Science maps for information retrieval

Bascur Cifuentes, J.P.

Chapter 6

Conclusion

In this dissertation we have attempted to partially fill the knowledge gap that exists in the literature on the performance of science maps for information retrieval. In the current chapter, we will answer the research questions that we presented in Chapter 1 and explore potential further research on this topic.

6.1 Answers to research questions

Research question 1: How can science maps be designed to support information retrieval?

We answered this question in Chapter 2 by implementing the tool SciMacro (Scientific Macroscope), which allows a user to navigate a science map of academic documents in a way that is conductive for information retrieval by using the principles of the Scatter-Gather method. From this research, we learned that there are no significant hindrances for implementing information retrieval in a science map, at least in the way we implement it. However, we found two minor challenges. The first is how to communicate relevant information using the bubble chart visualization of the science map, which we addressed by placing the related clusters together and minimizing white space. The second challenge is how to let the user control the granularity of the clusters, which we addressed by letting the user decide on the number of clusters they desire. Then, in the back end, we produced several clustering solutions with different resolutions until we found one that generated a good distribution of cluster sizes for this number of clusters (this is the slowest step and it has the greatest potential for improvement). After we had found this resolution, we merged the clusters until we got the number of clusters that the user desired.

Research question 2: How effective are science maps for producing systematic reviews?

We found in Chapter 3 that science maps are more effective than Boolean queries for about half of the evaluated systematic reviews, which is a good performance given the stringent conditions of the experiment (i.e., because the Boolean queries define the relevant documents, the baseline has perfect recall). This, plus our finding that the intersection between the sets of documents retrieved by the Boolean query and the ones retrieved by science maps is small, shows that one approach cannot replace the other, and ideally both should be used together for greatest effectiveness. We also found that science maps can correct for some shortcomings of the Boolean queries, like finding documents that the original authors missed. An interesting observation is that there was no topical difference between the set of systematic reviews where science maps performed better than the Boolean queries and the set of systematic reviews where they performed worse. This observation motivated research question 3.

Research question 3: Do science maps represent some topics better than others?

We found in Chapter 4 that some ontological categories of topics are systematically clustered better than others, in particular the ontological topic categories "Diseases" and "Organisms", and that this happens in both citation and text similarity networks. Therefore, the answer to this research question is positive. For information retrieval tasks, this means that it is possible to know beforehand if a science map approach is likely to be helpful, which makes science maps a more reliable information retrieval tool. We were surprised that citation and text similarity networks perform well for the same topic categories because this suggests that the clusters of both maps would be about more or less the same topics. However, we also found differences between these networks. For higher granularity and Coverage (i.e. higher Coverage means higher recall), citation networks yield better results than text similarity networks, and vice versa. We believe this might be due to the simplicity of the text similarity metric that we used (i.e. it only measures shared words between documents and does not measure more subtle similarities like semantic similarity). It seems that creating good clusters at higher granularity and Coverage is more difficult than at lower, and so a more sophisticated text similarity metric might be needed.

Research question 4: How can the representation of specific topics be improved in a science map?

We answered this question in Chapter 5 by using different types of academic documents networks, based on data from different sources, to create science map clusters. This allowed us to influence which topic categories were the best clustered in a science map. Given that both text and citation networks yield similar results in terms of which topics are best clustered (as we found in response to research question 3), we used a text similarity network as a baseline (instead of a citation network or both networks). We compared the new networks with the baseline network to measure both the changes regarding the cluster quality of the topics and changes regarding which topics are best clustered in the new network. The biggest improvement in clustering effectiveness happened in topics related to geographical entities in the document authors network. The other noteworthy improvements were health topics in the Facebook users network, biotechnology topics in the patent families network, government and social topics in the policy documents network, food topics in the Twitter conversations network, and nursing topics in the Twitter users network. However, most of the topics that achieved the highest clustering effectiveness in their networks still achieved lower clustering effectiveness than in the text similarity networks, which defeats the purpose of improving the clustering effectiveness of the topic. A notable exception was the network that mixed text similarity with Twitter conversations. The topics obtained in this network had a clustering effectiveness comparable with text similarity, and even better for topics about food. Apart from this exception, we have not found a way to influence which topics are better represented in a science map without decreasing the quality of the clustering.

Overarching research question: What is the effectiveness of science maps for information retrieval, and how can we enhance it?

We studied science maps that are based on document clusters, using documents mostly from the biomedical field of science. These science maps have been shown to be effective for finding the relevant documents of systematic reviews, and to perform particularly well on topics that belong to the ontological topic categories "Diseases" and "Organisms". The effectiveness of a science map can be enhanced by turning the map into an interactive visualization of the clusters, where the user can create a new visualization based on the documents in selected clusters and control the granularity of the map.

6.2 Further research

Follow ups to our findings

As discussed in the introduction, the research agenda set out in this dissertation is focused on evaluating and improving science maps for information retrieval. With regard to evaluation, we limited ourselves to systematic reviews and academic topics, but further research can also explore other information retrieval tasks, such as exploratory search tasks. With regard to improvement, we found that using different networks from different sources has the potential for influencing which topics are best represented, but only the network that mixed text similarity with Twitter conversations could achieve a performance that is as good as the performance of citation or text networks. We believe that a performance similar to the latter networks might be achieved by further refinement, for example by cleaning the data before creating the network (like removing bot users from Twitter), by creating the network with a different methodology (like normalizing the weights of the edges), or by mixing networks with a different criterion (like weighing one network more than the other). The issue of which ontological categories of topics are best represented in a science map has received only limited attention in the literature [76, 131], and future research in this area could provide new insights.

Clustering

A relevant topic that we did not research in this dissertation is the clustering algorithm [74] . The Leiden algorithm is the most popular one, but the MALBA algorithm [75] was created specifically to outperform the Leiden algorithm in field delimitation, and future research could use the methodology that we developed in Chapter 5 to compare them.

Large Language Models

Thanks to accelerating developments in large language models (LLMs), we believe that the text representation of documents will take a more prominent role in the creation of science maps. We can imagine that there could be a fine-tuned text embedding model for each of the ontological categories of topics that we analyzed (for example, there are over 6,000 pre-trained Sentence Transformer models available in the Hugging Face website [132]). Another area where these text processing methods can be used is in the cluster labeling, as shown by van Eck and Waltman [159], who labeled clusters by providing ChatGPT with their top 250 most cited documents. Additionally, entity recognition, which allows us to extract data directly from the documents, could improve science maps in unforeseen ways. Also, even though we did not compare it directly, our results in Chapter 5 strongly suggest that text similarity networks based on text embedding create better clusters than networks based on less advanced text processing methods.

Beyond text representation, LLMs are also relevant to science maps due to developments in retrieval-augmented generation (RAG), a method that retrieves documents to improve the quality of question answering of LLMs. The use of RAG for academic information retrieval is still an emerging field of study [22], but recent results show promise [9]. Also, Asai et al. [8] developed OPENSCHOLAR, a RAG tool specific for academic search. We believe that RAG does not replace science maps, but instead they complement each other, with science maps visualizing the RAG results and putting them in context. This search approach is already implemented in platforms such as Zeta Alpha [183].

Granularity

An important open issue in science mapping is the choice of granularity, understood as the level of detail of the map, usually corresponding to the size of clusters. There is no agreed-upon answer in the field, and accordingly, this dissertation addressed granularity in several different ways rather than fixing it to a single definition. In Chapters 2 and 3 it was controlled by a hypothetical user and by a user model, respectively. In Chapter 4 and 5 it was used to make fair comparisons, with the former centered on map granularity (size of clusters) and the latter focused on topic granularity (number of selected clusters). Other researchers have proposed different strategies: Sjögårde and Ahlgren [142] searched a granularity that would group the references of a review article into a single cluster, Held and Gläser [75] developed an algorithm to determine an adequate level based on network properties, and Ficozzi et al. [61] explored maximum granularity by representing each document individually, physicalized as a 100-square-meter floor mat. These diverse approaches show that granularity remains an open question, but also that it is central to making science maps useful for information retrieval.

Prototyping

We believe that the critical next step in research for science maps for information retrieval is the further development of prototypes. This would allow evaluating the performance of science maps with real users. This has the added benefit that, by showing concrete uses of science maps, it can bring additional interest to continue and support the research and sustainability of the software. We find this important because most of the proposals for academic information retrieval tools that we found in literature, even the ones we found promising, are currently unusable due to lack of maintenance. This could be achieved by collaborating with already existing academic information retrieval platforms, such as Web of Science, Scopus, Dimensions, Zeta Alpha, Semantic Scholar, Google Scholar, or OpenAlex. However, it is worth pointing out that evaluating the performance of science maps with real users is not a trivial task. Such evaluation of interactive information retrieval requires careful experimental design and the participation of field experts [94].

Trends

In this dissertation we have provided evidence and advice on how to make information retrieval with science maps a more viable option for academic users. Fortunately, since the start of our research, we have seen that bibliometrics enhanced information retrieval has gained popularity among researchers, and the open science movement is lobbying to make the metadata of academic documents openly available, which will make science maps more viable. We hope that our research will further strengthen these developments and will help support and popularize science maps for information retrieval.