



A Semantic Representation of the Citation Structure

Marcin Skulimowski^(✉) 

Faculty of Physics and Applied Informatics, University of Lodz,
Pomorska 149/153, 90-236 Lodz, Poland
marcin.skulimowski@uni.lodz.pl

Abstract. A scientific citation is usually represented as a relation between two publications without any precise meaning and inner structure. In fact, the structure of a citation, which is usually not represented explicitly, can be quite complicated. In our previous papers, we have proposed so-called expanded citations which allow representing the structure of a citation in a machine-readable way. In this short paper, we present and discuss selected structures of citations. In particular, we consider their meaning and possible application.

Keywords: Citation relationship · Semantic publishing · Digital libraries

1 Introduction

A citation is a relation between two scientific publications¹. We can visualize it as an arrow from a node representing a citing publication to a node representing a cited publication. A collection of articles and citations between them form a directed graph called a *citation graph* or *citation network* [1]. A citation network analysis provides useful data for many research information systems. However, a citation is more than merely a relationship between two papers without any precise meaning and inner structure. Consequently, there is a vast amount of literature on the creation and analysis of citation content data (e.g. [2,3]). For example, the *citation context* can provide us with knowledge about the reasons for a citation. This knowledge allows us to add meaning to an arrow representing a citation. To this end, we can use the CiTO ontology, enabling *characterization of the nature or type of citations* [4]. In our opinion, we can do something more. After reading two papers (citing and cited), we can add meaning to relations between parts of papers (e.g. concepts, definitions, figures) linked by a citation. It is possible because we know which parts (entities) from a cited paper are *used* in a citing paper and how they are *used*. Moreover, we can name relations between

¹ Our considerations apply to any scientific publication. Throughout this paper, the publications are also referred to as papers, articles or books. We do not distinguish between them.

papers and entities from these papers. In this way, the *structure of a citation* emerges. The structure which is usually known to a reader but is not represented explicitly and machines cannot process it. Until recently, such a representation has not been possible. Nowadays, however, using semantic technologies, we can represent the *structure of a citation* in a machine-readable way. We have proposed such a representation based on the so-called *expanded citations* in our previous papers [5,6].

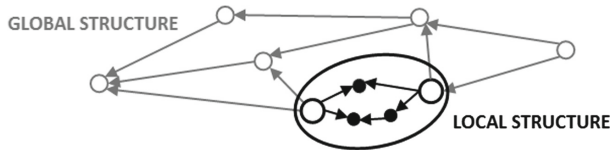


Fig. 1. The global and local structure of a citation network (\circ - publications, \bullet - entities from publications).

There is a vast amount of literature on citation networks and their *global* structure (see, e.g. [7]). In this paper, we are interested in the *local* properties of a citation network. Namely, we are going to present and shortly analyze the structures of individual citations containing not only papers but also entities from them (Fig. 1). The paper is organized as follows. Section 2 gives a brief overview of expanded citations. In Sect. 3, we analyze selected structures of citations. In particular, we discuss their meaning and consider whether the structures can be useful in the evaluation of scientist's work. The paper ends with a short discussion and the outline of future work.

2 Expanded Citations

A bibliographic citation links two articles (see Fig. 2a). Egghe and Rousseau [1] state: *the fact that a document is mentioned in a reference list indicates that in the author's mind there is a relationship between a part or the whole of the cited document and a part or the whole of the citing document*. Most studies have focused on a relationship between entire papers [1,4,7]. The point is that using expanded citations we can represent in a machine-readable form a relationship between parts of papers. Consequently, instead of one relation (*cites*) between two papers, we consider more relations between these papers and also between parts of them called *concepts* (see Fig. 2b). A *concept* we define as any entity (part) of a paper named with a URI (Uniform Resource Identifier) [5]. We assume that it is possible to assign a URI to each entity from a scientific publication (for details - see [5]). In the rest of this paper, a concept from a publication X we denote by C_X .

We are now ready to introduce the main definition of this paper (see [6] and references therein). Let A and B be two publications. We say that a citation

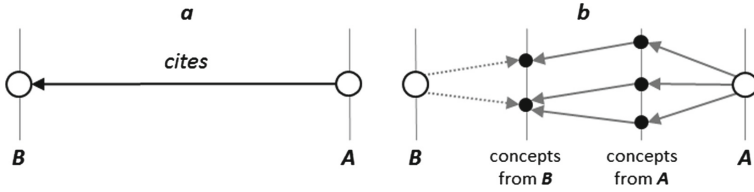


Fig. 2. A citation A cites B (a) and its exemplary structure (b).

$A \rightarrow B$ (A cites B) is *expandable* if there exist concepts C_A and C_B , relations r , r_A , r_B represented by object properties from some ontology O and the following RDF (Resource Description Framework) triples:

$$B r_B C_B. C_A r C_B. A r_A C_A. \tag{1}$$

We call the set of triples (1) an *expanded citation*. Moreover, an *expanded citation* created for a standard citation we call its *expansion*.

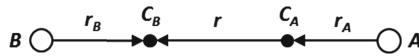


Fig. 3. An expanded citation.

Note that, using one expanded citation we can describe in RDF one “path” between nodes A and B representing articles (Fig. 3). Consequently, to represent a citation structure that can be made up of several “paths”, we may need a few *expansions* (compare Figs. 2b and 3). Moreover, note that in order to create expanded citations, we need terms from appropriate ontologies to add semantics to relations between publications and concepts. An example of such an ontology is CiTO [4].

3 Citations and Their Structures

Let us now consider three examples of citations and their structures².

- (I) Citing entity: DOI 10.2478/p1c-2013-0010
 Cited entity: ISBN 10 0195070038
 Citation context: *This process is called “knowledge construction” (e.g. Rogoff, 1990).*
 The author refers to the concept of *knowledge construction* introduced in the cited article. We present in Fig. 4-I an expansion for this citation.

² For simplicity, we do not use URIs in our examples. We also do not use object properties from any particular ontology.

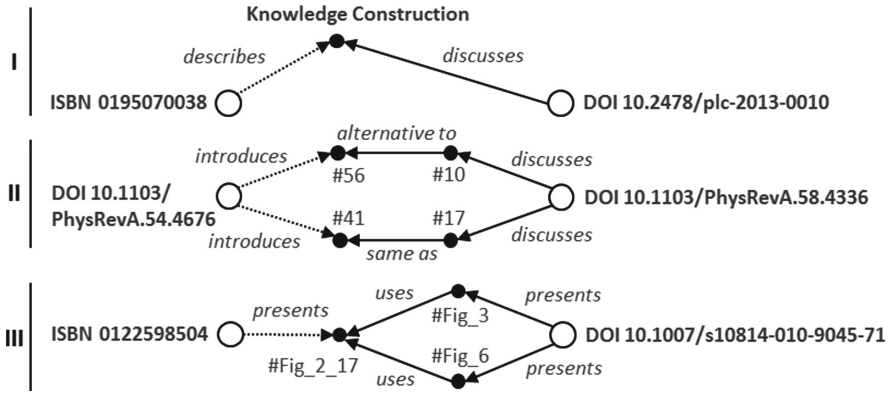


Fig. 4. The structures of citations from examples A-C.

- (II) Citing entity: DOI 10.1103/PhysRevA.58.4336
 Cited entity: DOI 10.1103/PhysRevA.54.4676
 Citation contexts: *Grot, Rovelli and Tate [13] introduced a regularized self-adjoint operator and considered the full expression (10) for possible application to more general states having both positive and negative momenta but vanishing in the proximity of $p = 0$.
 Grot, Rovelli, and Tate [13] (...) produce a “regularized” self-adjoint time operator T_ε with eigenstates (...), $\langle p|T_\pm\rangle_\varepsilon = \dots$ (17).
 Formulas (10) and (17) from the citing paper are in well-defined relations with formulas (56) and (41) from the cited paper. Hence, we represent the structure of this citation by two expanded citations - see Fig. 4-II.*
- (III) Citing entity: DOI 10.1007/s10814-010-9045-7
 Cited entity: ISBN 10 0122598504
 Citation contexts: *Fig. 3 Ground plans Ground plans of prehispanic houses from Oaxaca redrawn and adapted from the following sources: (...) (Flannery and Winter 1976, Fig. 2.17)(...) Fig. 6 Two extensively and meticulously excavated houses.(...) Flannery and Winter 1976, Fig. 2.17(...).*
 Figures 3 and 6 use figure 2.17 from the cited paper. The structure of this citation is presented in Fig. 4-III.

The structures of citations presented in Fig. 4 do not exhaust all possibilities of connection between two papers. Let us consider what structures, in general, are possible in this case. We limit ourselves to structures of citations with at most *two expansions*. All possible structures, in this case, are presented in Fig. 5. The citation presented above in example (I) has structure **1** (*1-chain*). Citations having this structure are often used in *Introduction*, *Related Works* or *Discussion* sections. Note that, in structure **1**, paper *A* refers directly to C_B . If C_B is somehow used in *A*, then there may exist a concept C_A in some relation with C_B . A citation has then structure **2** (*2-chain*). Structure **3** (*diamond*) corresponds to the situation when *A* directly refers to two concepts from *B*. However, these con-

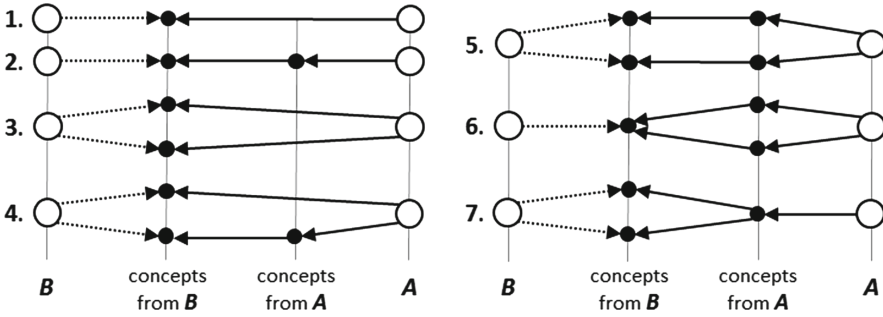


Fig. 5. The structures of citations with at most 2 expansions.

cepts are not “linked” to any concepts from A. In turn, structures 4 (*pentagon*) and 5 (*hexagon*) contain concepts from A related to concepts from B. Structure 5 contains two concepts from A. This structure appeared already in the above example (II). In structure 4, a concept from B is related to a concept from A. Moreover, another concept from B is only *discussed* or *mentioned* in paper A. It is important to note that two expansions of a citation may *overlap*. For example, they may contain the same concept. In structure 6 two different concepts from A are linked to the same concept from B. By analogy to *bibliographic coupling* [7] we can say that the two concepts from A are *conceptually coupled* because they both are linked to the same concept from B. Figures #Fig_3 and #Fig_6 from our example (III) are *conceptually coupled* (see Fig. 4-III). In structure 7 a concept from A is linked to two concepts from B. In this case, by analogy to *co-citation* [7], we can say that two concepts from B are *co-used* by a concept from A. It seems reasonable to assume that all structures 1–7 may appear in practice. So one may ask the question: *how can we use them?* Let us now consider whether the structures of citations can be useful in the evaluation of scientist’s work.

Nowadays, in the evaluation of a scientist’s work, only the presence of a citation is taken into account [1]. The structure (meaning) of a citation is ignored. However, the structure may contain important information. It is reasonable to assume that for authors particularly valuable are those publications of others in which concepts (e.g. propositions, approaches, formulas) from their publications are somehow *used*, i.e. they are related to (new) concepts from citing publications. We may say that such *used* concepts *contribute to the progress of science*. On the contrary, citations of the form: *In the literature, there are many examples of...* placed in the *Introduction* section are of less value. Nowadays, these citations are treated equally. The knowledge of a *citation structure* enables us to distinguish between them. Without considering the details of the structures, we can assume that *the more concepts (from a citing paper) in a citation structure, the higher the value of a citation*. Using this “rule”, we can sort the structures presented in Fig. 5 (in importance increasing order) by their values as follows: **1, 3, 2, 4, 7, 6, 5**. Note that the above considerations are valid, assuming that citations are *positive* or at least *neutral*. However, this is not necessarily the

case. There are also *negative* citations which may indicate problems or flaws in the work or an opposing viewpoint. For example, in paper *A*, two counterexamples to a statement proposed in *B* can be given. This negative citation has structure **6**, which is very valuable according to the above list. Thus, for negative citations, the proposed “rule” of citation evaluation does not apply.

4 Discussion and Future Work

Expanded citations allow representing the structure of a citation in a machine-readable way. This possibility opens new perspectives for processing (searching and visualizing) of scientific domains and the evaluation of a scientist’s work (see [6] and references therein). However, the picture is still far from completeness. Further studies are needed to estimate how large is the class of expandable citations. Does the size of this class depend on the scientific domain? In order to create an expanded citation, we need appropriate ontologies. Consequently, future work should also determine ontologies and terms from them useful in expanded citations. Another critical issue for future studies is to determine the extent to which machines can support the creation of expanded citations. The results obtained in the automatic classification of citation function using CNNs (Convolutional Neural Networks) and NLP (Natural Language Processing) suggest that machine support in the creation of expanded citations cannot be ruled out [8].

References

1. Egghe, L., Rousseau, R.: Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science. Elsevier Science Publishers, Amsterdam (1990)
2. Kogalovsky, M., Krichel, T., Lyapunov, V., Medvedeva, O., Parinov, S., Sergeeva, V.: Open citation content data. In: Garoufallou, E., Sartori, F., Siatiri, R., Zervas, M. (eds.) MTSR 2018. CCIS, vol. 846, pp. 355–364. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14401-2_34
3. Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., Zhai, C.: Content-based citation analysis: the next generation of citation analysis. *J. Assoc. Inf. Sci. Technol.* **65**(9), 1820–1833 (2014)
4. Peroni, S., Shotton, D.: FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *J. Web Semant.* **17**, 33–43 (2012)
5. Skulimowski, M.: On expanded citations, In: Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business, i-KNOW, pp. 38:1–38:4. ACM (2014)
6. Skulimowski, M.: The flows of concepts. In: Proceedings of the International Conference on Knowledge Management and Information Sharing, KMIS, vol. 3, pp. 292–298 (2015)
7. Fang, Y., Rousseau, R.: Lattices in citation networks: an investigation into the structure of citation graphs. *Scientometrics* **50**, 273–287 (2001)
8. Bakhti, K., Niu, Z., Yousif, A., Nyamawe, A.S.: Citation function classification based on ontologies and convolutional neural networks. In: Uden, L., Liberona, D., Ristvej, J. (eds.) LTEC 2018. CCIS, vol. 870, pp. 105–115. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95522-3_10