This repository   Search   Pull requests   Issues   Gist

jtleek / **readingpapers**

⊙ Watch ▾   5   ★ Star   82   ⑂ Fork   11

‹› Code   ⊙ Issues **0**   ⑂ Pull requests **0**   ▤ Wiki   ⟋ Pulse   ᵢₗᵢ Graphs

A guide to reading scientific papers

⊙ 1 commit	⑂ 1 branch	⬚ 0 releases	1 contributor

Branch: **master** ▾   New pull request   New file   Find file   HTTPS ▾   https://github.com/jtleek   Download ZIP

jtleek first commit   Latest commit `69775c1` 6 days ago

▤ README.md   first commit   6 days ago

▤ **README.md**

Reading academic papers

I have written a few guides for people in academics including:

- How to write your first paper
- How to review a paper
- How to share data
- How to write an R package

The other day on Twitter Amelia McNamara asked for one in a similar vein on reading papers for students. So this is my attempt to do that and I'd love contributions from other folks too.

Why should you read papers?

The academic paper is still the primary way of distributing new knowledge to the world. There are other ways too, with code, or blogs, or twitter. But academic papers are still the gold standard and where the vast majority of new scientific discoveries are reported.

What should you read?

Well this depends a lot on what you are interested in. There are a few broad categories of journals depending on what you are looking for.

- **Science magazines**: Journals like Nature, PNAS, and Science publish papers that are supposed to be "breakthroughs" of interest to a "general audience". This means that the papers tend to be written at a slightly less technical level and so are often more readible to people outside of a particular field. It also means that a lot of the most important papers get published in these journals. The flip side is that people often stretch hard in the interpretation of their data to make it seem like a "breakthrough" and get it into one of these journals. This means that the rate that papers are retracted is also very high in these journals. So read what you see there with a healthy grain of salt.

- **Health magazines**: In the health sciences there is a similar set of journals like the New England Journal of Medicine and the Journal of the American Medical Association. They have the same benefits and caveats as the Science magazines, but with a more health flavored bent.

- **Mega journals**: Journals like PLoS One and Peerj also publish in a wide range of areas and a ton of papers. The review criteria here is that it must be "correct" but not necessarily a "breathrough". So the heterogeneity in the papers is high. If a paper seems too good to be true, again it is worth taking with a grain of salt.

- **Field-specific Scientific journals**: Most scientific journals are not megajournals or magazines. These journals tend to be very field specific and tend to be much heavier on the details. This is where most science is published. The papers tend to be less focused on "breakthroughs" but are also less consistently risky to trust in these journals. In my area the journals might be something like Biostatistics or Biometrics.

- **Conference papers**: In some fields, like computer science, people tend to publish in short, peer reviewed conference papers. These papers tend to be quite technical - conferences like NIPS publish similar papers to very technical journals in other fields. Conference papers tend to be lighter on the detail and tend not to come with software/code so they can be a little harder to read and a little harder to use, but they are often talking about the very latest, coolest ideas in their subfield.

How to find what to read

The best places to find published academic papers are:

- Journal websites are a good place to start. Here is a list of journals.
- You can also read papers in biomedical sciences on aggregator sites like Pubmed Central

One problem with journal websites in particular is that many of the papers are behind a paywall - you have to pay to read them (see the next section). Increasingly you can find the latest papers on soemthing called a pre-print server. These papers aren't peer reviewed yet, but a large fraction of them ultimately end up in peer-reviewed journals. The nice thing about these papers is that they are frequently the latest research and free to read. Two good preprint servers are bioRxiv and arXiv.

Open access and #icanhazpdf

One thing that is super frustrating if you aren't at a university or research insitute is that many papers you might want to read cost money. They cost money because journals are what's called closed-access and they depend on making their money from readers/subscribers. In general papers will be free in:

- Open access journals (journals that make their money from authors instead of readers) like PLoS journals, Peerj, etc.
- Preprint servers like bioRxiv and arXiv
- Aggregators like Pubmed Central from funders that require papers to be free
- On authors websites where sometimes people post a free version.

If you run into a paper that costs money the first thing to do is check and see if the authors have published a pre-print and then check their website. If you still don't have any luck you could email the authors directly to ask for a copy of their paper (they are usually happy to oblige).

A more modern approach that has sprung up is something called *#icanhazpdf* which is a way you can crowdsource a pdf of a paper you can't read. If you have a twitter account, post a link to the paper, the hashtag *#icanhazpdf* and your email address and often someone will be willing to find and send you a copy of the paper. When you have the copy, delete your tweet, as this approach is technically a violation and could get you in trouble. Mostly journals won't care if you don't do this over and over with tons of papers, but be warned that journals can be nasty/lawyer up when their interests are being threatened.

How much should you read?

Academic papers come out all the time. Thousands are published every year, including hundreds in any given specific area. Unless you devote yourself full time to reading academic papers you won't be able to keep up with them all. I believe in the idea that you should read papers that you find interesting. Science is awesome and you shouldn't waste your time on the

boring parts if you can avoid it.

In general there are two main ways to find papers that I like. The way I used to do it was set up an aggregator with the RSS feeds from journals that I like, then I use the following (approximate) rates of reading parts of papers.

- 100% - read the title
- 20-50% - read the abstract
- 5-10% - look at the figures/captions
- 1-3% - read the whole paper

The new way that I do it is follow bioRxiv and a bunch of other people who have similar interests on Twitter. I use the above percentages for papers tweeted from aggregators and if I see a paper tweeted by 2-3 people I trust I usually end up reading that paper.

Reading a paper - the abstract/title

Different people will have different strategies. First I read the title and the abstract to get a sense for (a) why the paper is interesting according to the authors and (b) what are the main results in the paper. I do this to see if I think the paper is worth spending the time to read any deeper. I don't judge the quality at all from these components, just whether the paper is interesting or not.

Reading a paper - the figures

If I think the paper is interesting based on the title/abstract then the next thing I do is look at the figures and figure captions. As I've mentioned in my guide to writing papers the figures should tell a coherent story and should have figure captions that explain what is going on.

Hopefully the papers you are reading have figures that are this easy to read. I'm usually looking for the "story" that the authors are trying to tell. In the case of statistical or computational papers I'm also looking for comparisons to previous approaches and how this method stacks up.

Reading papers - the introduction

I usually skip the introduction. This is often an extended version of the abstract and often contains more opinion than fact about how awesome a particular result is.

The one exception I have to this rule is if I don't know the scientific area very well. Then I read the part of the introduction that reviews previous work in the area and if I don't understand something, I go chase down the references from the introduction and read through those to "get up to speed".

Reading papers - the methods/supplemental material

If I decide to read a paper carefully I spend the majority of my time reading the methods and supplemental material. This is where most of the real "science" is. It tells you how they did the experiments, how they analyzed the data, and how they support their conclusions. I'm looking for a few things when I read the methods section at a high level including:

- Do they explain clearly exactly which data they collected?
- Do they explain clearly exactly what analysis they performed on those data?
- Do they point to where I can get the data and code so I can verify these things?
- Do they explain every step in a process or skip over steps and reference previous papers?

Unfortunately, after that you sort of have to know the area to judge more critically whether the things they are doing are good or not. This comes with practice or with expertise in an area and can't be summarized very easily into succinct guidelines.

Reading papers - the results

I find that if the authors have done their job and made their figures tell the story and clear, then I usually spend less time reading the results section. The key results are usually in the figures, but I still glance over this section to see if there is any claim/idea that I missed from reading the figures. In general, I compare this section very carefully to the methods to make sure that the results seem well justified compared to what they say they did in the methods section.

Reading papers - Conclusions

Just like with the introduction, I often skip the conclusions. It is usually just a recap of what happened in the rest of the paper with a bit of guess work as to how the results might fit into the broader scheme.

Hype

One thing to keep in mind is that science is very often slow, steady, and incremental. But there is a lot of pressure on scientists to come up with "breakthroughs" (sometimes called the "i got a big one here" fallacy). When reading a paper if the authors claim they have cured cancer, discovered life on mars, or unified relativity and quantum theory then you should assume that they are full of it unless conclusively demonstrated otherwise.

Explain it to someone else

Reading academic papers can be a great way to catch up on knowledge. But in general I don't feel like I understand what is going on in a paper until I can explain the paper to someone else. So I try to discuss papers I think are really important with other people. The best way to do this is in a journal club or some other forum where you can put up figures from the paper and try to explain what is going on yourself.

Find out if others have read it

A lot of the papers I find interesting other people also find interesting. One nice way to learn a little more about a piece of scientific research is to see if it has been discussed on blogs. One thing to keep in mind is that blogs often have an agenda, so you should read the posts with a heavy dose of skepticism as well. Still, they can provide useful perspective on papers you've read.