A Graph-based Visualization approach for Related Work Exploration, Claim Detection, Citation Recommendation, and Research Footprint Exploration

Nishan Chatterjee

April 15, 2023

1 Introduction and Research Goals

Scientific papers are important for researchers and scholars to track trends and build on pre-existing methodologies to improve and integrate with their research. With the increase in scientific papers, it is difficult for scholars to read all the papers to extract emerging or noteworthy knowledge [1]. Paper citations can take place for reasons, not excluding methodology, experiment settings, experimental results, conclusions, equations, etc.

The following is a proposal for creating an end-to-end approach which allows researchers to: (a) Explore Related work by highlighting the relationship between each research paper with their settings being highlighted allowing for an improved contextualization, (b) Claim Detection that checks the relation between the research that they cite by highlighting sections of the paper (including a statement, methodology, figures, conclusions, etc) [2], (c) Automatic Citation Recommendation while providing clarifications for which work is being referred to (also borrows on work from (a) and (b), and Research Footprint Exploration, that highlight paper insights like author-coauthor relationships [3], self-citing, statistics about the paper and authors like the h/g/i10-indices that can help gauge the research benefits of a paper and associated research.

The proposed approach attempts to integrate methodologies of existing methodologies like Relation Classification, Automatic Text Highlighting, Task Driven Knowledge Graph Filtering, Paraphrasing, and Knowledge Graph Visualization. The proposed approach attempts to address both the tracking of research in context with the ever-growing academic field of Natural Language Processing and Artificial Intelligence, as well as provide insights to authors to reduce paper rejections because of inadequate literature review, duplicate work, out-of-date research, over self-citing, etc [4].

2 Related Work

Paper modelling can help scholars master the key information in scientific papers, and relation classification (RC) between entity pairs is a major approach to paper modelling [1]. Classification approaches have shown that it's possible to classify citations depending on their semantic relation type [5][6][7]. [8] also highlight citation recommendation methods using Knowledge Graphs to suggest researchers on related research by suggesting relevant research. Other key methods like Plagiarism Detection have also been used to help researchers gain insights into finding and detecting comparable research. Paraphrastic tools have also been used for automatic Plagiarism Detection [9]. Task-driven Knowledge Graph filtering methods have also been shown to improve potential solutions where there are too many candidate solutions [10], [11].

Previous and ongoing work at building large-scale relational graphs in the past like Connected Papers [12] and the "Web of Science" [13] help researchers graphically explore and represent research relevant to their work. Similarly, "Touchgraph Navigator" is another approach that performed cluster computation to reveal inherent groupings, along with co-citations and co-occurrence analysis to clarify dense networks [14]. Finally, in 2012, Elsevier's Application Development Challenge in China was won by the graduate students from Institute of Computing Technology at Chinese Academy of Sciences and Fudan University for their "Aerial Viewer" application that allowed users to quickly understand a compilation of related industry research papers, authors and organizations. The app generated a detailed, multi-dimension framework of information including statistical data, journals, conferences,

published papers, and visual maps of frequently used keywords to improve understanding of an entire research footprint [15].

3 Methods: Components of the approach

- Related Work Exploration
  - A graph-based visualization of all citations, and related works (and the citations in the related works, and their related works and so on).
  - Browse through academic papers in a visual high-dimensional format instead of a linear list of text format that most academic websites offer.
  - Interactive filtering of these high-dimensional network-graphs of papers switching from paper-content-view to relational network view.
  - The approach models the evolution of a particular methodology over the years and in what way (for example, the evolution of different neural network architectures, embedding techniques, etcetera), and which part of the research from an old work a new work builds upon and in what way.
  - Automatically providing code links similar to the work of Catalyzex.com, and citing research topics which suffer from reproducibility (thereby also providing students with topics and a platform to work on for improving the Replication Crisis).
- Claim Detection
  - The knowledge graph contains relations between two papers not only based on citations but also attempts to classify the semantic relation between these citations.
  - Nuance highlighting: A paper can contain specific conditions where the research holds, and there can be supporting arguments for a particular parameter/hyperparameter/architecture setting that needs highlighting to support the claim. Under the nuance detection task, the approach should encapsulate and highlight for

instance experiment conditions from claims and table results, as well as highlight settings/experiment conditions (if mentioned in the paper), where a particular observation doesn't hold. [1]

- Citation Recommendation
  - Lookup function based on Related Work Exploration and Claim Detection mentioned immediately above.
- Research Footprint Exploration
  - Filtering of data based on papers indicating paper statistics, author statistics, and relation modelling to understand what kind of research different research groups are participating in.
  - Allowing researchers access to tag papers suffering from explainability or fairness by creating research areas, especially for researchers who have not worked in the field for longer periods.

4 Challenges

The topic, however, presents several challenges. The following challenges are listed in no specific order of priority, since this would depend on whichever goal from "Methods: Components of the approach" we start with.

- Measuring the effectiveness or usefulness of such an approach.
- The task of relevant text highlighting requires annotations (the task built on top of relation classification). While this can be crowdsourced, especially from people who're already working on the field, as [2] highlights, highlighting is effective in reducing classification effort but does

[1] Note: This is a very specific example use-case as can be seen in this paper [16], to quote the "4.2 Results and Comparison", "explicit embeddings may be favoured over implicit ones whenever the explicit vector representation is required by some algorithms which require more than the dot product. Indeed, graph kernels are limited to kernel methods such as SVM. On the other hand, though, implicit ones are usually defined between two graphs whereas most of the explicit methods require the whole dataset to compute the graph embedding. So, implicit methods may be favoured other explicit one whenever the access to the whole dataset is limited.". This can be highlighted as a nuance condition since it's an either or case where explicit methods may or may not be better than implicit embedding methods in certain cases.

not improve accuracy - and in fact, low-quality highlighting can decrease it.

- Automatic paper-relationship extraction, i.e. the task of finding out what specific improvements or changes a work makes over existing research is a complex topic that may be extremely difficult to accomplish, if at all possible.
- The most simplistic approach to the task of Nuance detection would be a combination of looking at the abstract, the results, the conclusion and discussions and then referencing the tables to check experimental results, but that may be insufficient.
- Relational Nuance modelling would be a difficult challenge.
- There's a variety of research papers from survey papers to experiment papers. Not all the tasks highlighted above hold for every one of them and contextualization can be an issue.
- Finding a seamless UI transition from text in papers to graphical overviews.
- A Metric to determine the quality of the graph-network created.
- Filtering the graph can be a problem. For example, researchers new to the field also face challenges of domain and resource identification that applies to the filtering setting proposed in component 4 of the research goals of modelling the approach. In addition, classification tasks in relevant text highlighting also include the need of being sample efficient while looking at the different predictor variables used for classifying a task.

References

[1] Zhongbo Yin, Shuai Wu, Yi Yin, Wei Luo, Zhunchen Luo, Yushani Tan, Xiangyu Jiao, and Dong Wang. Relation classification in scientific papers based on convolutional neural network. In *Natural Language Processing and Chinese Computing*, pages 242–253. Springer International Publishing, 2019. doi: 10.1007/978-3-030-32236-6_21. URL https://doi.org/10.1007/978-3-030-32236-6_21.

[2] Jorge Ramírez, Marcos Baez, Fabio Casati, and Boualem Benatallah. Understanding the impact of text highlighting in crowdsourcing tasks, 2019. URL https://arxiv.org/abs/1909.02780.

[3] Yang Chen, Cong Ding, Jiyao Hu, Ruichuan Chen, Pan Hui, and Xiaoming Fu. Building and analyzing a global co-authorship network using google scholar data. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, page 1219–1224, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee. ISBN 9781450349147. doi: 10.1145/3041021.3053056. URL https://doi.org/10.1145/3041021.3053056.

[4] Siobhan Bowler. Common reasons why academic papers are rejected by journal writers. https://www.deakin.edu.au/__data/assets/pdf_file/0011/269831/reasons_papers_rejected-_24.08.pdf.

[5] Sara Mifrah and Ben Habib. Semantic relationship study between citing and cited scientific articles using topic modeling. pages 1–8, 10 2019. doi: 10.1145/3372938.3372943.

[6] Stephen P. Harter, Thomas E. Nisonger, and Aiwei Weng. Semantic relationships between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science*, 44(9):543–552, October 1993. doi: 10.1002/(sici)1097-4571(199310)44:9⟨543::aid-asi4⟩3.0.co;2-f. URL https://doi.org/10.1002/(sici)1097-4571(199310)44:9<543::aid-asi4>3.0.co;2-f.

[7] Yicong Liang, Qing Li, and Tieyun Qian. Finding relevant papers based on citation relations. In *Web-Age Information Management*, pages 403–414. Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-23535-1_35. URL https://doi.org/10.1007/978-3-642-23535-1_35.

[8] Arthur Brack, Anett Hoppe, and Ralph Ewerth. Citation recommendation for research papers via knowledge graphs, 2021. URL https://arxiv.org/abs/2106.05633.

[9] Alberto Barrón-Cedeño, Marta Vila, M. Martí, and Paolo Rosso. Plagiarism meets paraphrasing: Insights for the next generation in automatic

plagiarism detection. *Computational Linguistics*, 39(4):917–947, December 2013. doi: 10.1162/coli_a_00153. URL https://doi.org/10.1162/coli_a_00153.

[10] Mojtaba Nayyeri, Sahar Vahdati, Xiaotian Zhou, Hamed Shariat Yazdi, and Jens Lehmann. Embedding-based recommendations on scholarly knowledge graphs. In *The Semantic Web*, pages 255–270. Springer International Publishing, 2020. doi: 10.1007/978-3-030-49461-2_15. URL https://doi.org/10.1007/978-3-030-49461-2_15.

[11] Veronika Henk, Sahar Vahdati, Mojataba Nayyeri, Mehdi Ali, Hamed Shariat Yazdi, and Jens Lehmann. Metaresearch recommendations using knowledge graph embeddings. 2018.

[12] Connected papers. https://www.connectedpapers.com/about.

[13] Web of science - citation map. https://images.webofknowledge.com/WOKRS529AR7/help/TCT/h_citation_map.html.

[14] Touchgraph navigator. https://www.touchgraph.com/navigator.

[15] Dynamic visualizer of research footprints wins elsevier's application development challenge in china. https://www.elsevier.com/about/press-releases/archive/science-and-technology/dynamic-visualizer-of-research-footprints-wins-elseviers-application-development-challenge-in-china.

[16] Donatello Conte, Jean-Yves Ramel, Nicolas Sidère, Muhammad Muzzamil Luqman, Benoît Gaüzère, Jaume Gibert, Luc Brun, and Mario Vento. A comparison of explicit and implicit graph embedding methods for pattern recognition. In Walter G. Kropatsch, Nicole M. Artner, Yll Haxhimusa, and Xiaoyi Jiang, editors, *Graph-Based Representations in Pattern Recognition*, pages 81–90, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-38221-5.