)$	5

pattern of pairwise propagation of ranking among variables. For example, the *actor uses concept to annotate instance* relation in Table 1 involves three variables and propagates ranking among them according to rules 3 and 4 from the preceding section. Specifically, it propagates small amounts of ranking from the actor to the concept and to the instance (which the actor *uses*) and a larger amount of ranking from the instance to the concept (which *annotates* the instance). In general, we define a relation  $R \subseteq V \times V \times \mathbb{R}$  as a set of abstract weighted edges between pairs of variables. Such an edge represents a type of path along which ranking is permitted to flow, while its weight allows fine-grained control over the *extent* of flow. This set of abstract edges is a *pattern* of ranking propagation which is to be pre-defined by the programmer for each relation.

By combining a relation  $R$  with a set  $B$  of bindings, we obtain a multiset of weighted edges which propagate ranking among specific terms in  $T$ :

$$\begin{aligned} PropEdges(R, B) = \{ & (t_1, t_2, w) : (\exists v_1, v_2 \in V, w \in \mathbb{R}) \\ & ((v_1, t_1) \in B) \\ & ((v_2, t_2) \in B) \\ & ((v_1, v_2, w) \in R) \} \end{aligned}$$

That is, the set of propagation edges for  $R$  and  $B$  are those which can be formed by replacing variables in the abstract edges of  $R$  according to the variable-to-term pairs in  $B$ . This combination of an abstract relationship with specific terms has a counterpart in SPARQL queries, in which variables in a query are resolved to specific terms in its solution, and in relational databases, in which column names resolve to specific cells in rows of data. It is easy to imagine a process which iterates through all of the results of a query or all of the rows in a table, applying *PropEdges* to each result or row to generate a graph of all possible propagation edges:

$$G_{prop} = (T, \bigoplus_{(R,B) \in S} PropEdges(R, B))$$

Here,  $S$  represents the set of all relation-binding pairs, or concrete semantic relationships which make up the social network. Depending on the application, these pairs may be drawn from a SPARQL query, a SQL database, or any other source of tabular, relational data.  $G_{prop}$ , then, is the weighted, directed multigraph of propagation edges generated by means of  $S$  over the set  $T$  of terms. This is the graph on which we will actually run PageRank to derive ranking results.

## 2.3 Applying PageRank

The propagation graph  $G_{prop}$  is an intermediate result in our calculation of ranking. To derive a final result, we have only to apply a weighted form of PageRank to this graph. The fact that  $G_{prop}$  is a *multigraph* presents no additional challenges: to transform it into an ordinary weighted graph  $G'_{prop}$ , we simply merge parallel edges, adding their weights together. Formally, we compute PageRank by iteratively solving for the vector  $\pi \in \mathbb{R}^{|T|}$  in:

$$\pi = (1 - d)E + dG'_{prop}\pi$$

where  $d$  is the decay factor (typically chosen to be 0.85) and  $E \in \mathbb{R}^{|T|}$  is a vector representing a source of ranking. If  $E$  is uniform over all  $t \in T$ , then the resulting  $\pi$  is a *global* measure of reputation in our semantic social network. However, by biasing  $E$  in favor of particular terms, any number of so-called *personalized*[10] PageRanks can be applied. Computing a personalized PageRank ranking (biased, for example, towards actors, concepts and instances which are *trusted* by a particular user) over  $G_{prop}$  brings MultiRank closer in spirit to trust-based mechanisms in recommendation systems [2] and shared content repositories [8].

## 3. INCENTIVE FROM REPUTATION

Our technique has been thoroughly described in the sections above. Having once computed the “reputation” vector  $\pi$ , the application is free to use it in application-specific ways. For instance, an ordering of actors by decreasing reputation can be used as a “Top X” list to which actors may aspire, raising the quality of the social network in the process. Similarly, actors may strive to get their own artifacts into “best of” lists of various kinds. These rankings, in turn, may help to ensure that the top actors in the network get the attention they deserve, and that the top artifacts, such as the elements of well-designed ontologies, are consistently re-discovered and re-used. Although in this paper we have focused on the ranking technique itself, we believe that the prescriptive value of subjectively accurate ranking results is an ample foundation for incentive mechanisms to motivate users to improve their own resources and connections.

## 4. RELATED WORK

There have been a number of Semantic Web tools which make use of PageRank. For example, the Swoogle search engine’s OntoRank[4] is a variation of PageRank for ontologies. OntoRank takes a number of types of semantic links into account when calculating ranking, weighting links selectively according to these types. Similarly, the Semantic Web Search Engine’s ReConRank[6] extends a graph of RDF resources with contextual edges, forming a compound graph which includes relevant provenance information. However, both of these technologies operate upon existing binary-relational semantic networks, whereas MultiRank is designed for relations which are not necessarily binary, introducing the notion of propagation patterns to first construct a “virtual” binary-relational network before applying PageRank to it. This technique was motivated by the notion of semantic-social hypergraphs in Peter Mika’s tripartite ontology model, and builds upon previous work [12] in applying single-relational network analysis algorithms to multi-relational networks.

## 5. CONCLUSION AND FUTURE WORK

We have presented a PageRank-based reputation ranking system for semantic social networks, illustrating it with an actor-concept-instance model. The technique itself is very general: we make only the basic assumption that the structure of the network can be represented as a collection of multi-way relationships, such as the solution to one or more SPARQL queries or the contents of one or more tables in a relational database. MultiRank borrows from PageRank the implicit *reputation* of resources as expressed in network structure alone, while it adds application-specific *propagation patterns* which direct the flow of reputation according to a human's intuitive understanding of the social network. The algorithm proceeds in two stages: a loading stage in which relational data is processed row by row to derive a virtual binary-relational graph, and a computational stage in which the PageRank algorithm is applied to the virtual graph to generate ranking results. Due to the simplicity of the model and favorable performance characteristics of PageRank, we believe that a software implementation of MultiRank will provide a cheap and effective way to add reputation-based functionality to any of a variety of semantic social networks. Such an implementation is currently under development, building on the open-source Java Universal Network/Graph Framework (JUNG).<sup>7</sup> We intend to test this software in the near future using more than one social network data set, including a large dump of Freebase event logs. This will help us to estimate performance and memory usage, as well as to gauge the sensitivity of the computed ranking results with respect to the *weight* values of propagation patterns in different application scenarios. At that point, we should be ready to deploy and evaluate MultiRank-based incentive mechanisms in a live semantic social network environment.

## 6. ACKNOWLEDGEMENTS

This work has been supported by Rensselaer's Tetherless World Constellation. We would also like to thank Jim Hendler, Deborah McGuinness, Marko A. Rodriguez and Li Ding for their valuable feedback on various drafts of this paper.

## 7. REFERENCES

- [1] Thomas B. Adler and Luca de Alfaro, *A content-driven reputation system for the Wikipedia*, WWW '07: Proceedings of the 16th international conference on World Wide Web (New York, NY, USA), ACM Press, 2007, pp. 261–270.
- [2] Reid Andersen, Christian Borgs, Jennifer Chayes, Uriel Feige, Abraham Flaxman, Adam Kalai, Vahab Mirrokni, and Moshe Tennenholtz, *Trust-based recommendation systems: an axiomatic approach*, WWW '08: Proceeding of the 17th international conference on World Wide Web (New York, NY, USA), ACM, 2008, pp. 199–208.
- [3] Sergey Brin and Lawrence Page, *The anatomy of a large-scale hypertextual web search engine*, Computer Networks and ISDN Systems **30** (1998), no. 1–7, 107–117.
- [4] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, Scott R. Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs, *Swoogle: a search and*

*metadata engine for the Semantic Web*, CIKM '04: Proceedings of the thirteenth ACM conference on Information and knowledge management (New York, NY, USA), ACM Press, 2004, pp. 652–659.

- [5] S. Golder and B. Huberman, *The structure of collaborative tagging systems*, Journal of Information Science **32** (2006), no. 2, 198–208.
- [6] Aidan Hogan, Andreas Harth, and Stefan Decker, *ReConRank: A scalable ranking method for Semantic Web data with context*, In 2nd Workshop on Scalable Semantic Web Knowledge Base Systems, 2006.
- [7] Audun Josang, Roslan Ismail, and Colin Boyd, *A survey of trust and reputation systems for online service provision*, Decision Support Systems **43** (2007), no. 2, 618–644.
- [8] Deborah L. McGuinness, Honglei Zeng, Paulo P. da Silva, Li Ding, Dhyanes Narayanan, and Mayukh Bhaowal, *Investigations into trust for collaborative information repositories: A Wikipedia case study*, MTW (Tim Finin, Lalana Kagal, Daniel Olmedilla, Tim Finin, Lalana Kagal, and Daniel Olmedilla, eds.), CEUR Workshop Proceedings, vol. 190, CEUR-WS.org, 2006.
- [9] Peter Mika, *Ontologies are us: A unified model of social networks and semantics*, International Semantic Web Conference, LNCS, Springer, 2005, pp. 522–536.
- [10] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, *The PageRank citation ranking: Bringing order to the Web*, Tech. report, Stanford Digital Library Technologies Project, 1998.
- [11] Eric Prud'hommeaux and Andy Seaborne, *SPARQL query language for RDF*, Tech. report, W3C, January 2008.
- [12] Marko A. Rodriguez and Joshua Shinavier, *Exposing multi-relational networks to single-relational network analysis algorithms*, Tech. Report LA-UR-08-03931, Los Alamos National Laboratory, 2008.

<sup>7</sup><http://jung.sourceforge.net/>