TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 04/14 pp642-649
Volume 15, Number 6, December 2010

A Novel Ranking Framework for Linked Data from Relational Databases

ZHANG Jing (张 静)**, MA Chune (马春娥), ZHAO Chenting (赵晨婷), ZHANG Jun (张 军), YI Li (易 力), MAO Xinsheng (毛新生)

IBM China Development Laboratory, Beijing 100193, China

Abstract: This paper investigates the problem of ranking linked data from relational databases using a ranking framework. The core idea is to group relationships by their types, then rank the types, and finally rank the instances attached to each type. The ranking criteria for each step considers the mapping rules and heterogeneous graph structure of the data web. Tests based on a social network dataset show that the linked data ranking is effective and easier for people to understand. This approach benefits from utilizing relationships deduced from mapping rules based on table schemas and distinguishing the relationship types, which results in better ranking and visualization of the linked data.

Key words: linked data; ranking; relational databases (RDB); mapping rules

Introduction

A critical requirement for the evolution of the current web of documents into a web of data (and ultimately a semantic web) is the inclusion of the vast quantities of data stored in relational databases (RDB)[1]. Companies have much data stored in RDB. Many companies have already realized the importance of linked data and publish their data into linked databases such as the CIA Factbook (http://www4.wiwiss.fu-berlin.de/factbook), the Colibrary Web API (http://collab.di.uniba.it/Colibrary/books), and the Linked Movie DataBase (http://www.linkedmdb.org). Linked data from a RDB usually has a simple browsing end point. When users click the URI of a resource, most browsers usually just display all its relationships, with little organizing and ranking of the relationships. Some popular resources may have hundreds of related resources, so users cannot easily find interesting resources since they are buried in the large number of relationships. To improve the usability, the most urgent need for a linked data browser is to effectively organize and rank the large number of relationships. Although linked data can come from many sources, this first stage will focus on ranking linked data from RDB due to the large number of such databases.

One important feature of linked data from RDB is that the relationships are usually deduced from mapping rules (http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/) defined based on table schemas. These rules may reflect the relevance or importance of the relationships. This means that the ranking in linked data from RDB will be quite different from that for data from other linked data sources. In addition, linked data from RDB also has the general feature of linked data based on relationships of different types. Unlike homogeneous linkages, hyperlinks in the data web can have several types of relationships. For example person A has a friend type of relationship with person B and has a relative type of relationship with person C. Thus, the ranking criteria in linked data differs greatly from the document web. Therefore, a method is needed to rank heterogeneous relationships using leveraging mapping rules in linked data from RDB.

Received: 2010-09-16; revised: 2010-10-12
** To whom correspondence should be addressed.
E-mail: zhjingbj@cn.ibm.com; Tel: 86-13810654581

The paper describes a novel framework for ranking heterogeneous relationships in linked data sourcing from RDB. The framework groups and ranks relationships based on their types, and then ranks relationship instances attached to each type. Most users will firstly be attracted by the relationship types, and then move down to the resources belonging to their interested type. The relationship type ranking measures their relevance by considering the mapping rules defined for relationships and the user click-through effect. The ranking of relationships belonging to different types not only considers the relevance between sourcing and targeting resources based on the mapping rules, but also considers the global popularity of the targeting resources in the heterogeneous data web.

Previous studies have analyzed the relationships in linked data. Harth et al.[2] proposed ranking individual resources by calculating the authority of the data sources. ReConRank[3] is an online PageRank based on a topical subgraph for ranking resources related to input keywords. Triple Rank[4] ranks linked data by tensor decomposition. However, no one has investigated the special features of linked data sourcing from RDB, which are a critical source of linked data. SemRank[5] and Context-aware ranking[6] both rank the associations between two given resources. This study ranks all the resources related to one given resource. Ding et al.[7] ranked rdf documents on terms for searching scenarios. Alani et al.[8] ranked ontologies. Toupikov and Umbrich[9] ranked linked data sets. To our knowledge, no one has investigated linked data sourcing from RDB, which is a critical source of linked data.

1 Motivating Example

The motivating example is an academic social network ArnetMiner (http://www.arnetminer.org), which describes papers, conferences, authors, and their relationships. Figure 1 shows the concept model and Fig. 2 shows the table schema in ArnetMiner.

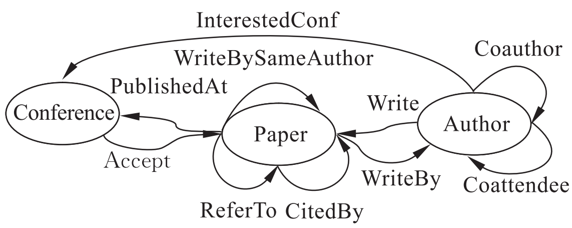

Fig. 1 Concept model in ArnetMiner

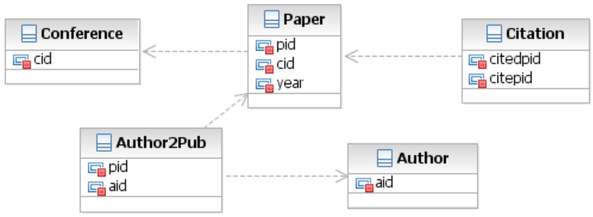

Fig. 2 Table schemas in ArnetMiner

The algorithm first generates a default mapping file using D2R tools (http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/) (which is the most popular tool for publishing RDB into linked data) based on the ArnetMiner DB schema (see Fig. 2). The foreign keys are transformed directly into relationships, e.g., Write (author A writes paper B), PublishedAt (paper B is published at conference A), and CitedBy (paper A is cited by paper B), then the mapping file is manually modified by adding other relationships according to the concept model (see Fig. 1), e.g., Coauthor (author A and author B co-write a paper), Coattendee (author A and author B attend the same conferences), and InterestedConf (author A is interested in conference B). Coauthor is defined by a D2R mapping (a declarative language for describing the relation ships between RDB schemas and the RDF vocabulary or OWL ontologies), which is shown in Fig. 3.

```
map:Coauthor a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:author;
  d2rq:property vocab:Coauthor;
  d2rq:refersToClassMap map:author;
  d2rq:join "author.aid = author2pub.aid";
  d2rq:join "author2pub.pid = author2pub2.pid";
  d2rq:join "author2pub2.aid = author2.aid";
  d2rq:alias "author2pub AS author2pub2";
  d2rq:alias "author AS author2";
  d2rq:condition "author.aid <> author2.aid"
```

Fig. 3 D2R mapping for Coauthor

A simple linked data browser was then built as in Fig. 4. The figure shows the representation for author Stefan Decker (identified by URI http://localhost/resource/person/19222), with the literal properties and values listed on the left, with object properties and their objects on the right. Object properties are treated as relationships of the sourcing resource to the targeting resources. There are many relationships for the resource Stefan Decker. However, why position WritePaper before Coauthor? Why position person 2558 at first for Coauthor? It will benefit users a lot if more relevant and popular relationships are ranked first.

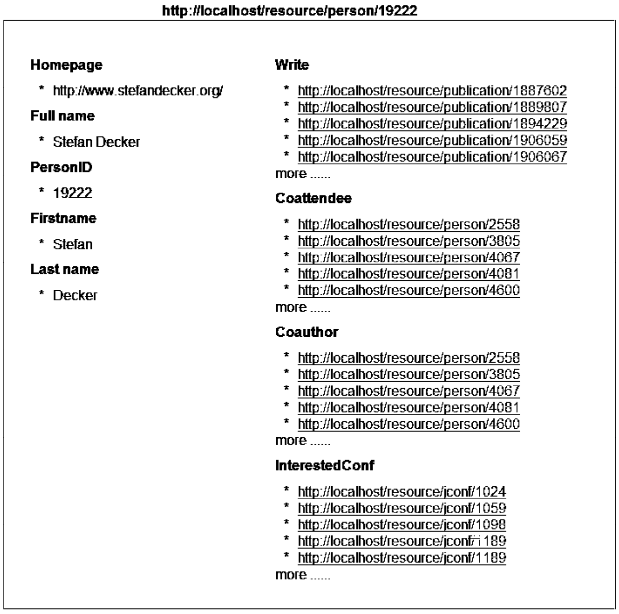

Fig. 4 Representation of author Stefan Decker

A unique feature of this kind of data is that relationships are usually deduced from mapping rules. These mapping rules may reflect the relevance or importance of relationships, which can drive the different ranking strategies from other linked data sources. Therefore, the ranking problem for linked data from RDB can be formalized as:

Problem 1 (Relationship ranking) Given one resource s and all its associated relationships $\{r\}$, with each relationship r attached to one type $t \in T$ (e.g., WritePaper or Coauthor) and each type t attached to a D2R mapping rule u, the objective is to define a ranking strategy F to rank $\{r\}$.

2 Ranking Framework

2.1 Framework overview

Figure 4 shows that one resource may contain different types of relationships and each type may have a number of relationship instances. These heterogeneous relationships are organized and ranked using a two-level ranking framework, which first defines a strategy F^t for ranking the relationship types $\{t\}$, and then defines a strategy F^r for ranking the relationship instances $\{r\}$ associated with each type t.

Most people organize things from concept to instance and from abstract to concrete. Only if users are attracted by interesting relationship types will they drill down to the resource instances.

2.2 Ranking relationship types

D2R tools are used to declare a relationship type based on a D2R mapping rule. The mapping rules for the relationship types shown in Fig. 4 are given in Figs. 5-7 (where Coauthor is illustrated in Fig. 3):

```
map:Coattendee a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:author;
d2rq:property vocab: Coattendee;
d2rq:refersToClassMap map:author;
d2rq:join "author.aid = author2pub.aid";
d2rq:join "author2pub.pid = paper.pid";
d2rq:join "paper.cid = paper2.cid";
d2rq:join "paper2.pid = author2pub2.pid";
d2rq:join "author2pub2.aid = author2.aid";
d2rq:alias "author2pub AS author2pub2";
d2rq:alias "paper AS paper2";
d2rq:alias "author AS author2";
d2rq:condition "author.aid <> author2.aid";
```

Fig. 5 D2R mapping for Coattendee

```
map:InterestedConf a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:author;
d2rq:property vocab: InterestedConf;
d2rq:refersToClassMap map:conference;
d2rq:join "author.aid = author2pub.aid";
d2rq:join "author2pub.pid = paper.pid";
d2rq:join "paper.cid = conference.cid";
```

Fig. 6 D2R mapping for InterestedConf

```
map:Write a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:author;
d2rq:property vocab: Write;
d2rq:refersToClassMap map:paper;
d2rq:join "author2pub.aid = author.aid";
d2rq:join "author2pub.pid = paper.pid";
```

Fig. 7 D2R mapping for Write

The relationships are ranked by their relevance to the sourcing resource. The relevance can be measured by (1) the influence of concept hops and (2) the influence of the number of user click throughs.

Concept hops Concept hops indicate how many concepts a relationship type contains, e.g., Coattendee (author A and author B attend same conferences) is defined as "author→paper→conference→paper→author", which contains five concept hops and Write (author A writes paper B) is defined as "author→paper", which contains two concept hops. Concept hops can be easily reflected by the number of d2rq:join clauses in the mapping file. A small number of hops in a relationship type indicate a more explicit relationship. In general, people prefer more explicit relationships, like Write and Coauthor. However, the system cannot exclude the situation where users wish to get hidden or indirect paths, for example, terrorist cells will remain distant and avoid direct contact with one another to avoid possible detection[6]. In this case, users may prefer more implicit relationship types containing more concept hops. Hence, the influence of concept hops I_t^{ch} is defined as follows:

$$I_t^{\text{ch}} = \frac{1}{|l|+1}, \quad I_t^{\text{ch}'} = 1 - \frac{1}{|l|+1} \tag{1}$$

where $|l|$ indicates the number of d2rq:join associated with t. Users can select the equations for explicit or implicit relationships for their interest. The relevance of the four relationship types in Fig. 4 can be calculated using I_t^{ch} as:

$$I_{\text{Write}}^{\text{ch}} = \frac{1}{2}, I_{\text{Coauthor}}^{\text{ch}} = \frac{1}{3}, I_{\text{InterestedConf}}^{\text{ch}} = \frac{1}{3}, I_{\text{Coattendee}}^{\text{ch}} = \frac{1}{5}.$$

Alternatively, the relevance of the four relationship types can be calculated using $I_t^{\text{ch}'}$ as:

$$I_{\text{Write}}^{\text{ch}'} = \frac{1}{2}, I_{\text{Coauthor}}^{\text{ch}'} = \frac{2}{3}, I_{\text{InterestedConf}}^{\text{ch}'} = \frac{2}{3}, I_{\text{Coattendee}}^{\text{ch}'} = \frac{4}{5}.$$

Number of user click throughs The relevance of the relationship types can also be influenced by the number of user click throughs. More clicks on the instances for a relationship type indicate the sourcing resource is more relevant. The influence of the number of click throughs is

$$I_t^{\text{click}} = |N_t| \times \left(- \sum_{r:\,\text{type}(r)=t \,\&\, n_r>0} \frac{|n_r|}{|N_t|} \log \frac{|n_r|}{|N_t|} + 1 \right) \tag{2}$$

where $|N_t|$ indicates the total click-through number of all the relationship instances associated with the relationship type t, $|n_r|$ indicates the number of click throughs of instance r associated with type t, type(r) $= t$ means that the type of instance r is t and $n_r > 0$ means that the system only states the instances with the number of click throughs larger than 0. The equation includes two parts, where the left part $|N_t|$ indicates that more clicks indicate more relevance of the relationship type while the right part in parenthesis represents the click diversity[10] of the relationship type. If more diverse instances belonging to one relationship type are clicked, the relationship type is more relevant. This assumption is based on the intuition that people usually prefer concepts with diverse instances preferred by many people, rather than a single instance. The plus 1 in the parenthesis prevents a zero diversity from leading to a zero click-through influence.

Taking the author Stefan Decker for example with the click-through data in Table 1:

Table 1 Click-through data for Stefan Decker's relationships

Relationship type	Relationship instances	Click number
Writer	http://localhost/resource/publication/1889807	5
	http://localhost/resource/publication/1906067	4
	http://localhost/resource/publication/2067790	3
Coauthor	http://localhost/resource/person/3805	10
	http://localhost/resource/person/4081	1
	http://localhost/resource/person/4600	1
Coattendee	http://localhost/resource/person/1133	6
	http://localhost/resource/person/11253	2
InterestedConf	http://localhost/resource/jconf/1098	3

The relevance influenced by the number of click throughs is calculated from Eq. (2) as:

$$I_{\text{write}}^{\text{click}} = 12 \times (0.468 + 1) = 17.616,$$

$$I_{\text{coauthor}}^{\text{click}} = 12 \times (0.246 + 1) = 14.952,$$

$$I_{\text{coattendee}}^{\text{click}} = 8 \times (0.244 + 1) = 9.952,$$

$$I_{\text{interestedConf}}^{\text{click}} = 3 \times (0 + 1) = 3.$$

Although the total number of clicks for Coauthor is the same for Write, the click distribution on instances is more diverse for Write than for Coauthor, which indicates that the type Write contains diverse instances preferred by users, so the final click-through number influence for Write is larger than for Coauthor.

2.3 Ranking relationship instances

One relationship type may be associated with many relationship instances. The ranking of relationship instances actually should rank the targeting resources of relationship instances. The targeting resources are ranked by the relevance influenced by the number of real paths deduced from the D2R mapping rule defined for the relationship type and the global popularity of the targeting resources in a heterogeneous graph.

Path number influence Each relationship type t is associated with a D2R mapping rule as in Figs. 3, 5-7. A rule is used to deduce real paths from the sourcing resource to each targeting resource. The ranking intuition is that more real paths deduced for one targeting resource indicate more relevance of the targeting resource to the sourcing resource. Take Coauthor for example (author is written as a and author2pub as ap):

$$\begin{aligned}&\text{a.aid}_1 = \text{ap.aid}_1 \to \text{ap.pid}_1 = \text{ap2.pid}_1 \to \\ &\quad \text{ap2.aid}_2 = \text{a2.aid}_2,\\ &\text{a.aid}_1 = \text{ap.aid}_1 \to \text{ap.pid}_2 = \text{ap2.pid}_2 \to \\ &\quad \text{ap2.aid}_2 = \text{a2.aid}_2,\\ &\text{a.aid}_1 = \text{ap.aid}_1 \to \text{ap.pid}_3 = \text{ap2.pid}_3 \to \\ &\quad \text{ap2.aid}_2 = \text{a2.aid}_2,\\ &\text{a.aid}_1 = \text{ap.aid}_1 \to \text{ap.pid}_1 = \text{ap2.pid}_1 \to \\ &\quad \text{ap2.aid}_3 = \text{a2.aid}_3,\\ &\text{a.aid}_1 = \text{ap.aid}_1 \to \text{ap.pid}_4 = \text{ap2.pid}_4 \to \\ &\quad \text{ap2.aid}_3 = \text{a2.aid}_3,\\ &\text{a.aid}_1 = \text{ap.aid}_1 \to \text{ap.pid}_5 = \text{ap2.pid}_5 \to \\ &\quad \text{ap2.aid}_4 = \text{a2.aid}_4,\end{aligned}$$

where the equation using "$\to$" (e.g., a.aid_1=ap.aid_1) is deduced from one d2rq:join clause.

The Coauthor rule defined in Fig. 3 leads to three real paths from author_1 to author_2, two paths from author_1 to author_3, and one path from author_1 to author_4. Then author_2 should be more relevant to author_1 than to author_3 or to author_4, because author_2 coauthors more papers with author_1. Therefore, the equation for calculating the path number influence (PNI) is defined as:

$$\text{PNI}_{ti} = \frac{|p_{ti}|}{|P_t|} \tag{3}$$

where PNI_{ti} indicates the influence of the path number to the i-th targeting resource of a relationship type t, $|p_{ti}|$ indicates the number of paths from the sourcing resource to the i-th targeting resource with type t and $|P_t|$ is the total number of paths from the sourcing resource to all the other resources with type t. Suppose there are a total of three targeting resources associated with type Coauthor, the PNI of the three authors can be calculated as:

$$\text{PNI}_1 = \frac{3}{6} = 0.5, \text{PNI}_2 = \frac{2}{6} = 0.3, \text{PNI}_2 = \frac{1}{6} \approx 0.167.$$

The path for some relationship types contain only one "$\to$", e.g., for relationship type Write:

$$\begin{aligned}&\text{author2pub.aid}_1 = \text{author.aid}_1 \to \\ &\text{author2pub.pid}_1 = \text{paper.pid}_1,\\ &\text{author2pub.aid}_1 = \text{author.aid}_1 \to \\ &\text{author2pub.pid}_2 = \text{paper.pid}_2.\end{aligned}$$

In this case, each targeting resource can be deduced from only one real path. PNI is then the same for every resource belonging to Write. Therefore, the targeting resources should be ranked by other criteria, e.g., the "year" of the paper. The real projects need an administration function to configure this kind of ranking criteria.

Global popularity in heterogeneous graphs Besides the influence of the number of paths on the relevance, the global popularity of a targeting resource also indicates the relevance. A resource is more popular when the global popularity is higher, which is the number of times a resource is linked by other resources. This idea is the same as Google's PageRank[11], except that in linked data, the relationships are more heterogeneous than the homogeneous hyperlinks in the document web. The weighting matrix in PageRank reflects the probability that web surfers follow from one web page to one of its outlinking pages. Since the hyperlink is homogeneous, the weighting is averaged among all the outlinks. However, in the data web, the weighting is set according to the different relationship types.

The weighting strategy to calculate the relevance of relationship types can be used to set the weights. The number of paths for relationship instances also affects the weighting.

Then weighting can be calculated by

$$w_{ij} = \sum_{t \in \text{Type}_{ij}} I_t \cdot \text{PNI}_{tj} \tag{4}$$

where w_{ij} indicates the probability of users navigating from resource i to j, Type_{ij} is the collection of relationship types from i to j, and t is one of the types. I_t is the influence of relationship type t (normalized to [0,1)), which is calculated by combining the concept hop and click-through number influence. PNI_{tj} is the influence of the number of paths from resource i to j belonging to type t.

The theory used in PageRank shows that the weighting of the links coming from one node should add up to 1, i.e., $\sum_j w_{ij}=1$. This condition guarantees convergence of the algorithm. The current weighting calculation also satisfies this condition:

$$\sum_{j\in \mathrm{Target}_i} w_{ij} = \sum_{j\in \mathrm{Target}_i}\left(\sum_{t\in \mathrm{Type}_{ij}} I_t \cdot \mathrm{PNI}_{ij}\right) =$$
$$\sum_{t\in \mathrm{Type}_{ij}} I_t \cdot \sum_{j\in \mathrm{Target}_{it}} \mathrm{PNI}_{ij} = \sum_{t\in \mathrm{Type}_{ij}} I_t \cdot 1 = 1 \qquad (5)$$

where Target_i is the collection of out-linking resource instances of i and Target_{it} is the collection of out-linking resource instances of i belonging to type t. Then the weighting method in a heterogeneous data web satisfies the PageRank convergent condition.

Consider the graph in Fig. 8:

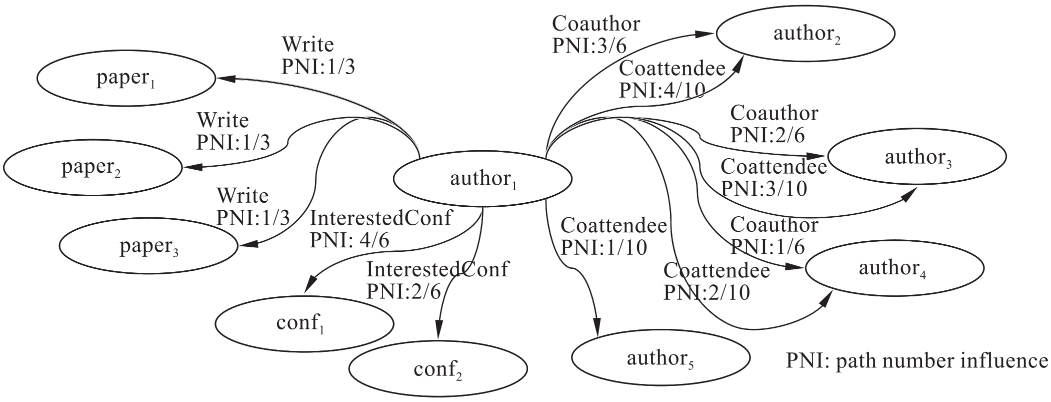

Fig. 8 Heterogeneous graph of academic data web

Suppose the influence scores for each type are Writer: 0.5, Coauthor: 0.3, Coattendee: 0.1, and InterestedConf: 0.1, the weightings are then as listed in Table 2.

Table 2 Weightings calculated for graph in Fig. 8

Relationship instance from i to j	w_{ij}
$\mathrm{author}_1 \rightarrow \mathrm{author}_2$	0.3×3/6+0.1×4/10=0.190
$\mathrm{author}_1 \rightarrow \mathrm{author}_3$	0.3×2/6+0.1×3/10=0.130
$\mathrm{author}_1 \rightarrow \mathrm{author}_4$	0.3×1/6+0.1×2/10=0.069
$\mathrm{author}_1 \rightarrow \mathrm{author}_5$	0.1×1/10=0.010
$\mathrm{author}_1 \rightarrow \mathrm{paper}_1$	0.5×1/3=0.167
$\mathrm{author}_1 \rightarrow \mathrm{paper}_2$	0.5×1/3=0.167
$\mathrm{author}_1 \rightarrow \mathrm{paper}_3$	0.5×1/3=0.167
$\mathrm{author}_1 \rightarrow \mathrm{conf}_1$	0.1×4/6=0.067
$\mathrm{author}_1 \rightarrow \mathrm{conf}_2$	0.1×2/6=0.033

The result shows that the sum of w_{ij} is equal to one. The global popularity is then calculated using the PageRank algorithm:

$$p_i = (1-\alpha)\cdot\frac{1}{|V|} + \alpha\cdot\sum_{j\in \mathrm{Target}_i} w_{ij}\cdot p_j \qquad (6)$$

where p_i is the global probability of resource i, $|V|$ indicates the total number of resource instances in the graph, and α represents the damping factor, usually being set to 0.85 as in PageRank.

The final ranking score of a targeting resource is calculated by combining the path number influence and the global popularity. The combination method can be a linear combination or some products. This will not be discussed here.

Nie et al.[12] also proposed an algorithm to calculate the popularity of objects in a heterogeneous graph, but they only considered the influence of relationship types and did not distinguish instances attached to the same type. These are distinguished here based on their path number influence.

3 Preliminary Result

The ranking framework was evaluated using the Arnetminer dataset[13], which is a system for extracting and mining academic social networks. A small part of the dataset is sampled, including 10 771 papers, 14 210 authors, 1438 conferences, 29 596 author2pubs, and 14 804 citations in the database (see Fig. 2).

A default D2R mapping file was generated using D2R with several other relationship types added manually. Since recent user click-through data was not available, the relationship types were ranked based only on the concept hop influence. In the academic domain, the analysis prefers more explicit relationships, therefore, the ranking results for the relationship types are (the number in parenthesis represents the calculated score):

PublishedAt(1)=Accept(1)>Write(0.5)=
ReferTo(0.5)=CitedBy(0.5)=WriteBy(0.5)>
Coauthor(0.33)=InterestedConf(0.33)=
WriteBySameAtuhor(0.33)>Coattendee(0.2).

The ranking results are then reasonable, because users browsing the profile of one person usually mostly want to read his/her papers, are less interested in his/her coauthors, and are least interested in his/her coattendees. Similarly, when browsing the content of a paper, users most want to check whether it was published at an authoritative conference, and if authoritative, the users want to see the references, and are less interested in papers written by the same authors.

Then the relationship instances were ranked according to the one that in Section 3.3 by multiplying the PNI by the global popularity (GP). GP was calculated by setting the weightings based on the PNI and the relationship type influence calculated above. The top 5 resources for each relationship type belonging to the author "Stefan Decker" are shown in Table 3.

Table 3 Top 5 resources for each type of Stefan Decker

Write	Coauthor	InterestedConf	Coattendee
EDUTELLA: A P2P networking infrastructure based on RDF PNI:1.0, GP:4.472E-4	Dieter Fensel PNI:0.103 GP:2.020E-4	ISWC PNI: 0.1290 GP:0.0034	Ian Horrocks PNI: 0.030 GP:1.335E-5
Description logic programs: Combining logic programs with description logic PNI:1.0, GP:3.278E-4	Rudi Studer PNI:0.085 GP:2.142E-4	WWW PNI: 0.0652 GP:0.0032	Carole A. Goble PNI: 0.007 GP:2.676E-6
On2broker: Semantic-based access to information sources at the WWW PNI:1.0,GP:3.087E-4	Ian Horrocks PNI:0.034 GP:4.428E-4	EKAW PNI:0.0652 GP:0.0012	Steffen Staab PNI:0.018 GP:6.992E-6
Ontobroker: Ontology based access to distributed and semi-structured information PNI:1.0, GP: 2.829E-4	Steffen Staab PNI:0.034 GP:3.795E-4	IEEE Intelligent Systems PNI:0.0323 GP:0.0014	Daniel S. Weld PNI: 0.003 GP:1.258E-6
Enabling knowledge representation on the Web by extending RDF schema PNI:1.0, GP:2.058E-4	Frank van Harmelen PNI:0.034 GP:1.191E-4	FLAIRS Conference PNI:0.0323 GP:0.0010	William W. Cohen PNI:0.003 GP:1.181E-6

The type Coauthor shows that the number of coauthored papers with Dieter Fensel is much greater than with other authors, so the PNI is mainly related to Dieter Fensel's ranking. Ian Horrocks, Steffen Staab, and Frank van Harmelen have the same PNI, however, their different GPs determine their rankings, with Ian Horrocks linked by more resources to given high GP. For the type InterestedConf, ISWC is ranked higher than other conferences because Stefan Decker focused on the semantic web. Other benchmarks will be collected further to evaluate this approach.

4 Conclusions

This paper presents a two-step framework to rank relationships for a given resource for linked data from RDB. Tests show that the approach can improve the ranking experience for humans, which indicates that the approach utilizing relationships deduced from mapping rules based on table schemas and distinguished by relationship types is effective for linked data ranking.

References

[1] Sahoo S, Halb W, Hellmann S. A survey of current approaches for mapping of relational databases to RDF. http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport_01082009.pdf, 2009.

[2] Harth A, Kinsella S, Decker S. Using naming authority to rank data and ontologies for web search. In: Proceedings of the 8th International Semantic Web Conference. Berlin: Springer-Verlag, 2009: 277-292.

[3] Hogan A, Harth A, Decker S. ReConRank: A scalable ranking method for semantic web data with context. In: Second International Workshop on Scalable Semantic Web

Knowledge Base Systems. 2006.

[4] Franz T, Schultz A, Sizov S. TripleRank: Ranking semantic web data by tensor decomposition. In: Proceedings of the 8th International Semantic Web Conference. Berlin: Springer-Verlag, 2009: 25-29.

[5] Anyanwu K, Maduko A, Sheth A. SemRank A: Ranking complex relationship search results on the semantic web. In: Proceedings of the 14th International Conference on World Wide Web. New York: ACM, 2005: 117-127.

[6] Aleman-Meza B, Halaschek C, Arpinar B I. Context-aware semantic association ranking. In: The 1st International Workshop on Semantic Web and Databases. Berlin, 2003: 33-50.

[7] Ding Li, Pan Rong, Finin W T. Finding and ranking knowledge on the semantic web. In: Proceedings of the 4th International Semantic Web Conference. Berlin: Springer-Verlag, 2005: 156-170.

[8] Alani H, Brewster C, Shadbolt N. Ranking ontologies with AKTiveRank. In: Proceedings of the 5th International Semantic Web Conference. Berlin: Springer-Verlag, 2006: 1-15.

[9] Toupikov N, Umbrich J. Ding! Dataset ranking using formal descriptions. In: Linked Data on the Web Workshop in Conjunction with 18th International World Wide Web Conference. 2009.

[10] Mann S G, Mimno D, McCallum A. Bibliometric impact measures leveraging topic analysis. In: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries. New York: ACM, 2006: 65-74.

[11] Page L, Brin S, Motwani R. The pagerank citation ranking: Bringing order to the web. Stanford Digital Library working paper SIDL-WP-1999-0120 (version of 11/11/1999). http://www-diglib.stanford.edu/cgibin/get/SIDL-WP-1999-0120, 2009.

[12] Nie Zaiqing, Zhang Yuanzhi, Wen Jirong. Object-level ranking: Bringing order to web objects. In: Proceedings of the 14th International Conference on World Wide Web. New York: ACM, 2005: 567-574.

[13] Tang Jie, Zhang Jing, Yao Linmin. ArnetMiner: Extraction and Mining of Academic Social Networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2008: 990-998.