VLDB 2014
Proceedings of the Very Large Database Endowment, Volume 7
Hangzhou, China

Reviews For Paper

Track	Research -> May 2013
Paper ID	81
Title	GQBE: Querying Entity-Relationship Graphs by Example Tuples

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question

Overall Recommendation	Weak Accept
Are there specific revisions that could raise your rating in the previous question?	Yes
Summary of the paper (what is being proposed and in what context) and a brief justification of your overall recommendation. One paragraph	This paper propose a complete method to query knowledge graphs using a QBE approach. The technique covers all stages of query processing: (1) Identifying the query features, (2) Scoring the answers, (3) top-k query processing for efficiency. TH=he proposed techniques are strong and the experimental evaluation is thorough. The paper is however very dense and would benefit from editing. Some connections with existing work on XML relaxations are needed.
Three (or more) strong points about the paper (Please be precise and explicit; clearly explain the value and nature of the contribution).	(s1) Complete system for querying knowledge graphs by example (s2) Clear motivations. This type of systems seems likely to be very useful. (s3) Thorough experimental evaluation
Three (or more) weak points about the paper (Please indicate clearly whether the paper has any mistakes,	(w1) Paper is too dense and compact (w2) The example should be used to illustrate the whole querying process from beginning to end (w3) The novel contributions are not clearly identified

missing related work, or results that cannot be considered a contribution; write it so that the authors can understand what are seen as negative aspects	
Relevant for PVLDB	YES
Novelty (Please give a high novelty ranking to papers on new topics, opening new fields, or proposing truly new ideas; assign medium ratings for delta papers and papers on well known topics but still with some valuable contribution).	With some new ideas
Significance	Improvement over existing work
Technical Depth and Quality of Content	Solid work
Experiments	Very nicely support the claims made in the paper
Presentation	Reasonable: improvements needed
Detailed Evaluation (Contribution, Pros/Cons, Errors); please number each point	(d1) The paper is very thorough and well written but it crams too much information in the 12 pages. It is obvious the authors have played with the font size and vertical spaces to fit everything in the allowed space. The result is a dense paper that is often difficult to follow: equations are inlined, figures are so small they are hard to read, there is no transition text between sections, which makes the paper hard to follow. I would recommend that the authors remove some of the material, proofs and complexity discussions could be moved to a TR and referenced in the paper. (d2) Probably because of the space issue, the running example is dropped

	at some point in Section 4 and replaced by a much less helpful example using ABCDE as edge names. But even that example is not consistently used in the remainder of the paper, Section 5.2 for instance would be much easier to follow with an example. (d3) The introduction identifies three challenges, but not the novel contributions of the work. (d4) This is my most major comment on the technical content of the paper. The scoring strategy, and to some extend the top-k query processing part, share a lot of similarities with work on XML relaxations (see "Structure and Content Scoring for XML", Amer-Yahia et Al. VLDB 2005 and subsequent work). In particular, the idf-like scoring of section 3.2, the lattice representation, and exploration have strong similarities with that work. (d5) When two answers are projected from multiple answers, why take the highest score. Shouldn't these answers get a boost from the fact that they are reinforced by several distince tuples? (d6) Section 6 covers an interesting problem but looks like an afterthought. It is also not clear to me that the ideas presented in that section are verified experimentally. I would like to see that section expanded, but given the space restrictions, this might be a good candidate to cut and expand for a journal version of the paper.
If revision is required, list specific revisions you seek from the Authors	(r1) Identify novel contributions. (r2) Discuss connections with work on XML approximate scoring (r3) Improve the presentation: - increase size of figures, add vertical spaces - add transitions between the different sections - move proofs and detailed discussion to TR - expand on the (unified) running example to illustrate each step of the proposed techniques

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question

Overall Recommendation	Weak Reject
Are there specific revisions that could raise your rating in the previous question?	No

Summary of the paper (what is being proposed and in what context) and a brief justification of your overall recommendation. One paragraph	The paper considers data modeled in the form of entities stored in a graph. The authors claim that graph queries are difficult to formulate. So they propose a system to support query-by-example over graphs. The biggest novelty in the system is mapping query keywords to a sub-graph query. Then that graph query is used to provide results similar to it. There is an inherit conflict about this paper. I am not sure the motivation captures a real user need with regards queries typed. Furthermore, assuming support for such query types is a necessity, it is not clear the proposed solution will work with the scale of data or how it compares to other query-by-example techniques, not necessarily using graphs, but addressing the same user need.
Three (or more) strong points about the paper (Please be precise and explicit; clearly explain the value and nature of the contribution).	- mapping of entity based keyword queries to reduced query graphs - statistical pre-computation of important nodes in a graph - the use of mechanical turk in the experimental section
Three (or more) weak points about the paper (Please indicate clearly whether the paper has any mistakes, missing related work, or results that cannot be considered a contribution; write it so that the authors can understand what are seen as negative aspects	- Insufficient motivation, not clear what is the concrete user scenario targeted - Data size and experiments not clear will work when targeting the web type of users - Weak experimental approach with regards other solutions as baselines
Relevant for PVLDB	YES
Novelty (Please give a high novelty ranking to papers on new topics, opening new fields, or proposing truly new ideas; assign	Novelty unclear

medium ratings for delta papers and papers on well known topics but still with some valuable contribution).	
Significance	No impact
Technical Depth and Quality of Content	Syntactically complete but with limited contribution
Experiments	OK, but certain claims are not covered by the experiments
Presentation	Excellent: careful, logical, elegant, easy to understand
Detailed Evaluation (Contribution, Pros/Cons, Errors); please number each point	1) Scenario and motivation: Query by example has been around a long time. Web search has also been around. The authors claim there is a need for graph query by example that looks like web search but retrieves information for similar entities. Web queries do contain entities, however users look for information on those entities. They are not used in seeing other related entities when not explicitly asked for such information. And when accessing a dedicated database system, there are techniques that already retrieve results, including query-by-example variations. The motivational scenario and its importance is not clear to me, and the proper scenario will dictate the right experiments to perform. My recommendation is to clarify the user problem more or perhaps limit the scope of the user scenario. 2) Query formulation: The main contribution in the paper is mapping the keyword query to a subgraph that is missing some relationships. The rest of the similarity retrieval sounds similar to techniques studied in past graph processing solutions (not necessarily published in DB conferences). The authors assume the query is entered in the form (jerry yang, yahoo!) but is actually means (results similar to "jerry yang founder of yahoo!") and they construct that meaning. The problem would be much easier if we assumed the users actually enter the "founder" word. Furthermore, it is not clear the user meant founder and not (jerry yang left yahoo) or some other "weaker" relationship. 3) Data size: Not clear the proposed solution will work for large data sizes and with what type of latencies. I wonder if it can be parallelizable. For this to be a realistic scenario in a web search type of alternative, we will be looking at potentially hitting a variety of entity databases addressed with keyword queries. So size and performance can be issues. 4) Alternate approaches: Assuming the motivation captures a real user need, it is not clear that a graph based solution will be the best. The query (jerry yang, yahoo!) can be addressed with other type of query-by-example

	solutions. It would be very convincing to see the authors compare their approach against such solutions as an end-to-end system. 5) Correct results: It is not clear to me how the results were labeled as correct. Given (jerry yang, yahoo), the authors claim (bill gates, microsoft) is correct and in the top 3. Similarly, all founders for all companies and their companies will be deemed correct. So as long as the "founder" node is mapped properly all nodes are correct. I am not sure what makes (bill gates, microsoft) better than any other founder / company pair. Hence, I am confused as to how one should interpret the results. To be fair, it is possible this confusion is because I am not able to follow the user need in the motivation. Therefore, I am not able to understand what would be the theoretically ideal results a system should thrive to produce. I guess a few more running examples with positive and negative results would be beneficial.

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question

Overall Recommendation	Weak Reject
Are there specific revisions that could raise your rating in the previous question?	Yes
Summary of the paper (what is being proposed and in what context) and a brief justification of your overall recommendation. One paragraph	The paper describes a system for applying query-by-example ideas to query semantic graphs. The paper is reasonably well-written (see below for small issues). The formalization of the problem is nice, and the solution has some interesting technical ideas, coupled with a working system. However, the experimental evaluation does not strongly support the complex architecture and metrics for the system, making the design seem somewhat arbitrary.
Three (or more) strong points about the paper (Please be precise and explicit; clearly explain the value and nature of the contribution).	1. The problem is interesting 2. The authors have a nice formalization of the problem using neighborhood graphs, very clearly defined. 3. I really like the lattice-based representation and pruning of query graphs: it's a neat idea. 4. The system seems to work relatively well.

Three (or more) weak points about the paper (Please indicate clearly whether the paper has any mistakes, missing related work, or results that cannot be considered a contribution; write it so that the authors can understand what are seen as negative aspects	1. The overall architecture of the system is opaque, with many seemingly arbitrary choices that don't have enough justification. 2. The experimental section does not truly demonstrate the utility of the system.
Relevant for PVLDB	YES
Novelty (Please give a high novelty ranking to papers on new topics, opening new fields, or proposing truly new ideas; assign medium ratings for delta papers and papers on well known topics but still with some valuable contribution).	Novel
Significance	Improvement over existing work
Technical Depth and Quality of Content	Solid work
Experiments	Obscure, not really sure what is going on and what the experiments show
Presentation	Reasonable: improvements needed
Detailed Evaluation (Contribution, Pros/Cons, Errors); please	Detailed Issues: 1. The paper very quickly jumps into the nitty-gritty, without giving the reader an overview of the system architecture: Overall, the way I understood it -- after a struggle -- the system starts with the query Q,

number each point	generates a Reduced Graph, then generates a Maximal Graph, balances it out, then constructs (incrementally) a lattice of all Query Graphs, matches Query Graphs against Answer Graphs (not entirely clear how), then spits out the top-k' (k'> k) answer graphs. It would really help to provide a system overview before jumping into the details. 2. Related to the above, a number of these steps are heuristic and (somewhat unnecessarily) complex in nature -- they require justification (either in text or via experiments): a. In Section 3.1, the unimportant edges are pruned from the neighborhood graph. I found it hard to get intuition as to why unimportant edges need to be removed, and whether there is any loss in doing so. Moreover, the definition for unimportance seemed recursive to me; I found it hard to follow. b. In Section 3.2, the metric of weighting edges seems intuitively fine, but why not any other metric? This requires experimental justification. c. In Section 3.2, the method for selecting edges to form the maximal graph seems unnecessarily complex: why not just add high weight edges until all the query entities are covered? d. In Section 3.2, what was the need to balance the maximal graph? What is lost if we keep the graph as is without balancing? e. In Section 5, I was not clear when and how answer graphs are evaluated and scored while evaluating query graphs in the lattice. Are there multiple answer graphs maintained for the same query graph, or is at most one answer graph maintained for each query graph -- I suspect the second, but in that case aren't you losing some potentially good answer graphs? f. In Section 5, how much are you gaining by pruning away the lattice? How much is gained by best-first search? How much is lost by approximately finding the best top-k (how much does k' impact the process?) 3. The experimental evaluation has a few issues: a. The ground set construction for the queries in experiment (A) seems strange to me -- wasn't the point of the system to uncover hidden relationships that are not obvious? Here, it looks like the ground set precisely seems to be the obvious (relatively simple) relationships. I suspect a straightforward graph query system that does a simple lookup of the query entities and relationships may do just as well as the complex GQBE system for such a case. b. It's nice that the authors used crowdsourcing to evaluate the utility of the system, but it seems like they only use MTurk to test the accuracy of the metric using which the top-k results are ranked (among themselves) by GQBE rather than whether the system actually produces interesting results in the top-k. That is, there may be items not found by GQBE that would be rated higher by human workers, but are not considered in the experiment. Issues in writing:

	1. Second Column, first page: "It requires extensive experiences" => extensive experience 2. Second Column, first page: "Its simplicity and improved user productivity make" => makes 3. First Column, second page: Typo: underlyingly 4. First Column, second page: "Note that query graphs...": I did not get the point of this paragraph, unclear. 5. First Column, second page: "In more details", rephrase. 6. Definition 5: What are v_1, ..., v_n 7. Page 8, Alogrithm => Algorithm, appears a couple of times 8. Second Column, Page 8: excepting => except
If revision is required, list specific revisions you seek from the Authors	Answers to questions above; an improved organization of the paper, as well as supporting experiments.