



# KEEPING UP WITH THE LITERATURE

Scientists need not fear missing out on developments in their field, thanks to software tools that highlight the latest key papers. **By David Matthews**

Every time Eddie Smolyansky had a few moments to himself, he tried to stay abreast of new publications in his field. But by 2016, the computer-vision researcher, who is based in Tel Aviv, Israel, was receiving hundreds of automated literature recommendations per day. “At some point the bathroom breaks weren’t enough,” he says. The recommendations were “way too much, and impossible to keep up with”.

Smolyansky’s ‘feed fatigue’ will be familiar to many academics. Academic alert tools, originally designed to focus attention on relevant papers, have themselves become a hindrance, flooding the inboxes of scientists worldwide.

“I haven’t even been reading my automated PubMed searches lately because it really is

overwhelming,” says Craig Kaplan, a biologist at the University of Pittsburgh in Pennsylvania. “I honestly cannot keep on top of the literature.”

But change is afoot. In 2019, Smolyansky co-founded Connected Papers, one of a new generation of visual literature-mapping and recommendation tools. Other services that promise to tame the information overload, integrating Twitter feeds and daily news as well as research, are also available.

## Origin story

Instead of serving up a daily list of new articles by e-mail, Connected Papers uses a single, user-chosen ‘origin paper’ to build a map of related research, based partly on overlapping citations. The service recently surpassed one

million users, Smolyansky says.

The maps are colour-coded by publication date, and users can toggle between ‘prior’, seminal, papers and later, ‘derivative’, works that build on them. The idea is that scientists can search for an origin paper that interests them, and see from the resulting map which recent papers have made a splash in their field, how they relate to other research, and how many citations they have accrued.

“You do not have to sit on the hose of papers and look at every paper that comes out for fear of missing it,” says Smolyansky. The tool is also helpful when scientists want to dive into an entirely new field, he adds, providing an overview of the essential literature.

Another visual-mapping tool is Open

## Work / Technology & tools

Knowledge Maps, a service offered by a Vienna-based not-for-profit organization of the same name. It was founded in 2015 by Peter Kraker, a former scholarly-communication researcher at Graz University of Technology in Austria.

Open Knowledge Maps creates its maps based on keywords rather than a central article, and relies on text similarity and metadata to work out how papers are related. The tool arranges 100 papers in similar subfields into bubbles whose relative positions suggest similarity; a search for articles on ‘climate change’, for example, might yield a related bubble about ‘risk cognition’.

Maps of these bubbles can be built in about 20 seconds, and users can change them to include the 100 most recently published papers of relevance, or other resources. Open Knowledge Maps includes not only journal articles, but also content such as data sets and research software. Its users have created more than 400,000 maps so far, says Kraker.

Amie Fairs, who studies language at Aix-Marseille University in France, is a self-proclaimed Open Knowledge Maps enthusiast. “One particularly nice thing about Open Knowledge Maps is that you can search very broad topics, like ‘language production’, and it can group papers into themes you may not have considered,” Fairs says. For example, when she searched for ‘phonological brain regions’ – the areas of the brain that process sound and meaning – Open Knowledge Maps suggested a subfield of research about age-related differences in processing. “I hadn’t considered looking in the ageing literature for information about this before, but now I will,” she says.

Yet despite her enthusiasm for the service, Fairs still tends to find new papers through alerts from Google Scholar, the dominant tool in the field; it’s easier to go “down the rabbit hole”, she explains, following a chain of papers that cite each other.

### Click to recommend

Google Scholar recommends papers depending on which articles users have authored and list in their profiles. The algorithm isn’t public, but the company says that the recommendations are based on “the topics of your articles, the places where you publish, the authors you work with and cite, the authors that work in the same area as you and the citation graph”. Users can manually set up extra e-mail alerts based on keyword searches or particular authors.

Aaron Tay, a librarian at Singapore Management University who studies academic search tools, gets literature recommendations from both Twitter and Google Scholar, and finds that the latter often highlights the same articles as his human colleagues, albeit a few days later. Google Scholar “is almost always on target”, he says.

Besides published articles, Google Scholar

might also pick up preprints as well as “low-quality theses and dissertations”, Tay says. Even so, “you get some gems you might not have seen”, he says. (Scopus, a competing literature database maintained by the Amsterdam-based publisher Elsevier, began incorporating preprints earlier this year, a spokesperson says. But it does not index theses and dissertations. “There will be titles that do not meet the Scopus standards but are covered by Google Scholar,” he says.)

Google Scholar does not disclose the size of its database, but it is widely acknowledged to be the biggest corpus in existence, with close to 400 million articles by one estimate (M. Guseinbauer *Scientometrics* **118**, 177–214; 2019). Open Knowledge Maps, meanwhile, is built on top of the open-source Bielefeld Academic Search Engine, which boasts more than 270 million documents, including preprints, and is curated to remove spam.

Connected Papers uses the publicly availa-

**“You do not have to sit on the hose of papers and look at every paper that comes out for fear of missing it.”**

ble corpus compiled by Semantic Scholar – a tool set up in 2015 by the Allen Institute for Artificial Intelligence in Seattle, Washington – amounting to around 200 million articles, including preprints. Smolyansky acknowledges this size discrepancy means that “very rarely” Google Scholar will find “some niche 1970s paper” that Semantic Scholar does not.

Semantic Scholar’s alert system, called an adaptive research feed, builds a list of recommended papers that users can train by liking or disliking the articles they see. To decide which papers are similar to those, it uses a machine-learning model trained on mutual citations, and on which articles Semantic Scholar users have viewed sequentially. It counts some 8 million monthly users.

### No more FOMO

Feedly, launched in 2008, also uses upvotes and downvotes to learn which new academic research is most relevant to the user, and benefits from an AI assistant that can be trained on specific keywords or topics. But Feedly isn’t aimed specifically at researchers – it aims to be an all-encompassing dashboard to monitor news, RSS feeds (which provide a way of alerting users to new content on websites), the online forum Reddit, Twitter and podcasts. A free version is available, but extra features, such as the ability to follow more than 100 sources and hide adverts, cost US\$6 or more a month (unlike most of the other tools mentioned here, which are entirely free; another paid option is ResearchGate +Plus, which boosts users’ visibility and

offers advanced statistics).

ResearchRabbit, which fully launched in August 2021, describes itself as “Spotify for papers”. Users get started by saving relevant papers to a collection. With each added paper, ResearchRabbit updates its list of recommended articles, mirroring how the music-streaming platform makes recommendations based on the songs users add to their playlists. The company behind it, based in Seattle, Washington, hasn’t revealed exactly how it assesses relevance, although it says it focuses on precise recommendations rather than floods of alerts. “We only want to send the most relevant papers to our users,” says chief executive Michael Ma.

Amber Brown Ruiz, a special-education and disability-policy doctoral student at Virginia Commonwealth University in Richmond, finds ResearchRabbit alerts to be more personalized than Google Scholar, which sometimes feeds her papers that are superficially similar to her own work but turn out to be far outside her discipline.

Ruiz also uses Connected Papers to find new articles. She finds it to be less automated than Google Scholar, which sends fresh papers by e-mail, “but you can manually go in and figure out which articles are the newest”, she says.

What all these tools have in common is that they use some sort of artificial intelligence to craft their recommendations. But some scholars enjoy the human touch, valuing recommendations from colleagues and contacts on Twitter, for example. ResearchGate, the long-standing platform that brands itself as a kind of social network for scientists, says it offers the best of both worlds (ResearchGate is in a content-sharing partnership with Springer Nature, which publishes *Nature*).

Founded in 2008, ResearchGate both e-mails recommendations of papers and serves them up through a rolling feed when users are logged in. (Users can also see a chronological newsfeed of papers posted by their ResearchGate contacts.) Although it doesn’t make its algorithm public, it uses information about a user’s publications and which publications they have viewed on the platform to understand their interests. It then calculates related articles on the basis of shared citations and extracted topics and keywords. ResearchGate currently includes some 149 million publication pages and has 20 million users.

“The secret sauce of ResearchGate is the combination of an active social network and a huge research graph,” says Joseph Debruin, ResearchGate’s director of product management, who is based in Los Angeles, California.

Five years after realizing he was drowning in new papers, Smolyansky is finally able to shake off that scientific ‘fear of missing out’. “You do not have to have that FOMO feeling,” he says.

**David Matthews** is a freelance writer based in Berlin.

**Clarification**

The text of this Technology feature has been modified to clarify that ResearchGate and SpringerNature are in a content-sharing partnership.