

How to Do (Good) Research? Starting from Scientometrics...

Shujun <u>LI</u> (<u>李</u>树钧)

Director, Institute of Cyber Security for Society (iCSS)

Professor of Cyber Security, School of Computing University of Kent, UK

http://www.hooklee.com/





Scientometrics ABC

What is <u>scientometrics</u>?





Contents [hide]

1 Historical development

Related changes Special pages

Permanent link

Page information

2 Methods and findings

Scientometrics is a science!

University of Kent Institute of Cyber Security for Society (iCSS)

- Some key journals
 - Scientometrics (Springer)
 - Journal of the American Society for Information Science and Technology (ASIS&T + Wiley)
 - Journal of Informetrics (Elsevier)
 - <u>Research Evaluation</u> (Oxford University <u>Press</u>)
 - Online Information Review: The international journal of digital information research and use (Emerald)
 - Learned Publishing (ALPSP + SSP)





- It helps us (better) evaluate scientific achievements and impact of our own research and others'.
- It links to general rules scientific communities follow.
- It tells us how we are assessed by our peers, professional bodies, university managers, government officials, etc.
- It relates to your personal goal(s) and your future career in academia.
- It appears and is often at the core of every step of how to do (good) research. – "Good" needs evaluating!
- It reveals the personal observation that science has been a (special type of) business. – We need to sell ourselves!
- And there are more...

- Yes and no. This is not a presentation about my own research on scientometrics, so it is not very formal and will involve many things beyond scientometrics.
- I will talk about my personal view on how to do (good) research in general, focusing on disciplines and research areas I am more familiar with (Computer Science, EEE, Mathematics, Cyber Security, and some related interdisciplinary areas).
- The goal is not about to show you what I think or know, but to inspire you to develop your own ideas.
- So discuss and debate with your peers and me.



Not everything that can be counted counts, and not everything that counts can be counted.



More likely: William Bruce Cameron in his 1963 book <u>Informal Sociology: A Casual</u> <u>Introduction to Sociological</u> <u>Thinking</u>

https://quoteinvestigator.com/2010/05/26/everything-counts-einstein/



- <u>San Francisco Declaration on Research Assessment</u> (DORA)
 - "the need to **eliminate the use of journal-based metrics**, such as Journal Impact Factors, in funding, appointment, and promotion considerations;
 - the need to assess research **on its own merits** rather than on the basis of the journal in which the research is published; and
 - the need to capitalize on the opportunities provided by online publication (such as relaxing unnecessary limits on the number of words, figures, and references in articles, and exploring new indicators of significance and impact)."
 - I will come to journal-based metrics later.
- As of May 2022, over 2,500 universities worldwide are signers, including the University of Kent.



Steps of Doing (Good) Research

- 1. Select a (good) research area (or topic)
 - How to find a (good) research area (or topic)?
- 2. Literature review read papers of others (and other references if needed)
 - How to find related work and get fulltext?
 - How to know which papers are more important?
- 3. Make your hand dirty trial and error
 - How to make work plan and adjust it according to what you success and/or failure you've got?
 - How to get help from other researchers? (Reproducible research campaign...)



University of







- 4. Develop your own idea(s) and verify it/them
 - How to verify it/them?
- 5. Write a compuscript to summarise your findings
 - Which word processing software should I use?
 - How to avoid ethical pitfalls?
- 6. Submit your compuscript to a selected venue
 - Making my work online before it is published anywhere?
 - Where to submit?
 - How does the review process work?
 - How to avoid ethical pitfalls?
 - What to do if rejected?







How to do (good) research: steps

- 7. After your paper is accepted
 - Prepare presentations (if needed)
 - Publicise your work online (if not yet) or update your online preprint (if done before submission)
 - Make camera-ready copy (add a permanent URL if you can)
- 8. After your paper is published
 - Keep your data and manuscript!
 - Track citations and impact of your work
 - Fix errors in published edition of your work
 - Update your online preprint
 - Extend your work and submit to a new venue?









Step 1: Select a (good) research area (topic)





- Follow the hot topics or your interests (if not both)?
- Focus on hard open problems or easy ones?
- Solve theoretical problems or meet practical needs?
- Work alone (on a smaller project) or collaborate (on a larger project)?
- Look for missing links in the field?
- Link two (or more) fields that are separate?
- Meet the gap between two steps?
- Disprove/Improve existing work?
- Create new research paradigms?



Step 2: Literature review – read papers of others







- Read books, PhD theses, surveys and SoK papers
 - SoK = Systematization of Knowledge (a new type of papers in cyber security, see <u>here</u> and <u>here</u>)
 - Pay attention to SLR (systematic literature review) papers
- Read recent research papers
 - Especially "Related Work" section
- Keep a close eye on newly published papers
 - TOC of important journals (there are e-alerts)
 - Programs of important conferences and workshops
 - Web pages and social media accounts of researchers
 - Seminars organised by relevant people and organisations
 - Your own social networks!

How to find relevant papers?





- Use online scientific services regularly
 - Indexing services (often with citations for forward snowballing – finding new papers citing a paper of interest)
 - <u>Google Scholar</u>, <u>Elsevier</u>'s <u>Scopus</u>, <u>DBLP</u> (Computer Science only, no citation data), <u>Clarivate</u>' <u>Web of Science</u>
 - Preprint servers
 - <u>arXiv.org</u> (Computer Science and more), <u>IACR</u>'s <u>Cryptology</u>
 <u>ePrint Archive</u> (Cryptology), <u>SSRN</u> for (Social Sciences), ...
 - Focus more on papers that have been peer reviewed and those of established researchers!
 - <u>ResearchGate</u> and other researcher-facing platforms
 - Curated bibliographies on relevant research topics
 - An example: <u>Best Paper Awards in Computer Science</u>





- Use scientific publishers' fulltext databases your organisation has a subscription to.
 - Some examples: <u>IEEE Xplore</u>, <u>ACM Digital Library</u>, <u>Elsevier ScienceDirect</u>, <u>SpringerLink</u>, …
- Some journals are fully or partially open-access.
- Search in Google using the paper's title (and when necessary add the authors' names)
 - Preprints, institutional archival websites, authors' websites
- Check the authors' web sites and ask them directly
- Ask your friends and collaborators for help

How to identity good papers?





- Look at trusted authors



- Trusted authors = Researchers who you trust (often big names, but be aware of the <u>Matthew effect</u>).
- Look at prestigious venues
 - Prestigious venues tend to publish only (but not always) good papers. (Why? – Papers are more rigorously peer reviewed by more critical eyes!)
- Look at citations
 - Good papers tend to (but not always) be cited more.
- Look at reviews and comments
 - Good papers are supposed to be received positively.
 - Examples: OpenReview, AMS's Mathematical Reviews, .





- Look at abstracts
 - A well-written abstract can tell you if a paper is good.
 - If the abstract is not well written, then it is unlikely a good paper.
- Look at the results
 - Key contributions listed in the Introduction section, figures and tables in a paper can gives you more insights about the paper in a few minutes.
- Look for recommendations
 - People you trust can recommend good papers to you.
- Form your own judgment!

- Look at the affiliations CAMBRIDGE
 - Prestigious affiliations tend to (but not always) have good researchers.
- Look at previous publication records
 - Good researchers tend to have a good publication record.
 - Naturally, this does not apply to newcomers who made ground-breaking results. (I will give an example later.)
 - How to judge one's publication record? Look at how well his/her publications are cited and reviewed.
 - Quantitative scientometric indicators: <u>h-index</u>, <u>g-index</u>, ...
 - Do NOT forget the remarks of Einstein and DORA!







- Look at recognitions and awards
 - Good researchers tend to (but not always) be recognised and awarded by the community.
- Look at reviews and comments
 - Well-reviewed and -commented papers tell us something about their authors.
- Look for comments from people you know personally
 - People you trust can recommend good researchers to you.
 - I will give two well-known examples on the next two slides.
- Form your own judgment!





University of Kent



- *h*-index (Jorge E. Hirsch, *PANS* 2005)
 - A researcher's *h*-index is *h* if he has *h* papers each of which has been cited ≥*h* times but any of other papers has been cited only less than *h* times.
 - How to calculate one's *h*-index?
 - Google Scholar Citations
 - Publish or Perish (Google Scholar based)
 - <u>Elsevier Scopus</u> and <u>ResearchGate</u> (can calculate *h*-index by excluding self-citations)
 - Semantic Scholar
 - Microsoft Academic Search
 - <u>AMiner</u>







- m-index or m-quotient (Jorge E. Hirsch, PANS 2005)
 - h-index / number of publishing years
- g-index (Leo Egghe, Scientometrics 2006)
 - A researcher's g-index is the maximum value g such that his g most cited papers have been cited (together) by ≥g² times.
 - A few highly-cited papers can make a big difference.
- *h_l*-index (Batista et al., Scientometrics 2006)
 - h-index / average number of authors in *h*-core papers
- *h_m*-index (P. D. Batista et al., *Scientometrics* 2006)





- Owners/Sponsors of journals: learned societies, universities and university presses, commercial publishers, researchers or group of researchers, ... (maybe jointly)
- Rule of thumb: prestigious owners tend to produce good journals.
- Prestigious universities and their presses: <u>Oxford</u>
 <u>University Press</u>, <u>Cambridge University Press</u>, <u>MIT Press</u>,
 <u>Princeton University Press</u>, …
- Prestigious learned societies (CS, EE and applied math):
 <u>IEEE, ACM, SIAM, IFIP, IET, EURASIP, IS&T, SPIE, IAPR, IACR, AAAI, IMLS, IEA, Eurographics, BCS (*The Computer Journal*), IMA, ...
 </u>















- Large/Old scientific publishers (and their brands/imprints):
 - <u>Elsevier B.V.</u> based in Amsterdam, The Netherlands: Academic Press, <u>Morgan Kaufmann</u>, North-Holland, …
 - <u>Springer Nature</u> based in Berlin, Heidelberg, Germany: Kluwer Academic Publishers, <u>Birkhäuser Verlag</u>, …
 - John Wiley & Sons, Inc. based in
 Hoboken, New Jersey, USA: <u>Wiley-Blackwell</u>, <u>Wiley-VCH</u>, …
 - <u>World Scientific Publishing</u> based in Singapore: <u>Imperial College Press</u> (joint with <u>Imperial College</u>)
 - <u>Taylor & Francis Group</u> based in UK: <u>CRC Press, LLC</u>, <u>Marcel Dekker</u>, ...
 - Emerald Group Publishing Limited based in Bradford, UK



WILEY

the language of science











- Scientific publishing becomes an increasingly easier and profitable business, so many new publishing companies have been flourishing since 1980s.
- Some new publishers become more visible over the years, but their reputation is still to be built up:
 - <u>IOS Press</u> based in Amsterdam, The Netherlands (1987-): known for its <u>Journal of Computer Security</u>
 - Hindawi Publishing Corporation based in Cario, Egypt (1997-): known as the former publisher of <u>EURASIP</u> journals
 - MDPI based in Basel, Switzerland (1996-)
 - Inderscience Publishers based in Geneva, Switzerland (1979-)
 - IGI Global based in Hershey, Pennsylvania, USA (1988-)
 - Now Publishers Inc. based in Boston, MA, USA (2003?-)





- Attention: Nowadays anybody can create a journal or a publishing company very easily, so keep your eyes open.
- Prestigious journals should always have an impressive editorial board filled with established researchers.
- Good journals are normally covered by well-known scientific indexing services and have good scientometric indicators (e.g. impact factors), especially <u>Clarivate</u>'s <u>Journal Citation Reports (JCR)</u> and <u>SJR (Scimago Journal & Country Rank)</u>.
 - Many journals covered by JCR are actually problematic venues (you'll see two examples later).
 - JCR does not cover all good journals, e.g., many new journals including those from decent publishers such as IEEE and ACM won't be covered until they have published enough issues.





- What are impact factors (IFs)?
 - Wikipedia: "The impact factor (IF) or journal impact factor (JIF) of an academic journal is a scientometric index calculated by Clarivate that reflects the yearly mean number of citations of articles published in the last two years in a given journal, as indexed by Clarivate's Web of Science."
 - The Wikipedia definition refers to Clarivate's service, but the concept has been used by other similar services.
- Let us check some journals' impact factors
 - Selected IEEE journals
 - Selected ACM journals
 - Selected non-IEEE/ACM journals in Computer Science
 - Selected journals in other disciplines
 - Selected journals that publish only or mainly survey papers





- Impact factors can be very misleading.
- Citation patterns differ from field to field:
 - medicine > physics > chemistry > biology > engineering (CS, EE, ...) > mathematics (<u>Igor Podlubny Scientometrics</u> <u>2005</u>)
- Journals publishing more (or only) reviews tend to have a much higher impact factor.
- Impact factors can be manipulated.
 - You will see two example later.
- Impact factors treat every citation equally no matter what this citation is about.





- New scientometric indicators are needed.
 - Clarivate's JCR has now different variants of their IFs to address some known issues.
 - Category Normalized Citation Impact (CNCI) ⇒ Journal Citation Indicator (JCI): JCI=1.5 means 50% more citation impact than the average in that category of journals
 - SCImago Journal Rank (SJR) Indicator



SCImago Journal & Country Rank

- use data from Elsevier's <u>Scopus</u>
- It mimics what Google <u>PageRank</u> algorithm does on web pages: citations from highly cited journals are weighted more than those from less cited journals.
- It has been used by some universities and research organisations for scientometric purposes.
- *h*-index is also extended to journals!





- Let us check the same journals' JCIs, SJR indicators and *h*-index values
 - Selected IEEE journals
 - Selected ACM journals
 - Selected non-IEEE/ACM journals in Computer Science
 - Selected journals in other disciplines
 - Selected journals that publish only or mainly survey papers





- Scientometric indicators do not work well. Do we have manually compiled journal list with subjective assessment?
- Yes, <u>Australian Research Council (ARC)</u> did this for the <u>ERA (Excellence in</u> <u>Australian Research Council</u> <u>Research for Australia</u>) initiative since 2006 until 2010.
- All journals are classified into five ranks: A*, A, B, C and "Not ranked".
- Unfortunately, the subjective ranking is highly controversial in Australian academia. In May 2011 the ARC decided to stop ranking journals in future years.
- If you want to have a look at Computer Science journals in ARC 2010 list, let me know.





Australian Academy of Science, Science Academy welcomes decision to drop ERA journal rankings, media release, 30 May 2011



- "It has been very distressing to see some universities using publications in highly ranked journals as the basis for funding, promotions, and even staff appointments."
- "The ranking of a journal as A* does not mean every paper in it is first rate, and some very good papers may appear in smaller journals."
- "People whose work is very relevant to Australian issues" rather than internationally, and those in new fields or collaborating between several universities, have been particularly disadvantaged."





- Other subjective rankings still being widely used:
 - Computer Sciences
 - <u>CORE (Computing Research and Education Association of</u> <u>Australasia) journal rankings</u> (updated yearly)
 - <u>CFC (China Computer Federation) journal and conference</u> <u>rankings</u> (last updated in 2019)
 - Business, Management and Information Systems
 - <u>ACPHIS (Australian Council of Professors and Heads of</u> <u>Information Systems) Information Systems Journal Ranking</u>
 - <u>ABDC (Australian Business Deans Council) Journal Quality</u>
 <u>List</u>
 - <u>CABS (Chartered Association of Business Schools)</u> <u>Academic Journal Guide 2021</u>





- Let us check selected Computer Science journals in CORE and CFC rankings
 - Selected IEEE journals
 - Selected ACM journals
 - Selected non-IEEE/ACM journals in Computer Science
 - Selected journals that publish only or mainly survey papers
How to identify good journals?













- Neither scientometric indicators nor purely subjective ranking should be used rigidly.
- We have to look at other aspects
 - Editorial board (ATTENTION: this can be fake!)
 - Reputation of authors
 - Quality of papers you've read
 - Recommendations from other researchers
 - ...
- There is space for more scientometric research $\textcircled{\odot}$
 - Better scientometric indicators
 - Better approaches to journal comparison

Be aware of junk journals!

- I detect junk journals according to
 - A journal whose website is hosted by Google Sites.
 - Don't laugh. There are MANY such fake journals out there!
 - Many are created by Asians especially Chinese, Indians, Pakistanis, etc. (You can guess the reason behind – economic drive!)
 - A journal whose name is so broad but no serious researcher around has heard about it or ever published papers on it.
 - A journal without a REAL peer review process.
 - A journal published/managed by a (brand) NEW organization you have never heard about.
 - Junk journals may NOT look like junk!











- Why conferences?
 - In Computer Science, conferences are very important because many (if not most) papers are published in conference proceedings, not in journals.
 - For some areas of computer science, top conferences are (much) more prestigious than journals and their proceedings are read by (much) more people!
 - Cyber security (including cryptology) is one of such fields with a strong tradition of trying to publish in (top) conference proceedings as much as possible.
 - Other fields include: theoretical computer science, programming, software engineering, operating systems, algorithms, AI, data sciences, information visualisation, ...





- Why conferences?
 - Computer Science conferences are normally smaller (in terms of its program), and with lower acceptance rate (<30%).
 - Top Computer Science conferences often have very low acceptance rate (<20% or even <15%).
 - In other disciplines, conferences are often big (or huge, with >1000 papers and participants), and with relatively high acceptance rate (>40%).
 - In some non-Computer Science disciplines, conference papers are NOT counted as scientific achievements.
 - In those disciplines, journals are often more prestigious but conferences offer researchers a good platform to publish preliminary work and meet peers.





- Journals vs. Conferences
 - Donald Geman, <u>Ten Reasons Why Conference Papers</u> <u>Should Be Abolished</u>, November 2007
 - Moshe Y. Vardi, <u>Editorial: Conferences vs. Journals in</u> <u>Computing Research</u>, *Communications of the ACM*, Vol.
 52, No. 5, p. 5, May 2009





- An old joke tells of a driver, returning home from a party where he had one drink too many, who hears a warning over the radio about a car careening down the wrong side of the highway. "A car?" he wondered aloud, "There are lots of cars on the wrong side of the road!" <u>I am afraid that driver is us, the computing-</u> <u>research community</u>...
- Lance Fortnow, <u>Time for Computer Science to Grow Up</u>, *Communications of the ACM*, Vol. 52, No. 8, pp. 33-35, August 2009





- In 2008, an interesting event was held: <u>Workshop on</u> <u>Organizing Workshops, Conferences, and Symposia for</u> <u>Computer Systems (WOWCS'08)</u>, USENIX
- Many issues about conferences were discussed at WOWCS'08. While the main topics are about how to better organise the technical program committee and improve the review process, it does mention the possibility of converting some conferences to journals if they do not offer the community-building element.
 - A new trend of moving towards multiple deadlines and a journal track/mode has been happening since then, including in cyber security.
- We will look at peer review policies later.





- Subjective rankings
 - <u>Australian CORE CS conference rankings</u> (updated yearly)
 - <u>CFC (China Computer Federation) journal and conference rankings</u> (last updated in 2019)
 - Australian ARC ranked conferences list in 2010
 - Some subjective rankings maintained by individual researchers
- Objective (automated) rankings
 - Italian GII-GRIN-SCIE (GGS) Conference Rating
 - Brazilian Ministry of Education's rankings (Qualis)
 - Top venues (including conferences) @ Google Scholar
 - <u>Top Computer Science conferences @ research.com</u>
 - Top CS conferences @ Microsoft Academic Search
 - Top CS venues @ Aminer

- Some key factors when judging the quality of a conference/symposium/workshop:
 - Reputation of organiser(s)/sponsor(s): IEEE and its societies, ACM and its SIGs (special interest groups), IACR, USENIX, IFIP TCs (technical committees), ISOC, IW3C2, SPIE, IS&T, IAPR, IET, AAAI, NIPS, INNS, IMLS, EURASIP, EATCS, Eurographics, BCS, GI, ...
 - Reputation of the technical program committee
 - History of the conference: the longer, the better
 - Acceptance rate: the lower, the higher the quality
 - Industry involvement: the more, the higher the quality

- ...







- Conferences with O(100) or even O(1000) papers and O(100) (big hidden constant) to O(1000) participants
 - Most IEEE conferences and some ACM conferences
 - Have to be multi-track to hold all papers in a few days.
 - Oral presentations and posters may not be decided upon quality.
 - Often accept special sessions organized by a group of researchers.
- Conferences with O(10) papers and O(100) participants
 - Most ACM conferences and almost all security conferences
 - Often single-track so that all papers are attended by all.
 - Posters (or work in progress, ...) are treated differently.
- Conferences with O(10) papers and O(10) participants
 - This type of conferences are often organised as workshops.
 - They are often co-located with other larger conferences.





- Fake conferences = Conference with a scientific name but with a purpose to earn money for the organizers.
- Fake conferences = Conference with a scientific name but with a purpose to promote somebody's research profile.
 - Being the chair of a conference is something one can put into CV.
- Fake conferences = Conferences with a scientific name but without a REAL peer review process.
- Fake conferences can be a source of amusements!
 - MIT students wrote a program called <u>SCIgen</u> to produce nonsense papers automatically. (Find more fun at <u>SCIgen web site</u>...)
 - They got an SCIgen paper accepted to <u>WMSCI 2005</u>!
 - Other SCIgen successes: <u>Dr. Herbert Schlangemann</u>'s SCIgen papers were accepted to CSSE 2009 and EBISS 2009!

The story of a fake researcher, a fake conference, and a fake journal

- A fake researcher
 - [redacted]
- gave a plenary talk at a fake conference organised by a fake organisation [redacted]
 - [redacted]
- and talked about "[redacted]".
 - He said: "... [redacted] ..."
- He published a paper at a fake journal <u>Nature and Science</u>.
 - [redacted]









Step 2: Two Case Studies



- Please read the entry on the RationalWiki to find more (link in the heading).
- I have some personal summary of this case study. Feel free to contact me if you want more details.



- I don't want to disclose the identity of the second researcher because he is still a professor at a decent university.
- It is not because I am afraid of him, but I don't want to spend time and money on malicious legal cases against myself. (This happened with other researchers.)



Step 3: Make your hand dirty – trial and error





- How to make work plan and adjust it according to what you success and/or failure you've got?
 - Long-term and short-term work plans
 - Set a publication target so you can work towards it (which can be changed later if you finally miss it)
 - Regular meetings with supervisors/collaborators
 - Be prepared for major changes to your work plans
- How to get help from other researchers?
 - Write to the authors of papers you are studying
 - For papers following the <u>reproducible research campaign</u>, the source code is always available...
 - Divide your work with your collaborators (if possible)...

Don't forget some important things



- Relevant laws in relevant jurisdictions
 - Cyber crime related law
 - Data protection
 - Copyrights
 - Contract laws
 - ...
- Ethical issues
 - We will come to this later.
- EDI
 - Equality, diversity and inclusiveness
 - Why does this matter? Be aware of potential biases.



Step 4: Develop your own idea(s) and verify it/them



How to...



- How to get new ideas?
 - Read, think, ask, experiment, reflect, repeat, ...
 - Improve existing methods
 - Combine techniques from different fields (you may need collective intelligence)
 - Get as much help from your supervisor(s)/collaborator(s) as possible (they may be busy, but keep communicating with them and telling them what you think).
 - You need to have luck.
 - It takes time, so be patient (while you are working hard).
 - You may need to switch to a different topic if you are stuck and no hope to get there (consult your supervisors).



- How to verify/validate your idea?
 - Do some experiments and see if and how it works!
 - Do (simplified) feasibility tests first to save your time.
 - Use a proper dataset (you may need to built your own if nothing is available – highly likely if your problem is very new or very specific).
 - Compare with the state-of-the-art methods.
 - Discuss with your supervisor(s)/collaborator(s).
 - If possible, ask non-experts to see if your idea makes sense for them.
 - Imagine what your future reviewers will look at your idea and your results.

. . .



- If your research has (potential) ethical issues, you need to consider carefully how to address them **BEFORE** you start your actual work.
 - Involving human participants is a main (not the only) reason.
- Many research organisations have a research ethics committee (called IRB – institutional review board –in some countries e.g. USA).
 - It is often mandatory for you to go through a formal review process. (Not the case for some countries, e.g., in Germany and China.)
 - There may be a fast-track process for simpler research tasks, e.g., anonymous online surveys.
 - Going through such a formal process is not just about protecting others and the organisation, but also you (the researchers)!
 - For the University of Kent, see <u>here</u> for more guidelines.



- Many journals, conferences, editors and reviewers will desk reject your paper.
- If your paper is accepted, it can still be withdrawn later.
 - IEEE S&P'21 Program Committee Statement Regarding The "Hypocrite Commits" Paper
- If your paper is published, it can still be retracted later.
 - This has been happening more and more frequently!
 - Check <u>Retraction Watch</u>!
- It will ultimately damage you and your co-authors' reputation.
- You may mislead readers, other researchers and practitioners, and this can cost lives and <u>time in prison</u>!
- So **NEVER** forget about research ethics!



Step 5: Write a compuscript to summarise your findings



How to write your campuscript: What typesetting software to use?



- Use a professional software for important work
 - TeX/LaTeX is the standard for computer science and mathematics, but not so for electronic engineering and a few subfields of computer science (e.g. computer-human interface).
- Don't know TeX/LaTeX yet?

- $T_E X / I_A T_E X$
- It deserves learning from now on....
- A recommended starting point: LaTeX WikiBook
- Install any TeX distribution you will see many useful documents shipped with it.
- Practice and ask for help from Texicians you know.



- Very professional and beautiful math (as a seamless part of the compuscript).
 - TeX was originally created for beautiful math equations.
- Automatic reference and citation management (via BibTeX).
- Automatically positioned floating objects (figures, tables, anything else you want them to be floating).
- Many powerful vector graphics drawing packages (<u>PGF/TikZ</u>, <u>PSTricks</u>, <u>MetaPost</u>, etc.)
- TeX/LaTeX source files are textual files!
- Thousands of TeX/LaTeX packages!



- Overleaf will be your choice!
 - Online collaborative TeX/LaTeX editor + online storage
 - Many LaTeX document templates (e.g. IEEE, ACM, ...)

Try out Overleaf v2!	My Projects Shujun Li 🗸
	ENEFITS + TEMPLATES PRICING COMPANY + HELP
Collaborative Writing and Publishing	
Join 750,000+ authors enjoying the easiest way to create, collaborate and publish online	
Start writing now! CREATE A NEW PAPER	
Sverleaf Sverleaf Sverlead Aver Svere Large Bullenaus ? C Filp	
Overlagt Overlagt <th< th=""><th><text><text><section-header><section-header><section-header><section-header><section-header><section-header></section-header></section-header></section-header></section-header></section-header></section-header></text></text></th></th<>	<text><text><section-header><section-header><section-header><section-header><section-header><section-header></section-header></section-header></section-header></section-header></section-header></section-header></text></text>

How to write your campuscript: Be aware of copyright issues...

C



- You should be careful in using copyright protected materials without a prior permission.
 - Such as images, videos and texts.
- Use safer or self-made materials if you can.
- Researchers are not lawyers, so they are not very sensitive to this issue: <u>The Lenna Story</u>
 - Do you know who owns the copyright of the "standard" test image "Lenna" and the use of this image violates the copyright law?



- <u>Do you know the owner ever warned SPIE about the</u> <u>unauthorized use of this image?</u>
- Do you know what happened afterwards?

65

How to write your campuscript: Be aware of copyright issues...

- Researchers are not lawyers, so they are not very sensitive to this issue: <u>The Lenna Story</u>
 - Eileen Kent, VP of new media at Playboy:
 "We decided we should exploit this, because it is a phenomenon."
- While the Lenna case has a win-win end, you may not be that lucky if you have used another copyright protected image or video.
- So be aware of possible consequences.
 - Once your paper is published, you cannot change it!





How to avoid ethical pitfalls: Research integrity is important!



- Do not be a plagiarist!
 - Do not steal ideas from others and claim them as yours!
 - Do not copy ideas from others and republish them!
- Do not be a self-plagiarist!
 - Do not publish your own work at multiple venues without prior approval by ALL the people who are concerned (editors, co-authors, publishers, etc.).
- Do not forge your data!
 - A common practice is to store your research data for ≥ 10 years.
- Always cite others' work if you use it!
- Do not pretend you don't know some existing work you actually know (be honest)!



- Do not be a plagiarist (legal requirement)!
 - Do not copy materials verbatim from copyright protected publications!
- Always use quotation marks if you cite more than a few sentences verbatim from a copyright protected publication!
- Seek/Clarify permission from the copyright holder if you want to use more than a few sentences verbatim from a copyright protected publication!
- "Copyright law protects the expression of ideas. If does not protect the ideas themselves."
 - So change your language to avoid violating other's copyrights!
- Read copyright transfer forms and know what you can do and what you cannot!

How to avoid ethical pitfalls: Research integrity vs. copyrights

- University of Kent Institut Cyber S for Soci (ICSS)
- Seek advice from a lawyer if you are not sure what to do!
 - Well this is expensive if you go to a lawyer office, but universities often have free legal consultants you can ask for help.
- Read the following paper to know your responsibilities as an CS/EE author and why copyrights matter!
 - Harold S. Stone, <u>Copyrights and Author Responsibilities</u>, *Computer*, vol. 25, no. 12, pp. 46-51, IEEE Computer Society, December 1992
- Read codes/rules of ethics/conduct of your learned society!
 - <u>IEEE Code of Ethics</u>, <u>ACM Code of Ethics and Professional</u> <u>Conduct</u>, <u>BCS Code of Conduct</u>, <u>IET Rules of Conduct</u>, ...
- Know more about the common practice in your field!
- Attend a training session about professional/scientific ethics if you feel a lack of knowledge!



- Who should appear in the co-authors list?
 - Rule of thumb: **anybody who contributed scientifically!**
 - People who helped you improve your work scientifically.
 - People who helped you improve scientific quality of the manuscript.
 - People whose (unpublished) work becomes part of your work.
 - Caveat: different fields have different cultures and conventions!
 - In some field (e.g. archaeology), providing materials is considered as scientific contributions.
 - In some fields, contributions of each co-authors are explained.
 - In many fields, supervisors and project leaders are always coauthors no matter how much contributions (maybe very little) they actually made. (This is a grey zone.)
 - In some fields (e.g. humanity and literature), only solo work is considered as major achievements.



- Who should be the **first and corresponding author(s)**?
 - Different fields have different cultures and conventions.
 - Medicine: the last authors is the most important and normally the PI.
 - Math (cover part of cryptography community): every co-author is equally important and co-authors are often ordered alphabetically.
 - CS and EE: co-authors are ordered according to the amount their actual contributions. (This can be difficult to determine.)
- Mark co-authors with equivalent contributions?
 - This is acceptable in some fields (e.g. medicine and biology).
- Acknowledge people who deserve a "thank you".
 - Including reviewers who help improve the quality of your work!
- Acknowledge funding bodies
 - Do not ignore this to save space for "more scientific" part because you are often (not always) obliged to thank them in publications.



- The (only) principal writer is always the first author.
- The person who contributed most is always the first author no matter if he/she is the principal writer of the paper.
- The person who proposed the key idea is always the first author no matter how much labour/time he actually spent.
- Some supervisors prefer being the last author.
- When there are more than one supervisor, an agreement is always needed among all co-authors.
- The PI of the project making the research possible is always the corresponding author.
- The co-author with the most stable contact information (i.e., a non-student) is always the corresponding author.



- ALL co-authors should be informed about the decision on the co-authors list and they all agree.
- The agreement should be secured before the first manuscript was submitted, NOT until it is accepted.
- In no case the co-authors list should be changed without a consensus among all co-authors.
- If a co-author does not agree to send your work for review, it is ethically wrong (100%!) to just remove their name and then submit!
- Be sensitive to this issue and be prepared of possible disputes among co-authors!
Case Study 1: <u>Dr Pál Schmitt</u>

- A successful fencer, winning two gold medals in 1968 and 1972 Olympics
- Obtained his doctoral (<u>dr. univ.</u>) degree in 1992 for his work on "An analysis of the programme of the modern Olympics"
- Hungarian President (2010 2012)
- Jan. 2012: plagiarism in his thesis (ca. 200 out of 215 pages) and thesis committee conflicts of interest exposed
- 29 March 2012: doctoral degree revoke by the <u>Semmelweis University</u>
- 2 April 2012: resigned from his presidency





- Doctorate in law from <u>Universität Bayreuth</u> with highest distinction "summa cum laude" (2007)
- Federal Minister for Economics and Technology of Germany (2009)
- Federal Minister of Defence of Germany (2009-2011)
- The most popular politician in Germany (2011)
- Feb. 2011: evidence about plagiarism appeared
 ⇒ Universität Bayreuth revoked his doctoral
 degree ⇒ <u>Open letter to German Chancellor
 Angela Merkel signed by 63,713 individuals</u>.
- March 2011: resigned from ALL his political posts (so politically died in Germany)





Case Study 3: Dr Jan Hendrik Schön

- Doctorate from Universität Konstanz (1997)
- 7 papers in Nature + 9 in Science (1998-2001)
- One paper every 8 days (2001)
- Otto-Klung-Weberbank Prize for Physics and Braunschweig Prize (2001)
- Outstanding Young Investigator Award of the Materials Research Society (2002)
- 2002: data forgery in almost all his work detected
- "Biggest fraud in physics in the past 50 years"
- June 2004: doctoral degree revoked by Universität Konstanz (Schön appealed and later sued the university but failed in September 2011)





Case Study 4: <u>Prof Woo-suk Hwang</u> (<u>黄禹锡教授</u>)

- The "5th" cloned cow in the world (1999)
- "1st" success of human somatic cell cloning (Science 2004)
- "Another" success of human somatic cell cloning (Science 2005)
- 1st cloned dog in the world (*Nature* 2005)
- 1st Supreme Scientist of South Korea (2005)
- 21 Nov. 2005: violation of ethics defined in the <u>Declaration of Helsinki</u> exposed
- 10 Jan. 2006: all Hwang's stem cells were found fakes by his university (so were both Science papers were)
- 20 March 2006: dismissed by his university
- 21 March 2006: Supreme Scientist title revoked
- 12 May 2006: charges of fraud, embezzlement and breach of bioethics law ⇒ suspended 18-month prison term (2010)





Case Study 5: Dr Haruko Obokata (小保方晴子博士)

- 2011: PhD in Engineering, Waseda University, Japan
- 2013: Two research papers published at the same issue of *Nature*.
- 2013: Head, Lab for Cellular Reprogramming, Riken Center for Developmental Biology, Japan
- March 2014: Riken Center for Developmental Biology confirmed data manipulation allegations against Obokata's *Nature* papers.
- July 2014: Both Nature papers were retracted.
- August 2014: Obokata's co-author Yoshiki Sasai committed suicide.
- December 2014: Obokata resigned.
- November 2015: Obokata's PhD degree was revoked (due to similar but not the same data manipulation allegations).
- January 2016: Obokata published Ano hi (あの日 That Day), in which she blamed another co-author Teruhiko Wakayama for the misconduct.
- March 2018: Obokata published <u>小保方晴子日記</u> (Haruko Obokata's Diary).







- Wikipedia has <u>an incomplete list of famous scientific</u> <u>misconduct cases</u>.
 - "A 2009 systematic review and meta-analysis of survey data found that about 2% of scientists admitted to falsifying, fabricating, or modifying data at least once." (Daniele Fanelli, PLOS ONE 2009)



Step 6: Submit your compuscript to a selected venue





- Making my work online before it is accepted/published anywhere?
 - This may protect your idea from being stolen by reviewers.
 (Yes, this did happen from time to time!)
 - There are many ways to publish your preprint publicly: arXiv, IACR Cryptology ePrint Archive, your institute's archival services (KAR for the University of Kent), media press (if your work interests them), ...
 - Some conferences have a double blind review policy, which may not allow you to publish your work before it gets accepted. Check the policy to clarify the rules!



- Where to submit?
 - Start from the best venues: you can get best comments, if you are lucky you get the highest reward, motivate you to set a higher bar of your own work, only at best venues you can meet best peers, ...
 - Try a relatively easy but still decent venue for your first publication only: you may need to build up confidence from a small success; ...
 - Set up a bottom line for your submissions: I will never send paper to a conference less known than ...
 - Send to conferences first and then extended editions to journals (this is a common practice in both CS and EE)



- Do not submit any work to more than one venue, unless it is allowed by all venues and you clearly inform the editors/TPCs of all venues.
- Do not submit any published work to a new venue without a significant amount of new materials.
- For many conferences, when you submit your paper you agree that you will attend the conference for presenting your work if it is accepted.
 - You may be blacklisted if you violate this rule later.
- You can always withdraw from one venue if you do want to switch to a different venue, but this is not a good practice...
 - Write to the editor/programme chair(s) for approval.
 - You need a good reason for this.



- Single-blind review
 - Papers are not anonymized, so the reviewers know the authors.
 - Reviews are anonymized, so the authors don't know the reviewers.
- Double-blind review
 - Both papers and reviews are anonymized.
- Triple-blind review
 - Double-blind review + No pre-prints or any public disclosure
- Rebuttal and revision
 - Authors are given a chance (or more) to rebut.
 - Adopted by almost all journals and some top conferences.
 - For journal submissions you will normally have a chance to further revise your manuscript, but not for conference submissions.
- On-site review
 - Post-conference proceedings of selected and revised papers



- Open reviews
 - Neither papers nor reviews is anonymised, so ...
 - Some journals had a try (e.g. *Nature* in 2006).
 - <u>OpenReview</u> (popular in AI venues) have reviews open and reviewers's IDs open to all.
- Invited papers
 - Some journals and conferences may invite established researchers to write (often survey) papers.
 - They are normally not peer reviewed or lightly reviewed.
- Shepherding
 - Some conferences offer conditional acceptances and allocate TPC members to help authors revise their submissions accordingly.
- Special issues (journal) and special sessions (conferences)
 - Different rules may be used by guest editors/session chairs.



- 1. A paper is submitted to the (online) submission system.
- 2. The EiC allocates the paper to an associate editor (or an editorial board member) whose expertise matches.
- 3. The associate editor accepts the allocation and sends invitations to potential reviewers.
- 4. Potential reviewers accept review invitations.
- 5. Reviewers return review reports.
- 6. The associate editor sends a recommendation to the EiC.
- 7. The EiC decides if the paper should be accepted or rejected or resubmitted after minor or major revision.
 - In some fields (mainly mathematics as far as I know), a consensus across the board must be reached.



- 1. A paper is submitted to the (online) submission system.
- 2. Immediately after the submission deadline, the TPC chair(s) allocate papers to TPC members.
 - Some conferences allow TPC members to bid first.
- 3. TPC members either review the allocated papers themselves or look for help from external reviewers.
- 4. TPC members & external reviewers return review reports.
- 5. TPC chair(s), the whole TPC or selected TPC members meet (physically or remotely) to select papers that will be accepted.
 - There may be multiple stages to gradually filter some papers out.
 - When there is a rebuttal phase, rebuttals from the authors will be used during the decision making process.



- Good papers can get rejected (again and again).
 - The following paper gives a list of many good examples. Juan Miguel Campanario, <u>Consolation for the Scientists:</u> <u>Sometimes it is Hard to Publish Papers that are Later Highly-Cited</u>, *Social Studies of Science*, vol. 23, pp. 342-362, 1993
- Bad papers can get accepted.
- There is randomness in the decision making process.
 - There is randomness about who will be your reviewers.
 - There is randomness about reviewers' recommendations.
 - There is randomness about final decision by the EiC/TPC.
 - For each conference, there is normally only a few papers that will be accepted without any controversy.
- What should I do? Improve, resubmit, be patient, ...



Step 7: After your paper is accepted



Things to do...



- Prepare presentations
 - For conferences you will always need to prepare something, either an oral presentation or a poster.
 - I always feel posters are difficult to make than oral presentations ©
- Make camera-ready copy (final edition)
 - Incorporate reviewers' comments
 - Add things missed from the submission
 - Add a permanent URL of your work so that your readers can get new materials (e.g. corrigendum) there
- Publicise your work online or update your online preprint
 - This can help increase citations of your work.
 - Make sure readers have the updated citation information.



Step 8: After your paper is published





- Keep your data and compuscript! (Why?)
- Track citations of your work. (Why? How?)
- Fix errors in published edition of your work. (Why?)
- Update your online preprint. (Why?)
- Extend your work and submit to a new venue?
 - This is allowed in some fields and by some learned societies/publishers.
 - For CS/EE, the normal rule is: add at least 30% new materials into a conference paper you may send the extended edition to a journal for possible publication.
 - Check the rules before start! They may differ!



- Some scientific services allow you to track new citations your paper have received.
 - You have been shown some of such services before.
 - Set up a Google Scholar Citations profile and keep it maintained!
- Some researchers can enforce citations ③
 - If your paper provides a database, you can set up a license so that people who use your database must cite your related paper(s).
 - If you are reviewers of a paper, you can "suggest" citing your relevant paper(s).
 - Is this ethical? It depends. (Why and when?)



Some useful links on Shujun's personal web site

Shujun's collection of links ©



- HTMS = How to Make Science

- Scientific Ethics; How to Do Good Research; Peer Review; How to Survive Academia; Story about Prestige Factor; More about Academia: Histories, Anecdotes, Fun
- Online SoK bibliography
- Scientific Services
- Scientific Journals
- <u>Scientific Conferences</u>
- Scientific Organizations
- <u>TeX/LaTeX Resource</u>
- Resources for researchers in the UK



Some sayings you may find useful











- Be honest, always. (不信不立,不诚不行。)
- Respect others, respect yourself. Do not treat others in ways that you would not like to be treated. (尊重他人,尊重自己。己所不欲,勿施于人。)
- Work hard; no pains, no gains. (业精于勤,而荒于嬉。不劳无获。)
- Be curious (and thus self-motivated); love what you are doing. (兴趣是成功的基石。)
- Be modest; everybody can be your teacher. (三人行 , 必有我师焉。择其善者而从之, 其不善者而改之。)



- Be patient; it is impossible to build Rome in a day. (无欲速,无见小利。欲速,则不达;见小利,则大事 不成。)
- Be realistic; don't do what you cannot do (or have no time to do). (人贵有自知之明。扬长避短。)
- Be ambitious and confident. I will be successful if I work hard. (不想做将军的士兵不是好士兵。)
- Start from the basics; pay attention to details. (故不 积跬步,无以至千里,不积小流,无以成江海。)
- Dubito, ergo cogito, ergo sum. (I doubt, therefore I think, therefore I am.) (我疑故我思,我思故我在。)



- Be open-minded; exposing yourself to more beyond your field may help in an unexpected way. (无专精则不能成,无涉猎则不能通也。)
- Nobody is perfect and there will always be (big or small) errors in your work. (金无足赤,人无完人。)
- Communication the human connection is the key to personal and career success. (沟通是成功之母。)
- One of the best ways to learn from others is to share with them. (分享是最好的学习。)