Enhancing Search Using Layered Graph Ranking of Multigraphs

Ján SUCHAL

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
suchal@fiit.stuba.sk

Abstract. Exploiting relationships between entities to improve search engines is a well known technique used by various graph ranking algorithms, such as PageRank or HITS. These algorithms however, typically use only one type of relationship. With the coming era of the Semantic Web, semantic data containing multiple types of relationships are becoming more common. In this paper we present a novel graph ranking method based on creating layered graphs from multigraphs. We also present a search engine prototype in the domain of scientific publications exploiting proposed layered graph ranking approach.

1 Introduction

Modeling relationships between entities using graphs and exploiting such graphs using ranking algorithms is a well known technique and has been proved useful in improving relevance of search and recommendation results. However, these graph ranking algorithms typically exploit only one type of entity and one type of relationship between entities, such as hyperlinks between web pages or friendship between people. With the upcoming era of the Semantic Web, semantic data containing several types of entities and relationships are becoming common, thus algorithms gaining advantage from such typed multigraphs are needed. However, typed multigraphs also bring new challenges not experienced in past graph ranking.

Section 2 introduces the context problem that arises from straightforward application of known graph ranking algorithms to typed multigraphs. Section 3 describes multigraph...
transformation into a layered graph. Also layered graph ranking algorithm is proposed which is used to solve context dependence of rank calculation. In Section 4 our prototype search engine in domain of scientific publications is introduced. Despite the fact that multigraph ranking is relatively new area of research, two interesting related approaches are presented in Section 5. Section 6 concludes this paper and presents two possible directions of future work.

2 The Context Problem

Figure 1 shows a partial schema of publication domain ontology with multiple types of entities and relationships between them.

Straightforward application of graph ranking algorithms to such multigraph reveals a problem unknown in graphs with one type of relationship between entities. The root of the problem is caused by the fact that different relationships between entities have the same weight. Unfortunately this assumption is too strong and far from truth in general case. For example when searching for similar publications one might prefer higher keyword similarity over authorship similarity. On the other hand, when searching for publications of same authors, it is clear that the authorship relationship dominates. As we can see, in general case, weights of relationships depend on the context of searching.

3 Layered Graph Ranking

Since graph ranking algorithms such as PageRank [4] are usually time-consuming processes, precomputation of ranks is often used to overcome this problem. Unfortunately with changing context, relationship weights change, thus ranks also change and precomputation is not possible.

Fortunately, by transforming full multigraph into a layered graph (see example on Figure 2) we can compute ranks for each layer, making ranks on each layer independent of the context.
This transformation is done by extracting only one type of relationship for each layer. In our domain of scientific publications (Figure 1) five such relationships/layers are possible.

For each layer ranks are precomputed beforehand and overall ranks $r$ are then computed as a linear combination of ranks on each layer.

$$r = \sum_{j=1}^{N} w_j r_j$$  \hspace{1cm} (1)

where weight vector $w = (w_1, w_2, ..., w_N)$ consisting of relationship weights represents current context.

**Fig. 2.** Example layered graph constructed from publication domain ontology multigraph shows three publications $P_1$, $P_2$ and $P_3$ and their connections through keywords $KW_1 - 4$, authors $A_1 - 3$ and topics $T1, T2$ on each respective layer.
4 Search Engine Prototype – Semantha

Search engine prototype (http://leela.fiit.stuba.sk/semantha/) was implemented to demonstrate capabilities of layered graph ranking approach.

A large scientific publication ontology used in project MAPEKUS\(^1\) extracted from ACM Digital Library\(^2\) has been used as source database. Three layers (keywords, topics and authorship) connecting publications are used in layered graph ranking.

Our search prototype searches for similar publications exploiting all of these layers. Similarity of publications on each layer is calculated using spreading activation search\(^2\) graph ranking algorithm with normalized distribution caching\(^5\).

Figure 3 shows typical result from our search engine showing similar publications with detailed information about rank values for each layer.

![Semantha search engine prototype GUI - Detail of similar publications displaying ranks for each layer.](image)

5 Related Work

PopRank algorithm powering Microsoft Libra academic search\(^3\) introduced by Zie et al.\(^3\) is an extension of PageRank algorithm applied to typed multigraphs with different

---

\(^1\) Project MAPEKUS - http://mapekus.fiit.stuba.sk/
\(^2\) ACM Digital Library - http://portal.acm.org/dl.cfm
\(^3\) Microsoft Libra academic search - http://libra.msra.cn/
weights of relationships. Weights of these relationships are learned using exhaustive
simulated annealing algorithm maximizing relevance on manually ranked data by domain
experts. For each weight vector full PopRank must be recomputed.

In Aleman-Meza et al. [1] patterns of paths in typed multigraphs (ontologies) are
weighted by users and user behavior. These weighted paths patterns are then exploited
to find other relevant instances in ontologies which are then used as a recommendation
to users.

6 Conclusions and Future Work

Semantic Web brings structured data with multiple types of entities and relationships.
Ranking such typed multigraphs is a new area of research with open and upcoming
problems.

We have described the context problem which lies in differentiating weights of dif-
ferent relationships depending on the context of searing. Since context might change in
time or even between users and graph ranking algorithms are time-consuming processes,
precomputation of ranks is a problem.

We have proposed a transformation of multigraph into multiple layers which allows
precomputation of ranks on each layer and fast calculation of final ranks depending on
chosen context vector. Proposed layered graph ranking is used in a freely available search
engine prototype (http://leela.fiit.stuba.sk/semantha/) in domain of scientific publications
that searches for similar publications based on authorship, keywords
and topic similarity.

Future work focuses mostly on two directions:

• Calculating ranks as combination of ranks in each layer of multigraph is an
  approximation of ranks on full multigraph calculated by PopRank. Quality of this
  approximation needs to be evaluated.

• Since calculation of final ranks is fast, context vector of relationship weights
  can be used to personalize search. Furthermore, evaluating relevance of results
  presented to user using implicit or explicit feedback can be exploited to find best
  matching context vector.

Acknowledgement: This work was partially supported by the Slovak Research and
Development Agency under the contract No. APVT-20-00710 and the State programme
of research and development "Establishing of Information Society” under the contract
No. 1025/04.
References


