Table 1: Methodology comparison

	(Gordon and Lindsay, 2002)	Our Approach
Process	Requires human intervention	Fully automatic
Concepts	Bi-grams only	Uni-grams, abbreviations and variable length n-grams
Techniques	Lexical Statistics	Lexical Statistics & Distributional Semantics
Output	List of bi-grams	Detailed contextualised semantics groupings

LA research. Therefore, the suggestions provided through our approach will greatly influence to uplift the process of research in LA. In comparison with the only existing CS-related LBD approach (Gordon and Lindsay, 2002) our approach utilises an improved methodology (Table 1).

To the best of our knowledge, this is the first non-medical LBD study that utilises neural word embeddings to detect the target C-concepts. Our initial results demonstrate the importance of exploring neural word embeddings to effectively identify potential cross-disciplinary knowledge associations buried in literature. We would like to further enhance our existing approach by considering the below-mentioned future directions that are categorised based on our four research questions (RQ) described in Section 4.

RQ 1 (Content Analysis): A subtle analysis of literature is needed to accurately capture the hidden knowledge associations. In our current experiment, we are considering concepts at keywords-level by identifying seed concepts. As an improvement, we would like to have an organised topic structure with different levels of granularity. In order to achieve that, we are intending to utilise semantic web technologies and pre-existing topical categories (e.g., Dewey Decimal Classification) to enhance the understanding of the content. Moreover, the identified topic structure will also be useful to provide a clearly structured, logical output to the user than merely listing the identified associations. Due to the lack of LBD research that

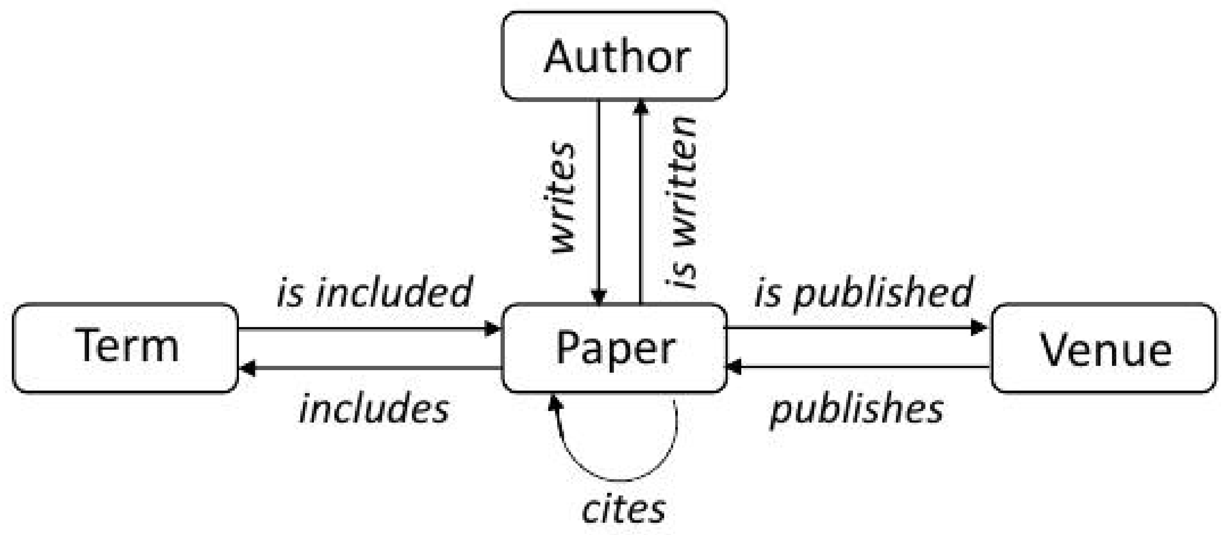

Figure 3: Entity & relation types (Shakibian and Charkari, 2017).

analyse the effects of topic modeling (Sebastian et al., 2017b), it is also important to study how topical information propagate among research publications to detect interesting, implicit knowledge associations. Another interesting future direction would be to utilise deep language understanding techniques to infer ontologies from the scientific literature automatically which can be utilised to identify more granular knowledge associations.

RQ 2 (Bibliometrics Analysis): In our current experiment, we are utilising the popular ABC model to discover the knowledge associations. However, the inference steps introduced through ABC model is simple and not foolproof. Therefore, in our future research studies, we are intending to analyse more complex inference steps to identify complex knowledge associations that cannot be identified through ABC model. To achieve that, we are aiming to integrate a graph-based approach by analysing the relationships among the four entity types (i.e. *author*, *term*, *paper*, *venue*) illustrated in Figure 3. In other words, we are intending to utilise different bibliographics-based link structures such as co-author relationships, direct citation links, co-word analysis, bibliographics coupling, and co-citation links to uncover complex knowledge associations. For example, when authors from disjoint research fields collaborate for a research, it implies a potential association between the two knowledge areas. This simple co-author relationship can be further expanded to more complex associations by analysing shared authors in the citations of the source and target literature, analysing authors in source literature that are cited by the target literature etc. Same as for the *author* entity, this procedure can be followed for the remaining entities (i.e. *paper*, *term*, *venue*) of the network schema in Figure 3 to derive more complex and implicit associations. With regards to *term* entity, the identified associations can be further expanded by leveraging topic modeling and