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 relationships 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, 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,2 Flickr,3 and CiteULike.4 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 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 Freebase5 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-
tion of PageRank, 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 \( G_{\text{prop}} \), 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 descriptive 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.

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 annotation 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

---

\[ PR(p) = \frac{1-d}{N} + \sum_{q \in B(p)} \frac{PR(q)}{N_q} \]

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 \((p,q)\).
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 $G'$, 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:

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

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).7 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

7http://jung.sourceforge.net/