



Issues about entitymetrics and paper-entity citation network

Fengjun Sun¹ · Yingqiu Li¹ · Guojun Sheng¹ · Xiaolin Yao¹

Received: 1 November 2021 / Accepted: 14 February 2022 / Published online: 27 February 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

To the editor

The journal *Scientometrics* published a paper presenting an entitymetric analysis on COVID-19 (the COVID-19 paper, Yu et al., 2021), and the key part of the paper is an entity–entity co-occurrence network based on the bio-entities extracted from titles and abstracts of COVID-19 publications. According to the original paper proposing entitymetrics (the entitymetrics paper, Ding et al., 2013), there are two aspects of entitymetrics: (1) the co-occurrence of entities as mentioned the COVID-19 paper, and (2) citation relationship between entities. The COVID-19 paper extracted the entity co-occurrence relations instead of the citation relations. However, we believe there are some issues that need to be discussed about the citation relationship aspect of entitymetrics.

The entitymetrics paper (Ding et al., 2013) assumes that if one paper cites another paper, then an entity in the citing paper will be considered to cite an entity in the cited paper, and entity citation network is built based on the assumption (see Fig. 1). The first issue about this assumption is that where should the cited entities be extracted. Both the COVID-19 paper (Yu et al., 2021) and the entitymetrics paper (Ding et al., 2013) extracted the entities from the abstracts and titles of the PubMed Central (PMC) papers. However, instead of the titles and abstracts, we might consider extracting the entities from the full text, specifically the citation context. An abstract is a summary of a research article, and the title is the name for the work. Both are provided by the authors, whereas the citation contexts are the description of the cited paper from the citing authors' point of view. Even though from the authors' perspective, it seems that the most important entities should be in the titles and abstracts, however, the citing authors' perception could be different with the cited authors. The entities extracted from the titles and abstracts might not be in the citation contexts, and vice versa. One paper contains a lot of concepts, entities, and ideas, but only a few of them are used or cited by the citing authors. What entities are used is directly revealed or

✉ Fengjun Sun
sunfengjun@neusoft.edu.cn

Yingqiu Li
liyinqiu@neusoft.edu.cn

Guojun Sheng
shengguojun@neusoft.edu.cn

Xiaolin Yao
yaoxiaolin@neusoft.edu.cn

¹ Dalian Neusoft University of Information, Dalian 116023, People's Republic of China

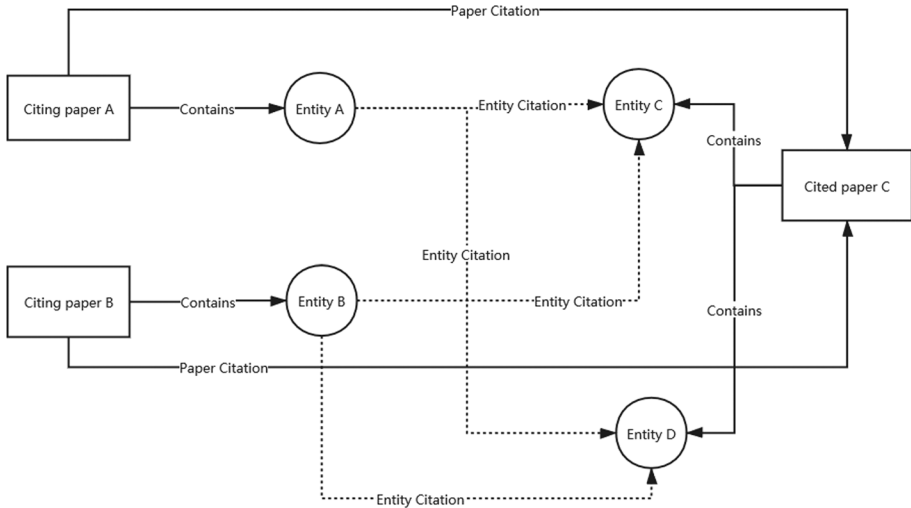


Fig. 1 The original entity–entity citation network based on paper–paper citation network. The figure is redrawn based on Fig. 3 from the entitymetrics paper (Ding et al., 2013)

described in the citation contexts. Thus, future studies about entitymetrics might consider using the entities extracted from the citation contexts as the cited objects. In one of the earliest studies about concept symbol (Small, 1978), to confirm the correctness of the theory, Henry Small conducted an experiment on the scientific literatures of chemistry by the scrutiny of the citation contexts of the citing documents. With the availability of the modern, machine-readable, and open-accessed full text of PMC papers, the time is ripe for large scale entity extraction and data analysis based on citation context dataset (Sun et al., 2021).

Another issue about the citation relationship aspect of entitymetrics is what the citing object should be. For the entitymetrics paper assumption, if the citing paper A has m entities and the cited paper B has n entities, then a m – n entity citation network will be built between the two papers. On the one hand, as we mentioned before, if the cited entities are only extracted from the corresponding citation context, then the number of the cited entities might be much smaller than n . On the other hand, among the m entities in the citing paper A, it is hard to identify which entity to be the citing entity. Even though the important entities might be in the titles, abstracts, conclusions, or other parts of the paper, it is still a little bit bold to link a citing entity to a cited entity extracted from citation context. Perhaps, the only certain fact is that the citing paper cites the cited entities extracted from its own citation contexts. In other words, instead of a unipartite network between entities, we might consider using a bipartite paper–entity network like the bipartite network between customers and products. In Fig. 2, for the cited paper C, both citing papers A and B cited entity D in their citation contexts; however, only paper A cites entity C and paper B does not cite entity C. Compared with Fig. 1, the original entity–entity citation network based on paper–paper citation network, paper to entity citation network is more concise. We do not have to figure out what entities are in the citing paper, and only entities in the citation context are extracted.

Entitymetrics can extend bibliometric methods by measuring the impact of knowledge entities in scholarly communication, and knowledge entities acting as carriers of knowledge in scientific articles have shown the potential to be the next big thing in scientometrics

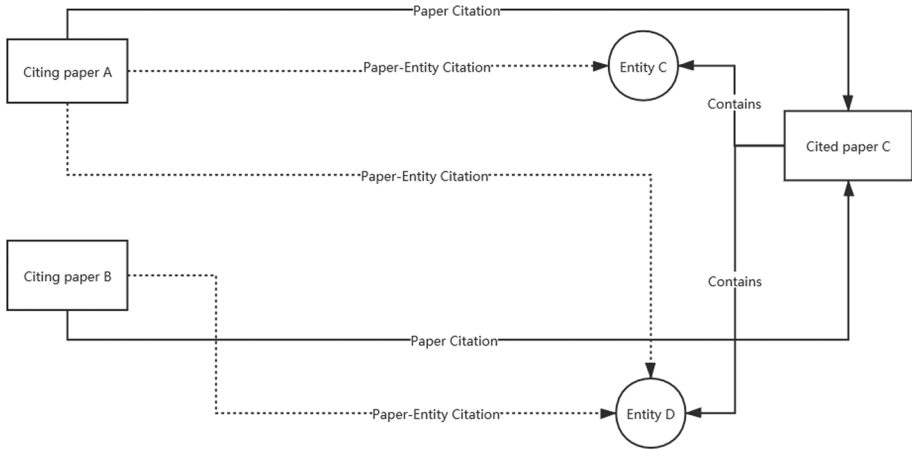


Fig. 2 Paper to entity citation network

and informetrics. However, the original entity–entity citation network can be extended with a paper–entity citation network by (1) replacing the original citing object entity with citing paper, and (2) limiting the cited entities to be the entities only extracted from citation contexts. The assumption is changed as such: if an entity is extracted from the citation context of a paper, then the entity is cited by the paper.

References

Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., & Chambers, T. (2013). Entitymetrics: Measuring the impact of entities. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0071416>

Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, 8(3), 327–340.

Sun, F., Li, Y., Sheng, G., & Yao, X. (2021). PubMed central citation context dataset (pp. 1063–1073). In *Presented at the proceedings of the ISSI 2021 international conference*.

Yu, Q., Wang, Q., Zhang, Y., Chen, C., Ryu, H., Park, N., et al. (2021). Analyzing knowledge entities about COVID-19 using entitymetrics. *Scientometrics*. <https://doi.org/10.1007/s11192-021-03933-y>