Millions of research papers are published in a year. How do scientists keep up?

April 27 2022, by Eva Botkin-Kowacki

Alessia Iancarelli, a doctoral student in Northeastern's Affective and Brain Sciences lab. Credit: Alyssa Stone/Northeastern University

If you want to be a scientist, you're going to have to do a lot of reading.

Science is an endeavor focused on building and sharing knowledge. Researchers publish papers detailing their discoveries, breakthroughs, and innovations in order to share those revelations with colleagues. And there are millions of scientific papers each year.

Keeping up with the latest developments in their field is a challenge for researchers at all points of their careers, but it especially affects early-career scientists, as they also have to read the many papers that represent the foundation of their field.

"It's impossible to read everything. Absolutely impossible," Ajay Satpute, director of the Affective and Brain Science Lab and an assistant professor of psychology at Northeastern. "And if you don't know everything that has happened in the field, there's a real chance of reinventing the wheel over and over and over again." The challenge, he says, is to figure out how to train the next generation of scientists economically, balancing the need to read all the seminal papers with training them as researchers in their own right.

That task is only getting more difficult, says Alessia Iancarelli, a Ph.D student studying affective and social psychology in Satpute's lab. "The volume of published literature just keeps increasing," she says. "How are scientists able to develop their scholarship in a field given this huge amount of literature?" They have to pick and choose what to read.

But common approaches to that prioritization, Iancarelli says, can incorporate biases and leave out crucial corners of the field. So Iancarelli, Satpute and colleagues developed a machine learning approach to find a better—and less biased—way to make a reading list. Their results, which were published last week in the journal *PLOS One*, also help reduce gender bias.

"There really is a problem about how we develop scholarship," Satpute

says. Right now, scientists will often use a search tool like Google Scholar on a topic and start from there, he says. "Or, if you're lucky, you'll get a wonderful instructor and have a great syllabus. But that's going to be basically the field through that person's eyes. And so I think that this really fills a niche that might help create balance and cross-disciplinary scholarship without necessarily having access to a wonderful instructor, because not everyone gets that."

The problem with something like Google Scholar, Iancarelli explains, is that it will give you the most popular papers in a field, measured by how many other papers have cited them. If there are subsets of that field that aren't as popular but are still relevant, the important papers on those topics might get missed with such a search.

Take, for example, the topic of aggression (which is the subject the researchers focused on to develop their algorithm). Media and video games are a particularly hot topic in aggression research, Iancarelli says, and therefore there are a lot more papers on that subset of the field than on other topics, such as the role of testosterone, and social aggression.

So Iancarelli decided to group papers on the topic of aggression into communities. Using citation network analysis, she identified 15 research communities on aggression. Rather than looking at the raw number of times a paper has been cited in another research paper, the algorithm determines a community of papers that tend to cite each other or the same core set of papers. The largest communities it revealed were media and video games, stress, traits and aggression, rumination and displaced aggression, the role of testosterone, and social aggression. But there were also some surprises, such as a smaller community of research papers focused on aggression and horses.

"If you use community detection, then you get this really rich, granular look at the aggression field," Satpute says. "You have sort of a bird's-eye-

view of the entire field rather than [it appearing that] the field of aggression is basically media, video games, and violence."

In addition to diversifying the topics featured by using this community approach, the researchers also found that the percentage of articles with women first authors dubbed influential by the algorithm doubled in comparison to when they focused only on total citation counts. (Iancarelli adds there might be some biases baked into that result, as the team couldn't ask the authors directly about their gender identity and instead had to rely on assumptions based on the author's name, picture, and any pronouns used to refer to them.)

The team has released the code behind this algorithm so that others can use it and replicate their citation network analysis approach in other fields of research.

For Iancarelli, there's another motivation: "I would love to use this work to create a syllabus and teach my own course on human aggression. I would really love to base the syllabus on the most relevant papers from each different community to give a true general view of the human aggression field."

  More information: Alessia Iancarelli et al, Using citation network analysis to enhance scholarship in psychological science: A case study of the human aggression literature, *PLOS ONE* (2022). DOI: 10.1371/journal.pone.0266513

Provided by Northeastern University