Dear Editor,

On the behalf of my co-authors, I enclose the revised manuscript entitled “Fast Subgraph Matching Strategies based on Pattern-only Heuristics”. We would like to thank the reviewers for their efforts to improve our work and we are grateful to them for their constructive, extensive, and insightful comments. We highlighted the modifications in red in the revised manuscript. Below we give a detailed point-to-point response.

Best Regards,

Rosalba Giugno

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Reviewer #2:**

This manuscript is a comparison of state-of-the-art subgraph isomorphism methods, especially in the context of analyzing biologically-relevant graphs. The manuscript includes adequate detail on the generation of relevant datasets used in the comparisons. Also, the datasets generated are fairly comprehensive for the biological use-cases described. And there is a nice introduction to the algorithms being compared. However, this topic has been well-studied before. And the main point that pattern heuristics are key to superior algorithmic performance appears to be a specialization of previous work illustrating that graph coloring reduces the subgraph isomorphism problem to polynomial complexity. Therefore, I would only recommend acceptance if the authors adequately address the following major issues.

**Major Issues**

**Reviewer question**

1) The authors' major result that certain heuristics move the subgraph isomorphism problem into polynomial time is basically the concept of coloring graphs, which is already known to reduce the subgraph isomorphism problem to polynomial time, when the coloring overcomes the NP-hardness of cycles in the graphs. The authors must discuss their work in the context of the following two publications: a) Alon, Noga, Raphael Yuster, and Uri Zwick. "Color-coding." Journal of the ACM (JACM) 42.4 (1995): 844-856; b) Kratsch, Stefan, and Pascal Schweitzer. "Isomorphism for graphs of bounded feedback vertex set number." Scandinavian Workshop on Algorithm Theory. Springer, Berlin, Heidelberg, 2010. The authors also need to point out (at least in response to this reviewer) how their analyses provide new information beyond these and similar publications.

*Authors Answer*

*We thank the reviewer for this useful comment. We have extended the manuscript by citing the suggested relevant works. Note that we do not state that our strategy solves the Subgraph matching problem in polynomial time. We report heuristics that in practice make the algorithm very fast. Furthermore, such heuristics are very general and their applicability does not require any restriction on pattern or target topologies. This paper presents a comparison of state-of-the-art subgraph isomorphism methods for the general case, where no polynomial algorithms are known. To our knowledge the cited works present polynomial solutions under some graph assumptions. As far as we know, there is no software implementing them to test and no related work where they are compared on general large graphs, which is the aim of this paper.*

**Reviewer question**

2) The subject of this manuscript is a well-studied topic. Here is very similar type of subgraph isomorphism comparison paper from 2012 that is well-cited: Lee, Jinsoo, et al. "An in-depth comparison of subgraph isomorphism algorithms in graph databases." Proceedings of the VLDB Endowment. Vol. 6. No. 2. VLDB Endowment, 2012. What are the authors adding beyond this type of comparison?

*Authors Answer*

*We thank the reviewer for this useful comment. The cited paper presents a deep comparison between algorithms that are no longer state-of-the-art (for example VF2). Our work compares state-of-the-art algorithms and tests them on larger datasets. In fact, in Lee’s work, the target sizes reach a maximum of 5000 vertices, while the dimensions of the patterns grew up to 24 vertices. In our work, on the other hand, the number of vertices of the targets reaches 20000, while that of the patterns 256. We have however added the reference and discussed this previous work in the Introduction.*

**Reviewer question**

3) To broaden the accessibility of the manuscript, this reviewer would suggest either adding a glossary of what all mathematical symbols used in the manuscript mean, probably as supplemental material or a reference to an online resource that provides a glossary of a superset of the mathematical symbols used in the manuscript. Wikipedia is a nice glossary of mathematical symbols:<https://en.wikipedia.org/wiki/List_of_mathematical_symbols> . However, you will see that some symbols have a variety of meanings depending on the context. If a mathematical symbol glossary of terms is provided, then specifically mention this in the manuscript.

*Authors Answer*

*We thank the reviewer for this useful comment. As suggested, we point the reader to the following mathematical glossary* https://en.wikipedia.org/wiki/List\_of\_mathematical\_symbols *to help him/her with the notation.*

**Reviewer question**

4) page 3, line 38: As the authors stated the subgraph isomorphism problem is to find all subgraphs of G\_t that match G\_p. So, the problem is not to find a single injective function f : V\_p -> V\_t , but to find all injective functions f\_i: V\_p -> V\_t.

*Authors Answer*

*We thank the reviewer for this useful comment. We have amended the definition 2.*

**Reviewer question**

5) This reviewer could not find the details on the implementation of the three algorithms RI, RI-DS and VF3 that was used to generate the results. What language were these algorithms implemented in? What graph representation(s) did they use?

*Authors Answer*

*We thank the reviewer for this useful comment. RI and RI-DS have been implemented in C while VF3 in C ++. We used the source codes of the algorithms kindly provided by the authors. Each graph is stored in a text file in the form required by each algorithm which lists the nodes and edges with their corresponding labels. The target graphs were generated in the RI format and converted with a script in Python to obtain the dataset in VF3 format.*

**Reviewer question**

6) How can the authors claim that just pattern-only heuristics are needed for improved algorithmic performance when the fundamental properties outlined in definition 2 indicates the need for alpha and beta annotation functions specific to both the target and pattern graphs?

*Authors Answer*

*We thank the reviewer for this useful comment that we clarify as follows. Definition 2 lists the constraints that the mapping of nodes must satisfy to be considered as a subgraph matching occurrence. Our assertion regards the heuristics that graph searching algorithms use to construct the mapping and check the satisfiability of the constraints. Tests on large graphs showed that searching heuristics mainly based on patterns make subgraph matching algorithms faster than those based on target graphs.*

**Minor Issues**

page 3, line 33: insert a "the" between "to" and "pattern."

page 3, line 44: point 3 seems to be covered by the statement that f is an injective function.

page 3, line 45: in point 4, is the backward E symbol unneeded? (Just asking the authors to check).

page 4, line 20: shouldn't the inequality i < |V\_p| actually be i <= |V\_p| ?

page 4, line 31: shouldn't the inequality N\_1(u) > N\_1(v) actually be |N\_1(u)| > |N\_1(v)| ? This would also affect several other inequalities in the nearby vicinity.

page 5, line 32: replace "by run" with "were produced by using".

*Authors Answer*

*We thank the reviewer for these useful comments and suggestions. We have addressed all comments.*

*page 3, line 33: insert a "the" between "to" and "pattern.": Sentence is slightly different now.*

*page 3, line 44: It is true, some points are implicit in what is said early, as point 1. We list all properties for clarity purpose.*

*page 3, line 45: No, it is not needed. Edges between mapped nodes can exist in the target but not in the pattern graph.*

*page 4, line 20: Thanks, we have corrected this.*

*page 4, line 31: Thanks, we have corrected this.*

**Reviewer #3:**

This paper presented an approach for subgraph matching, overall concept is good, methodology is also good, presentation and organisation may be improved.

*We thank the reviewer for these appreciations. We have revised the manuscript according to all reviewers* comments*.*