



HAL
open science

Unveiling the Hidden Dynamics of Knowledge Graphs: The Role of Superficiality in Structuring Information

Cédric Sueur

► **To cite this version:**

Cédric Sueur. Unveiling the Hidden Dynamics of Knowledge Graphs: The Role of Superficiality in Structuring Information. Peer Community In Network Science, 2024, 10.24072/pci.networksci.100113 . hal-04602417

HAL Id: hal-04602417

<https://hal.science/hal-04602417>

Submitted on 5 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unveiling the Hidden Dynamics of Knowledge Graphs: The Role of Superficiality in Structuring Information

Cédric Sueur^{1,2,3}

1 Université de Strasbourg, IPHC UMR7178, CNRS, Strasbourg, France

2 ANTHROPO-LAB, ETHICS EA 7446, Université Catholique de Lille, Lille, France

3 Institut Universitaire de France, Paris, France

Cite as : Sueur, C. (2024) Unveiling the Hidden Dynamics of Knowledge Graphs: The Role of Superficiality in Structuring Information. *Peer Community in Network Science*, 100113. [10.24072/pci.networksci.100113](https://doi.org/10.24072/pci.networksci.100113)

Knowledge graphs [1–4] represent structured knowledge using nodes and edges, where nodes signify entities and edges denote relationships between these entities. These graphs have become essential in various fields such as cultural heritage [5], life sciences [6], and encyclopedic knowledge bases, thanks to projects like Yago [7], DBpedia [8], and Wikidata [9]. These knowledge graphs have enabled significant advancements in data integration and semantic understanding, leading to more informed scientific hypotheses and enhanced data exploration.

Despite their importance, understanding the topology and dynamics of knowledge graphs remains a challenge due to their complex and often chaotic nature. Current models, like the preferential attachment mechanism, are limited to simpler networks and fail to capture the intricate interplay of diverse relationships in knowledge graphs. There is a pressing need for models that can accurately represent the structure and dynamics of knowledge graphs, allowing for better understanding, prediction, and utilisation of the knowledge contained within them.

The paper by Lhote, Markhoff, and Soulet [10] introduces a novel approach to modelling the structure and dynamics of knowledge graphs through the concept of superficiality. This model aims to control the overlap between relationships, providing a mechanism to balance the distribution of knowledge and reduce the proportion of misdescribed entities. This is the first model tailored specifically to knowledge graphs, addressing the unique challenges posed by their complexity and diverse relationship types. The innovation lies in the introduction of superficiality, a parameter that governs the probability of adding new entities versus enriching existing ones within the graph. This model not only addresses the multimodal probability distributions observed in real KGs but also offers a more granular understanding of the knowledge distribution, particularly the presence of misdescribed entities. The authors validated their model against three major knowledge graphs: BnF, ChEMBL, and Wikidata. The results demonstrated that the generative model accurately reproduces the observed distributions of incoming and outgoing degrees in these knowledge graphs. The model successfully captures the multimodal nature and the irregularities in the degree distributions, especially for entities with low connectivity, which are typically the majority in a knowledge graphs.

One significant finding is the impact of superficiality on the level of misdescribed entities. The study revealed that lower superficiality leads to a more uniform distribution of relationships across entities, thus reducing the number of entities described by few relationships. Conversely, higher superficiality results in a higher proportion of entities with minimal descriptive facts, reflecting a paradox where increasing the volume of knowledge does not necessarily reduce the level of ignorance. The authors

also conducted an ablation study comparing their model to traditional models like Barabási-Albert [11] and Bollobás [12]. The results showed that the proposed multiplex model with superficiality parameters consistently outperformed these traditional models in accurately reflecting the characteristics of real-world knowledge graphs.

This research provides a groundbreaking approach to understanding and modelling the structure and dynamics of knowledge graphs. By introducing superficiality, the authors offer a new lens through which to examine the distribution and organisation of knowledge within these complex structures. The model not only enhances our theoretical understanding of knowledge graphs but also has practical implications for improving data storage, query optimisation, and the robustness of knowledge induction processes.

The introduction of superficiality opens several avenues for future research and application. One potential direction is refining the model to account for localised perturbations in smaller knowledge graphs or specific domains within larger knowledge graphs. Additionally, longitudinal studies could further elucidate the evolution of superficiality over time and its impact on the quality of knowledge representation. Another promising area is the application of this model in real-time knowledge graphs management systems. By adjusting superficiality parameters dynamically, it may be possible to optimise the balance between entity enrichment and the introduction of new entities, leading to more robust and accurate knowledge graphs. In the broader context of knowledge engineering and data science, this model offers a framework for exploring the vulnerability of knowledge graphs and their susceptibility to various types of biases and inaccuracies. This understanding could lead to the development of more resilient knowledge systems capable of adapting to new information while maintaining a high level of accuracy and coherence.

Overall, the concept of superficiality and the associated generative model represent significant advancements in the study and application of knowledge graphs, promising to enhance both our theoretical understanding and practical capabilities in managing and utilising these complex data structures. It would be interesting to see how this can be extended to domains in social network analyses [13,14].

References

1. Nickel M, Murphy K, Tresp V, Gabrilovich E. 2015 A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE* **104**, 11–33.
2. Ehrlinger L, Wöß W. 2016 Towards a definition of knowledge graphs. *SEMANTICS (Posters, Demos, SuCESS)* **48**, 2.
3. Hogan A *et al.* 2021 Knowledge graphs. *ACM Computing Surveys (Csur)* **54**, 1–37.
4. Ji S, Pan S, Cambria E, Marttinen P, Philip SY. 2021 A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems* **33**, 494–514.
5. Bikakis A, Hyvönen E, Jean S, Markhoff B, Mosca A. 2021 Special issue on semantic web for cultural heritage. *Semantic Web* **12**, 163–167.
6. Santos A *et al.* 2022 A knowledge graph to interpret clinical proteomics data. *Nature biotechnology* **40**, 692–702.
7. Suchanek FM, Kasneci G, Weikum G. 2007 Yago: a core of semantic knowledge. pp. 697–706.

8. Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z. 2007 Dbpedia: A nucleus for a web of open data. pp. 722–735. Springer.
9. Mora-Cantalops M, Sánchez-Alonso S, García-Barriocanal E. 2019 A systematic literature review on Wikidata. *Data Technologies and Applications* **53**, 250–268.
10. Lhote L, Markhoff B, Soulet A. 2023 The Structure and Dynamics of Knowledge Graphs, with Superficiality. (doi:10.48550/arXiv.2305.08116)
11. Barabási A-L, Albert R. 1999 Emergence of scaling in random networks. *science* **286**, 509–512.
12. Bollobás B, Borgs C, Chayes JT, Riordan O. 2003 Directed scale-free graphs. pp. 132–139. Baltimore, MD, United States.
13. Sueur C, King AJ, Pelé M, Petit O. 2013 Fast and accurate decisions as a result of scale-free network properties in two primate species. In *Proceedings of the European conference on complex systems 2012* (eds T Gilbert, M Kirkilionis, G Nicolis), pp. 579–584.
14. Romano V, Shen M, Pansanel J, MacIntosh AJJ, Sueur C. 2018 Social transmission in networks: global efficiency peaks with intermediate levels of modularity. *Behav Ecol Sociobiol* **72**, 154. (doi:10.1007/s00265-018-2564-9)