TECHNISCHE UNIVERSITÄT DRESDEN

ACADEMIC SKILLS IN COMPUTER SCIENCE

Lecture 3: Judging Research Quality

Markus Krötzsch

Knowledge-Based Systems

TU Dresden, 16th April 2019

Goals for today

Learning goals of this lecture:

(1) Understand the strengths and weaknesses of peer review

(2) Get to know heuristics for judging the quality of research papers

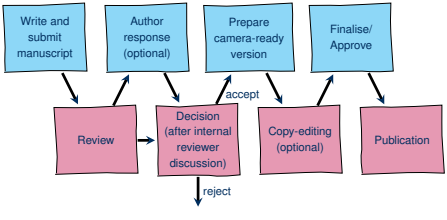
(3) Find out how people have tried to evaluate research publications

(4) Learn techniques to effectively search and access research publications

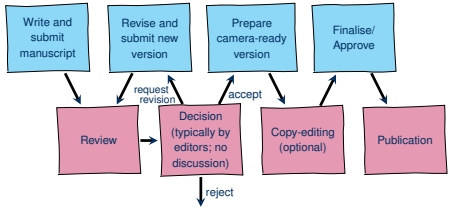
Review: Review

The most widely used method of quality control across all of academia

Conference/workshop reviewing:



Journal reviewing:



- Reviewers are volunteer experts, selected by the organisers/editors of a publication
- Should be competent, experienced, thorough, fair, and without conflicting interests
- Usually at least three reviews are gathered to make a decision

Peer review: strengths and weaknesses

Strengths

- Experts evaluate research close to their own area
- Some amount of social control & power balance (chair/editor/reviewers control each other)
- Scales well when a research area grows or shrinks

Peer review: strengths and weaknesses

Strengths

- Experts evaluate research close to their own area
- Some amount of social control & power balance (chair/editor/reviewers control each other)
- Scales well when a research area grows or shrinks

Weaknesses

- Reviewers might be inappropriately biased (favouring topics, disliking authors, etc.)
- Reviews can be of unsatisfactory quality (careless, uninformed, brief, etc.)
- Reviewers can exploit their position (e.g., by asking authors to cite their own works, or by stealing content from papers they reject)
- Significant effort (time spent by experts reviewing)

Peer review: strengths and weaknesses

Strengths

- Experts evaluate research close to their own area
- Some amount of social control & power balance (chair/editor/reviewers control each other)
- Scales well when a research area grows or shrinks

Weaknesses

- Reviewers might be inappropriately biased (favouring topics, disliking authors, etc.)
- Reviews can be of unsatisfactory quality (careless, uninformed, brief, etc.)
- Reviewers can exploit their position (e.g., by asking authors to cite their own works, or by stealing content from papers they reject)
- Significant effort (time spent by experts reviewing)

Both strength and weakness:

- Social norms and conventions on “how to do research properly” are reinforced (higher quality, more rigour, but also self-serving sub-communities, hostility to outsiders, and reluctance to challenge old ideas)

The perfect quality control?

Peer-review is not perfect:

- Errors in published papers are not uncommon, even for top venues
- Great papers are sometimes rejected

$\rightsquigarrow$ Even published papers must be read critically

The perfect quality control?

Peer-review is not perfect:

- Errors in published papers are not uncommon, even for top venues
- Great papers are sometimes rejected

$\rightsquigarrow$ Even published papers must be read critically

How to deal with errors:

- Authors can publish an error note (in the next issue of the journal, or at least on the own institute's page)
- In severe cases, accepted papers can be retracted by the publisher (usually in cases of fraud)
- Errors can be corrected in a future version (e.g., journal version of a conference paper)

$\rightsquigarrow$ Try to rely on most recent versions of works; prefer journal articles over conference publications

Informal publications

Today, technical reports can easily be published and accessed world-wide

- Nevertheless, they do not count as proper publications
- Research that only appeared in such reports is considered “unpublished”
- Books, M.Sc. and Ph.D. theses are treated similarly

Reasons for publishing technical reports:

- Make results known even before official publication
- Claim authorship of a new result before others do it
- Ensure public, archival availability of research (outside of publisher pay-walls)

Finding Literature:
(1) Judging Research Quality

The need to judge

Research publications can confuse and mislead the inexperienced:

- Anything can be published today
- Everybody tries to make their work look important, significant and useful
- Research moves fast: what was once reasonable can be invalid or irrelevant today

The need to judge

Research publications can confuse and mislead the inexperienced:

- Anything can be published today
- Everybody tries to make their work look important, significant and useful
- Research moves fast: what was once reasonable can be invalid or irrelevant today

Reading a research text always has two goals:

(1) **Learn:** understand what was done and why, get new insights, find out which ideas/sources/people are important in this field

(2) **Judge:** asses the quality of the work, identify weaknesses and limitations, rate how credible/transferable/trustworthy it is

$\rightsquigarrow$ if you are only a learner, you are missing a point (and are easily fooled)

How to know a “good” research paper?

There is no recipe!

- Research is not governed by fixed rules
- Any text on the internet might contain valuable and correct information

$\rightsquigarrow$ quality needs to be judged individually, based on experience

How to know a “good” research paper?

There is no recipe!

- Research is not governed by fixed rules
- Any text on the internet might contain valuable and correct information

$\rightsquigarrow$ quality needs to be judged individually, based on experience

However, some questions can help to estimate quality before reading in detail:

1. **Where was it published?**
2. **Who are the authors?**
3. **When was it published?**
4. **Are other researchers referring to it?**
5. **What do others say about it?**
6. **How is it presented?**

$\rightsquigarrow$ practically useful prejudices (“heuristics”); can help us to focus on the most promising works, but cannot replace critical reading of the actual paper

Q1: Where was it published?

Rationale: The best works are submitted to and accepted by the journals and conferences with the highest reputation.

- A self-fulfilling prophecy: if everybody believes that a venue is “the best” then all great works will first be submitted there (and many will be accepted)
- Problem: people don’t agree on the ranking and it varies from topic to topic; there will always be exceptions (bad paper/top venue or top result/mediocre venue)

Q1: Where was it published?

Rationale: The best works are submitted to and accepted by the journals and conferences with the highest reputation.

- A self-fulfilling prophecy: if everybody believes that a venue is “the best” then all great works will first be submitted there (and many will be accepted)
- Problem: people don’t agree on the ranking and it varies from topic to topic; there will always be exceptions (bad paper/top venue or top result/mediocre venue)

Checklist:

- Was it even reviewed? (journals, conferences, and some workshops; not: technical reports, books, theses, . . .)
- Reputation of venue? (established, well-known journal/conference? one-of workshop?)
- Topicality of venue
  (Examples: “Nature” is top in physics, not in computer science, but still would be credible there;
  if a journal is called “Amino Acids” then it is probably not a credible source for research in computational complexity)
- Reputation of publisher?
  (publishers like ACM, Springer-Nature, and Elsevier are responsible for the quality of their conferences/journals; the IEEE stamp seems to be available to any conference; other publishers such as CEUR and arXiv.org do not claim strong quality checks)
- Beware of fraud! (predatory publishers often name journals similar to established ones)

It takes time to learn about the reputation of venues in a specific field.

Q2: Who wrote the paper?

Rationale: Leading researchers in a field tend to produce papers that are highly relevant and credible.

- Indeed, they would not be leading if they would not do good work.
- Problems: it is hard to know if a researcher is a “guarantee for quality” (and there are few of those); some well-known researchers manage large research groups and “co-author” more publications than one person could even read; it would be wrong to assume that unknown authors cannot produce good work; reputation of host institutions should not be confused with reputation of researchers

Q2: Who wrote the paper?

Rationale: Leading researchers in a field tend to produce papers that are highly relevant and credible.

- Indeed, they would not be leading if they would not do good work.
- Problems: it is hard to know if a researcher is a “guarantee for quality” (and there are few of those); some well-known researchers manage large research groups and “co-author” more publications than one person could even read; it would be wrong to assume that unknown authors cannot produce good work; reputation of host institutions should not be confused with reputation of researchers

Checklist:

- Who are the senior/well-known researchers among the authors?
- How relevant might their contribution be here? (single author vs. group of many authors)
- How much and at which venues do those researchers publish? (are they known for quantity or quality?)
- Is this work actually in the area that they are known for?

It takes time to learn about the reputation of researchers in a specific field.

Q3: When was it published?

Rationale: Recent research is more up-to-date and well-informed than older work.

- One should assume that the authors knew the older works and improved upon them.
- Problems: some old work is highly relevant and won't be superseded; authors may not know about relevant previous works (many insights are "rediscovered" by uninformed researchers, possibly accepted by equally uninformed reviewers)

Q3: When was it published?

Rationale: Recent research is more up-to-date and well-informed than older work.

- One should assume that the authors knew the older works and improved upon them.
- Problems: some old work is highly relevant and won't be superseded; authors may not know about relevant previous works (many insights are "rediscovered" by uninformed researchers, possibly accepted by equally uninformed reviewers)

Checklist:

- When was the research done?
  (also recall that publications are finalised before they can be published; journals may have significant delays between submission/acceptance/publication.)
- Beware of misleading dates
  (e.g., references to Web pages may report the time of access or time of last change; books can have many editions from different years)
- What is the best-before date of the results?
  (mathematical theorems stay true in eternity, but can become irrelevant if common definitions change; empirical results, especially those related to practical applications, systems and performance age quickly and become irrelevant after a few years)
- How active is the field of research? (the more research happens in an area, the faster it changes)

Publication time is easy to interpret even for beginners; just use some common sense!

Q4: Are other researchers referring to it?

Rationale: If a work is important and useful, then many other research works will be based on it.

- Problems: How to find out how many other works are based on it? Very recent works cannot yet be used on many places. How many works use a result strongly depends on how many works are studying an area (recently).

Q4: Are other researchers referring to it?

Rationale: If a work is important and useful, then many other research works will be based on it.

- Problems: How to find out how many other works are based on it? Very recent works cannot yet be used on many places. How many works use a result strongly depends on how many works are studying an area (recently).

Checklist:

- Is the work cited in a positive way by other high-quality works?

  (note that one can also cite works to say that they are wrong/outdated)
- How often is the work cited overall?
- Who cites the work?

Citations are one of the visible forms of community appreciation in research, but can be grossly misleading (see later)

Q5: What do others say about it?

Rationale: Colleagues can help you to assess the quality of a work.

- Asking experts is an extremely valuable short-cut to learning about research quality
- Problems: personal views not usually published on the web; access to competent colleagues might be limited, and the opinions you get might not be representative

Q5: What do others say about it?

Rationale: Colleagues can help you to assess the quality of a work.

- Asking experts is an extremely valuable short-cut to learning about research quality
- Problems: personal views not usually published on the web; access to competent colleagues might be limited, and the opinions you get might not be representative

Checklist:

- Has the work been recommended to you?
- What do your co-authors/supervisors/teachers/colleagues think about this paper/researcher/topic?
- Can you find useful discussions in other works?
  (Example: high-quality(!) survey papers include balanced discussions of several approaches; but be careful with discussions in papers that want to advocate their own solution)
- If no local expert is available: can you ask somebody by email?

Benefit from the expertise of seniors; don't be shy in asking

Q6: How is it presented?

Rationale: High quality research will also have an excellent presentation, with good English, a pleasant narrative, and a thoughtful layout.

- Top researchers often care a lot about how their (best) works are presented.
- If it looks like junk, it probably is junk.
- Problems: only a coarse criterion, mostly to filter out the worst works

Q6: How is it presented?

Rationale: High quality research will also have an excellent presentation, with good English, a pleasant narrative, and a thoughtful layout.

- Top researchers often care a lot about how their (best) works are presented.
- If it looks like junk, it probably is junk.
- Problems: only a coarse criterion, mostly to filter out the worst works

Checklist:

- Are spelling, grammar, and typesetting flawless?
- Does the paper look professional and well structured?
- Is the narrative logical? Are all necessary details explained?

  (beginners might find it hard to distinguish “advanced research on a complex subject” from “insufficiently explained bogus”)

Requires some experience, but basic common sense helps.

Measuring research quality?

Preferential attachment

Many proxies of research quality suffer from phenomena of preferential attachment.

Preferential attachment

Many proxies of research quality suffer from phenomena of preferential attachment.
In other words:

"The rich get richer."

- Journals and conferences perceived as top ranked will receive better submissions
- Well-known authors are more frequently cited and will be more favourably reviewed
- Highly cited papers are read and cited much more often

Preferential attachment

Many proxies of research quality suffer from phenomena of preferential attachment.
In other words:

"The rich get richer."

- Journals and conferences perceived as top ranked will receive better submissions
- Well-known authors are more frequently cited and will be more favourably reviewed
- Highly cited papers are read and cited much more often

This can be good: It is useful to have agreed-upon top venues (whatever they are)
But often it is bad: Concentration of funding and influence; barriers of entry for outsiders

$\rightsquigarrow$ Don't be blinded by superficialities when judging the works of others!

Citations, the new currency of research

There is a strong desire to quantify research quality:

- Governments and university administrations want to rate and rank their researchers
- Researchers want to find “the best” works to refer to
- Funding bodies and hiring committees want to see “objective” evidence

. . . and the easiest quantity to measure are citations.

Citations, the new currency of research

There is a strong desire to quantify research quality:

- Governments and university administrations want to rate and rank their researchers
- Researchers want to find “the best” works to refer to
- Funding bodies and hiring committees want to see “objective” evidence

. . . and the easiest quantity to measure are citations.

How to count citations?

- Several online services count citations per paper (Google Scholar [free], Crossref/IO4C [free & open], Scopus [subscription], . . .)
- The numbers differ significantly:
  - Which publications are considered?
  - What is the data quality?
  - Are self-citations distinguished from citations of other researchers?
- Mostly accessed online or via Web API, but open data becomes available (IO4C)

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

- Highly cited papers: find papers with most citations

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

- Highly cited papers: find papers with most citations
- Citation counts: rank researchers by how often they get cited

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

- Highly cited papers: find papers with most citations
- Citation counts: rank researchers by how often they get cited
- H-index: number n of publications with at least n citations (typically computed per author, but anything else that defines a set of publications can be used)

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

- Highly cited papers: find papers with most citations
- Citation counts: rank researchers by how often they get cited
- H-index: number n of publications with at least n citations (typically computed per author, but anything else that defines a set of publications can be used)
- Impact factor: sum of citations of publications in previous two years, divided by the number of such publications, i.e., average number of recent citations per recently published article (typically computed for journals or conferences)

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

- Highly cited papers: find papers with most citations
- Citation counts: rank researchers by how often they get cited
- H-index: number n of publications with at least n citations (typically computed per author, but anything else that defines a set of publications can be used)
- Impact factor: sum of citations of publications in previous two years, divided by the number of such publications, i.e., average number of recent citations per recently published article (typically computed for journals or conferences)
- k-year impact factor: same as impact factor but using another number of years

Bibliometrics

Raw citation data can be used in many ways, e.g., to find papers that cite another paper (e.g., using Google Scholar)

But one of their main uses is to evaluate publications/researchers/research venues:

- Highly cited papers: find papers with most citations
- Citation counts: rank researchers by how often they get cited
- H-index: number n of publications with at least n citations (typically computed per author, but anything else that defines a set of publications can be used)
- Impact factor: sum of citations of publications in previous two years, divided by the number of such publications, i.e., average number of recent citations per recently published article (typically computed for journals or conferences)
- k-year impact factor: same as impact factor but using another number of years
- $i10$-index: number of papers with at least 10 citations

The faults of bibliometrics

Bibliometric evaluations of research are heavily criticised, for several reasons:

- Based on unreliable, subjective data: different data collection methods lead to wildly different results (e.g., Google Scholar vs. Scopus vs. Thomson Reuters)
- Heavily biased towards hype topics: the more researchers in an area, the more citations (e.g., machine learning vs. programming languages)
- Heavily biased towards certain paper types: introductory articles, tutorials, surveys, and reports about software tools are more highly cited than works that answer specific research questions
- Not meaningful across research areas: one cannot compare biology journals with computer science conferences – but where is the actual boundary?
- Impact factor focussed on fast-paced research: areas where impact happens on (much) larger time scales penalised (e.g., math)
- H-index favours the old (“age index”): grows over time, even if no new research happens
- Bibliometrics are easy to game: co-author syndicates, citation cartels, metric-optimised publishing, self-citation and self-promotion (especially effective for boosting journal impact factors)

Thomson Reuter's Impact Factor

The specific case of the widely known journal impact factor has further deficiencies:

- It is calculated by the company Thomson Reuters from raw data that it not publicly accessible.
- It is heavily affected by the (undisclosed) choice of which sections of a journal are considered research articles. This choice is negotiated (in private) with journals.
- It is essentially an average of citations. But citations follow a heavily skewed (almost exponential) distribution where few most cited papers are responsible for almost all citations. It is not sound to describe such a distribution by an average.
- It is easily gamed by journal editors by publishing "research articles" that cite many recent papers from their own journal.
- It was never intended to be more than a help for comparing journals, but has been abused to evaluate researchers and research programmes.

$\rightsquigarrow$ leading research institutions and funding bodies have discouraged the use of this metric for research evaluation

The bibliometrics trap

Main problem:

- **Initial situation:** bibliometrics are an imperfect but acceptable way to measure good research
- **Situation after a few years:** bibliometrics, rather than good research, becomes the primary working goal of many “researchers”

Bibliometrics become what is known as a **perverse incentive**.

The bibliometrics trap

Main problem:

- **Initial situation:** bibliometrics are an imperfect but acceptable way to measure good research
- **Situation after a few years:** bibliometrics, rather than good research, becomes the primary working goal of many “researchers”

Bibliometrics become what is known as a **perverse incentive**.

What to do about it?

- Measure quality not quantity (rely on human experts)
  - UK Research Evaluation Framework (nation-wide quality evaluation)
  - Responsible hiring/promotion/funding committees consulting invited experts
- Discourage the use of misleading metrics (increase awareness)
  - Many funding agencies discourage a focus on metrics
- Don’t make lives depend on metrics (fear makes people cheat)
  - Permanent positions allow researchers focus on honest research

Finding Literature:
(2) Effective search and retrieval

Searching for research

Searching literature is much easier today than it ever was

- Unified search of every possible publication
- Search by content, citations, and publication time
- Almost no delay between author and reader
- Many texts easily accessible online

. . . all thanks to the world wide web and modern search engines

Basic search process:

(1) Find publications on the Web (coming next)

(2) Judge if they are relevant/trustworthy/useful (previous part)

Finding literature (1)

For direct text-based search, major search engine providers are usually best:

Example: Google Scholar is a useful service for searching research texts (rather than the whole web). Features:

- Full text search (like usual web search)
- See authors, venue, publication time, and citations with results
- Links to PDF when available
- Search in citing documents
- Specify publication time

Like web search in general, the service's results are neither complete nor correct; rankings can be manipulated

Finding literature (2)

Specialised databases may offer more finely configurable search features and higher data quality.

Example: DBLP is the most important literature database in computer science.

- Freely available, high quality data (but no citation data)
- Simple, clean interface, centred on author or venue

Example: Scopus is a literature database and search service by Elsevier.

- Advanced search by many criteria (author, title, host institution, venue, funding body, . . .)
- More careful selection of sources than Google scholar
- Display citing documents; possibly filtering self-citations

Unfortunately, it is not free and offers a less-than-ideal user experience.

Other search techniques

Text search is not the only option:

- Citation-based: given an interesting paper, look at its citations (backward search), and search for other papers that cite it (forward search) [Google, Scopus]
- Wikipedia-based: look at what Wikipedia cites (use to find classical key references to established fields and definitions; be careful with very recent and niche topics)
- Person-based: given an interesting paper, figure out what its authors have published recently [DBLP, Google, author's web page]
- Venue-based: browse recent editions of a relevant conference or journal [DBLP, Scopus]

$\rightsquigarrow$ combine techniques for best results; interleave reading/rating and further search

Accessing literature

Many research papers in computer science are readily accessible:

- Google Scholar links to known legal PDF versions on the Web
- Even many historic publications have been scanned and OCR'ed
- Many authors use their right to publish preprints on their sites
- Some articles are free for the general public (open access)
- Other articles are available only through university subscription: access them from the university network (eduroam or VPN suffice)
- If all fails, you can email the authors or ask friends in other places (maybe other universities have access)
- Nevertheless, some works might still remain inaccessible ...

Accessing literature

Legal note: The use of the online services Library Genesis (libgen) and Sci-Hub to gain immediate free access to research articles and textbooks can be illegal (whether or not the Tor network is used to avoid tracking). The sites' practice of granting access to tax-paid research without collecting millions in license fees jeopardises the business of major publishing houses, and undermines their efforts of offering some of their content to a wider public under specific circumstances. Therefore, you should not use these sites, even if studies suggest that many students and academics, including from top universities, are doing this as part of their regular work.[a]

[a] Bohannon, John (28 April 2016). "Who's downloading pirated papers? Everyone". Science. 352 (6285): 508–512. doi:10.1126/science.aaf5664

Summary

Publication is not a definite certificate of high quality – critical thinking is needed

Some heuristics help us guess the quality of research papers (Where? Who? When? Who used it? What do others say? How does it look?)

Bibliometrics are a popular way of misunderstanding research as something that can be captured by a numeric value

Finding computer science literature works best by combining several strategies using Google Scholar, DBLP, and possibly Scopus

It is easy to access almost any publication on the web

What's next?

- Reading & understanding academic papers
- Writing reports