Dear managing board, dear recommenders,

Please find the attached revision of our manuscript entitled *"The Structure and Dynamics of Knowledge Graphs, with Superficiality"* to be considered for recommendation in PCI Network Science.
https://arxiv.org/abs/2305.08116 [version 3] (submitted under arXiv May 31, 2024)

First of all, we would like to apologize for the excessive time it took us to revise our manuscript. We would also like to sincerely thank the 3 recommenders for their feedback, which enabled us to significantly improve the article.

We now detail our main modifications in response to the suggestions made by the 3 reviewers, where :

- R1 refers to the review by Abiola Akinnubi.
- R2 refers to the review by Mateusz Wilinski.
- R3 refers to the review by Tamao Maeda.

**RELATED WORK AND BIBLIOGRAPHY**

*R1 "Author references or bibliograph contains references that seems incomplete."*
*R2 "The references use different styles: sometimes we have "et al." after the first author, sometimes after four authors and sometimes we have all five authors mentioned. I would suggest to "unify" it."*

We have unified the style of the bibliography, listing all authors and providing more detail on the articles referenced.

*R1 "References or Bibliography seems too long. Author should focus on reference that made it to text."*

All references are cited in the article, either in the main text or in the supplementary material. We must explain what knowledge graphs are, this is a part of the references. The other part is to show that we carefully considered all existing models with respect to the aim of knowledge graphs modeling. Finally, there are several references to data sets to guarantee the reproducibility of our experiments.

**CLARITY OF THE PRESENTATION**

*R1 "While the paper is novel in her thought process, the structure of the paper suffers greatly. For example, there are confusion as to whether the paper is literature review or research paper. Also, the paper transition from areas that may appear as literature review to areas the present the paper contribution is confusing or missing.*
*Authors needs to structure the paper to the appropriate format to allow easy*

*following."*

The titles of the article may not be self-explanatory, but we think the structuring is classic:

- Knowledge Graphs : Introduction with the contributions mentioned in the last paragraph
- Preferential Attachment and Multiplex Networks : Literature review
- Generative Model with Superficiality : Paper contribution with the introduction of the new model and the main experimental findings
- Impact of Superficiality on Ignorance : Discussion and conclusion

In addition, we modified the section "Generative Model with Superficiality" by adding a more structured layout to facilitate the reader's progression. We also improve the presentation of some concepts (see related work) and the presentation of our new model.

*R1 "Author seems to be calling knowledge base as knowledge graph these are two different concepts while they borrow information from each other."*

In the revised version, only the term knowledge graph is used (after the definition in the introduction). It's simpler to use just one term.

*R3 "p3. "a relationship must be enriched…".*
*"Created" or "established" instead of "enriched"?*
*I think enrich is used like "enrich a network", but not to a single edge (relationship)."*

The verb "enrich" is indeed not very clear. In the end, we chose the verb "complete", which we feel is more precise.

*R2 "Multiplex is not the same as multilayer, it is a special case of the latter."*

We have changed the sentence.

*R2 "Authors could spend a bit more time when describing the whole concept of knowledge graphs.[...] some extra figures and examples could make the whole manusciprt much clearer for a non-specialist or a general reader."*

We have added a figure showing an excerpt from a knowledge graph. This makes it possible to gather elements of examples given in the article in the same place.

*R3 "I recommend to add a list of each parameter with a brief explanation (a table, for example) to aid readers to understand."*

We have added a table in Figure 3 describing the input parameters of the generative model. We also provided the notations of the main studied connectivities.

*R3 "Superficiality refers to the fact that something is unimportant and superficial, which I though it does not match with the meaning of the notion you added in the model. Would you explain why you used this term?"*

We hope it is not the model that is superficial, but the knowledge modeled within the knowledge graph. Indeed, the higher this parameter σ, the more the facts added to the knowledge graph are spread over numerous entities, leading on average to a superficial knowledge of each entity.

**EXPERIMENTS**

*R2 “I am slightly puzzled by the results on comparing Authors' model with BA (their fits to data). Isn't BA a special case of the new described model? How can it be better for any instance then (outgoing ChEMBL for example)? Does it even make sense to make a comparison in this case? Is it similar for Bollobas?”*

The comparison with simplex networks makes sense when all links are considered without taking into account the type of relationship (a single layer). This approximation is naive and explains the poor reproduction of degree distributions. We've added a clarification in the ablation study.

*R2 "2. What model do Authors use when they say Bollobas? Bollobas-Riordan? Though there is one citation, it is not referred to in the description of the experiment (comparing different models)."*

We have clarified the model based on the reference below:
Bollobás, B., Borgs, C., Chayes, J. T., & Riordan, O. (2003, January). Directed scale-free graphs. In SODA (Vol. 3, pp. 132-139).

*R2 "4. Could Authors add the other models' fits to Fig. 1?"*
*R2 "2. Shouldn't the comparison between different models go to the main text?"*
*R3 "It might be better to visualize the edge weight distribution of two models and a real network (like Figure1) to show the characteristics of the new model."*

We have introduced the graphs corresponding to the two simplex models in the literature. In addition, we have added several elements to the description of the experiments, without repeating them in their entirety in the main text.

*R2 “Authors claim, unless I missunderstood something, that their model capture the multimodality, but Fig. 1 does not seem to support that, especially in all the outgoing cases.”*

Our approach does not completely reproduce the spectacular variations of real-world knowledge graphs. But unlike simplex models (whose points are completely “rectilinear”), our approach makes “waves” that cannot be explained by anything other than multimodality. In particular, it enables us to reproduce well the falls observed for low degrees (as for Wikidata). These drops are very important, as they concern a large proportion of entities.

Best regards,

Loïck Lhote, Béatrice Markhoff, Arnaud Soulet