Nandish Jayaram <j.nandish@gmail.com>

Fw: Major Revision Decision Re: TKDE-2014-05-0350

Jayaram, Nandish <nandish.jayaram@mavs.uta.edu> Thu, Sep 18, 2014 at 11:56 AM
To: "j.nandish@gmail.com" <j.nandish@gmail.com>

From: onbehalfof+s.parikh+ieee.org@manuscriptcentral.com <onbehalfof+s.parikh+ieee.org@manuscriptcentral.com> on behalf of s.parikh@ieee.org <s.parikh@ieee.org>
Sent: Wednesday, September 17, 2014 2:27 AM
To: Jayaram, Nandish; arijit.khan@inf.ethz.ch; Li, Chengkai; xyan@cs.ucsb.edu; Elmasri, Ramez A
Cc: s.parikh@ieee.org
Subject: Major Revision Decision Re: TKDE-2014-05-0350

RE: TKDE-2014-05-0350, "Querying Knowledge Graphs by Example Entity Tuples"
Manuscript Type: Regular

17-Sep-2014

Dear Dr. Jayaram,

We have completed the review process of the above referenced paper that was submitted to the IEEE Transactions on Knowledge and Data Engineering for possible publication and recommend that your paper undergo a Major Revision.

Your reviews are enclosed. We would suggest that you revise your paper according to the reviewers' comments and resubmit the paper for a second round of reviews. If you wish to revise the paper, please do so before 16-Dec-2014.

To revise your manuscript, log into https://mc.manuscriptcentral.com/tkde-cs and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

If you do not intend to submit a revised version of your paper, please let us know so that we can formally close your file.

Please be mindful when making your revisions that you still need to maintain the size limitations for papers submitted to TKDE. TKDE manuscript types and submission length guidelines (including the main text, the abstract, index terms, illustrations, references and bios) are found at,

http://www.computer.org/portal/web/peerreviewjournals/author#manuscript

Please note that double column will translate more readily into the final publication format. Our peer review double column templates can be found at,

http://www.computer.org/portal/web/peerreviewjournals/author#templates

Text in any color other than black is not acceptable. Your revised paper must include the following:

-abstract
-index terms
-author affiliation information
-main text
-references
-figure captions

-table titles
-brief biography of each author
(biographies are not required for concise papers or comments papers)

I will send your revised manuscript and your revision summary to the Associate Editor and reviewers for comments.

Should you realize that you forgot to include a file before submission, please email it to me at s.parikh@ieee.org. Please do not hesitate in contacting me should you have any questions about our process or are experiencing technical difficulties.

Thank you for your contribution to TKDE, and we look forward to receiving your revised manuscript.

Sincerely,

Sonal Parikh on behalf of Jian Pei, EIC
IEEE Transactions on Knowledge and Data Engineering
s.parikh@ieee.org

****

Editor Comments

Associate Editor
Comments to the Author:
In this paper, the authors introduce a Graph Query By Example (GQBE) system, which aim to automatically capture users' query intent (inferred from input query tuples) using a weighted maximal query graph. The reviewers have agreed with the important and novelty of the proposed problem and system. However, they also pointed out a list of important issues which need to be addressed. Given this, I would like the authors to make a major revision of the paper.

The following are some key concerns/issues from authors:

1. Reviewers have some concern/questions on the design choice on the technical solution. Both Reviewer 1 & 3 provides some suggestions on alternative approaches. I would suggest the authors to provide some further discussion on what are the naive approach, and what are the advantage of the proposed multi-stage approach? Reviewer 2 suspect the proof of Theorem 1 maybe incorrect. I hope the authors can double-check this and provide a feedback/high-level explanation on the correctness of Theorem 1 for the followup responses.

2. The experimental results need some further improvement. For instance, Reviewer 3 has some questions on the scalability on the real-world scale-free knowledge graph.

3. Reviewers have some suggestions on the overall organization of the paper. Particularly, to make the paper self-contained, the appendix can be added into the main paper.

Besides the major issues, the authors should also address the typos and minor issues as being pointed out by the reviewers.

****
Reviewer: 1

Recommendation: Author Should Prepare A Major Revision For A Second Review

Comments:
1) This paper proposes a new kind of graph query, which yields the results by example entity tuples. Such a kind of query can free end users from expressing complex graph queries.
2) Based on user's input tuples, this paper automatically derives a weighted hidden maximal query graph to capture user's intend, models the space of query graph by a query lattices, and designs a method to partially evaluate the query lattices to obtain the top-k answer tuples.
3) They conduct experiments on two real dataset, including freebased and dbpedia to illustrate the precision and efficiency of graph query by example.

I am concerned about the following issues:
1) Query by example in the relational database can express the predication in the query expression. Then, it seems more powerful than the query proposed in the paper, which only expresses result examples. It may lead to misunderstanding if the query proposed in this paper shares the same name as “the query by example”.

2) As for the evaluation method, please justify the transformation from user’s input tuples, neighborhood graph, maximal query graph, merging maximal query graph to query lattice. Some of transformations are time consuming. Do we really need such an expensive evaluation method? For example, we can cluster the terms in the knowledge graph offline first. As for the given two (or more) terms, we join two (or more) corresponding sets, rank the results, and report top-k results to users. The paper should explain why these straightforward methods do not work for the new query proposed in the paper.

3) In the knowledge graph, the labels on edges may be not exactly the same but semantically related, which results in the mismatch between query graph and data graph. These factors seems not be considered in the paper.

4) In the experimental part, the authors use the tables (freebase tables or Wikipedia tables) as the ground truth. Then again, could we extract the similar information (e.g., from tables in web pages) to answer example query without enumerating and evaluating different kind of queries?

Additional Questions:
1. Which category describes this manuscript?: Research/Technology

2. How relevant is this manuscript to the readers of this periodical? Please explain your rating under Public Comments below. : Relevant

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature. : 1) This paper proposes a new kind of graph query, which yields the results by example entity tuples. Such a kind of query can free end users from expressing complex graph queries.
2) Based on user’s input tuples, this paper automatically derives a weighted hidden maximal query graph to capture user’s intend, models the space of query graph by a query lattices, and designs a method to partially evaluate the query lattices to obtain the top-k answer tuples.
3) They conduct experiments on two real dataset, including freebased and dbpedia to illustrate the precision and efficiency of graph query by example.

2. Is the manuscript technically sound? Please explain your answer under Public Comments below. : Partially

1. Are the title, abstract, and keywords appropriate? Please explain under Public Comments below.: Yes

2. Does the manuscript contain sufficient and appropriate references? Please explain under Public Comments below.: References are sufficient and appropriate

3. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on? Please explain your answer under Public Comments below.: Yes

4. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic? Please explain under Public Comments below. : Satisfactory

5. Please rate the readability of the manuscript. Explain your rating under Public Comments below.: Readable - but requires some effort to understand

6. Should the supplemental material be included? (Click on the Supplementary Files icon to view files): No, it should not be included at all

7. If yes to 6, should it be accepted:

Please rate the manuscript. Please explain your answer.: Good

Reviewer: 2

Recommendation: Author Should Prepare A Major Revision For A Second Review

Comments:
In this article, the authors introduce a Graph Query By Example (GQBE) system, which automatically derives a weighted hidden maximal query graph based on input query tuples, to capture a user's query intent. This is in particular useful to promote the usage of graph queries. I can see the merits and importance of this work, but I feel this paper could be improved.

Below please find my detailed comments.

1. I would suggest the authors to remove Section 1.3, and merge its content with either Section 1.2 or Section 8.

Furthermore, although the workshop paper in GRADES'2014 was mentioned in the summary letter, it was ignored in the paper itself.

2. A journal paper needs to be self-contained. It is extremely inconvenient to see so many ``please refer to the Appendix...''.

Removing 1.3 should save some space, and the authors should at the very least provide the proof sketches in the paper.

3. I am not convinced with the proof of Theorem 1. Please provide a detailed proof. In addition, I would like to point out that ``maximal'' is claimed in the theorem.

4. In the experiments, I did not see the logic that "In aforementioned single-tuple query experiment (A), GQBE attained perfect P@25 for 13 of the 20 Freebase queries. We thus focused on the remaining 7 queries." Why not provide the experimental results of all the queries?

Minor comments:

1) approach + "doing"

2) Accuracy Based on User Study -> Accuracy Based on User Studies

Additional Questions:
1. Which category describes this manuscript?: Research/Technology

2. How relevant is this manuscript to the readers of this periodical? Please explain your rating under Public Comments below. : Very Relevant

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature. : In this article, the authors introduce a Graph Query By Example (GQBE) system, automatically derives a weighted hidden maximal query graph based on input query tuples, to capture a user's query intent. This is in particular useful to promote the usage of graph queries.

2. Is the manuscript technically sound? Please explain your answer under Public Comments below. : Appears to be - but didn't check completely

1. Are the title, abstract, and keywords appropriate? Please explain under Public Comments below.: Yes

2. Does the manuscript contain sufficient and appropriate references? Please explain under Public Comments below.: References are sufficient and appropriate

3. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on? Please explain your answer under Public Comments below.: Could be improved

4. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic? Please explain under Public Comments below. : Could be improved

5. Please rate the readability of the manuscript. Explain your rating under Public Comments below.: Readable - but requires some effort to understand

6. Should the supplemental material be included? (Click on the Supplementary Files icon to view files): Yes, as part of the digital library for this submission if accepted

7. If yes to 6, should it be accepted: After revisions. Please include explanation under Public Comments below.

Please rate the manuscript. Please explain your answer.: Good

Reviewer: 3

Recommendation: Author Should Prepare A Minor Revision

Comments:
1) The paper discusses an interesting type of knowledge graph queries. The introduction section clearly presented the problem, research motivation, and the approach intuition.

2) For the paper organization, there is no obvious difference between section 1.3 and section 8 (Related work). Probably they should be merged.

3) Many design choices seem pure heuristic and ad-hoc, more justifications are needed. For example, why do you choose to first construct a neighborhood graph followed by a query graph construction procedure, instead of directly constructing a query graph using mechanisms like spanning tree? The latter approach is much simpler and probably more efficient, considering the neighborhood graph is usually very dense. For the proposed MQG (maximal query graph), what makes it a good structure that well captures the ‘query semantics’ or ‘user intent’ is not clearly discussed or analyzed.

4) For the experiments, the data sets are highly reduced. The reduced data set is very sparse, for example the reduced Freebase’s average degree is only 1.68. The original Freebase has 45+ M entities and 2.6+ B facts, which is much denser than the reduced data set. For the denser original graph, it is unclear whether the proposed approach still works well considering the fact that the neighborhood graph would be a lot more complex as the data graph becomes denser. GQBE takes seconds to complete for many queries, it is acceptable. Would the performance number still be acceptable when applying the method to a full-scale real-world knowledge graph?

Additional Questions:
1. Which category describes this manuscript?: Research/Technology

2. How relevant is this manuscript to the readers of this periodical? Please explain your rating under Public Comments below. : Relevant

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature. : The authors proposed an intuitive way of querying a knowledge graph --- query the graph by examples. Even there are similar works in the IR or web search field, it is interesting to see how to query a knowledge graph just by user-provided sample entity-pairs.

2. Is the manuscript technically sound? Please explain your answer under Public Comments below. : Yes

1. Are the title, abstract, and keywords appropriate? Please explain under Public Comments below.: Yes

2. Does the manuscript contain sufficient and appropriate references? Please explain under Public Comments below.: References are sufficient and appropriate

3. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on? Please explain your answer under Public Comments below.: Yes

4. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic? Please explain under Public Comments below. : Could be improved

5. Please rate the readability of the manuscript. Explain your rating under Public Comments below.: Readable - but requires some effort to understand

6. Should the supplemental material be included? (Click on the Supplementary Files icon to view files): Yes, as part of the digital library for this submission if accepted

7. If yes to 6, should it be accepted:

Please rate the manuscript. Please explain your answer.: Good