

The Art and Science of Empirical Computer Science

Week 4: The Science of Impact

Jimmy Lin
October 3, 2022

Computational Artifacts

Position A: Early-stage researchers should do the minimal in open-sourcing computational artifacts that arise from their work. Doing anything more than the community norm is a waste of time and effort that could be better spent writing more papers.

Position B: Early-stage researchers should actively promote the adoption of computational artifacts that arise from their work, for example, contributing to popular open-source libraries. Even if this requires a lot of time (e.g., refactoring code into a production-ready state), such efforts are worthwhile.

Let's debate!

“The Science of Science”, Part II

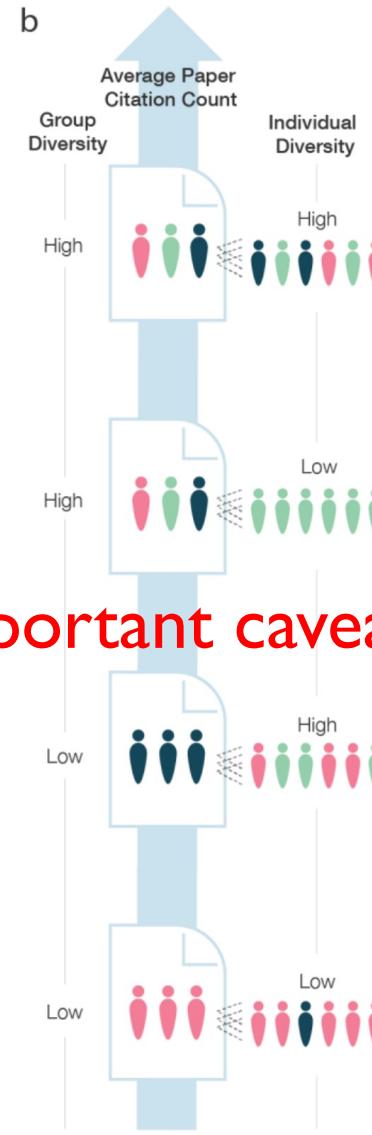
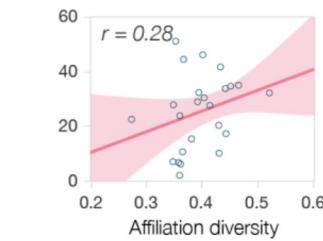
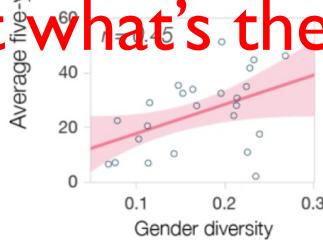
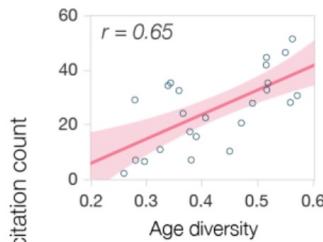
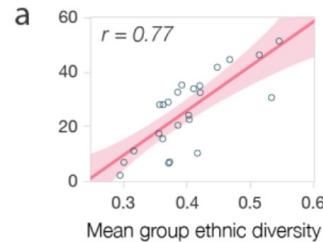
The Science of Collaboration

Chapter III: Team Assembly

Discussion Points

Models...

... and “reality”



But what's the important caveat?!

Figure 2.4.1 Team diversity and scientific impact. Panel a: an analysis of more than 1 million papers in 24 academic subfields (circles) shows that ethnic diversity correlates more strongly (r) with citation counts than do diversity in age, gender or affiliation. Panel b: Comparing team versus individual diversity reveals that diversity within the list of authors on a paper (team diversity) has a stronger effect on citation count than diversity in a researcher's network of collaborators (individual diversity). After Powell [58].

Simple (newcomer, incumbent) model

Incumbency parameter p : fraction of incumbents in a team.

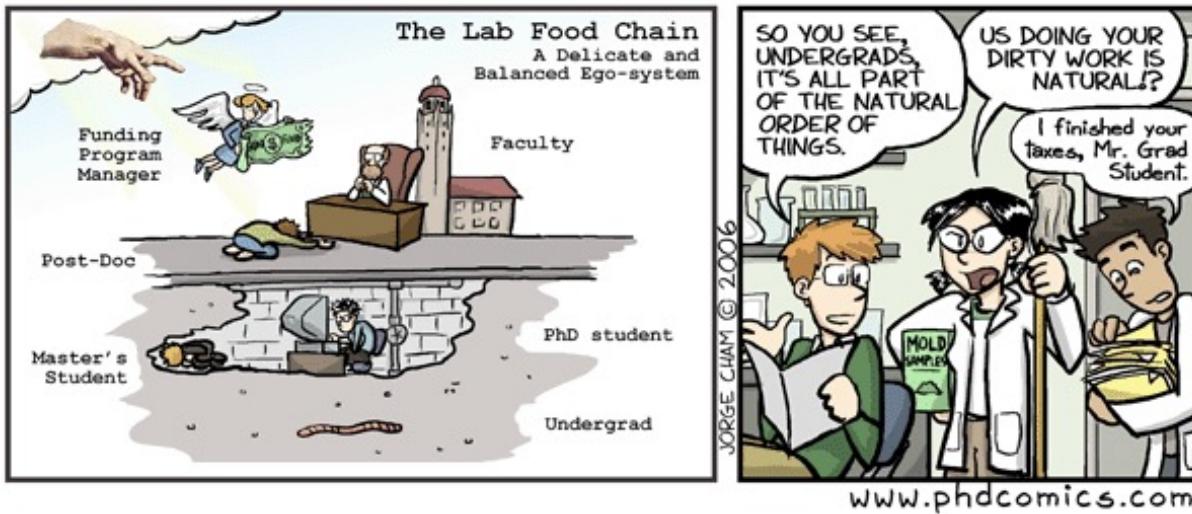
Diversity parameter q : probability incumbents collaborate with prior collaborators. (Higher q means more previous collaborators)

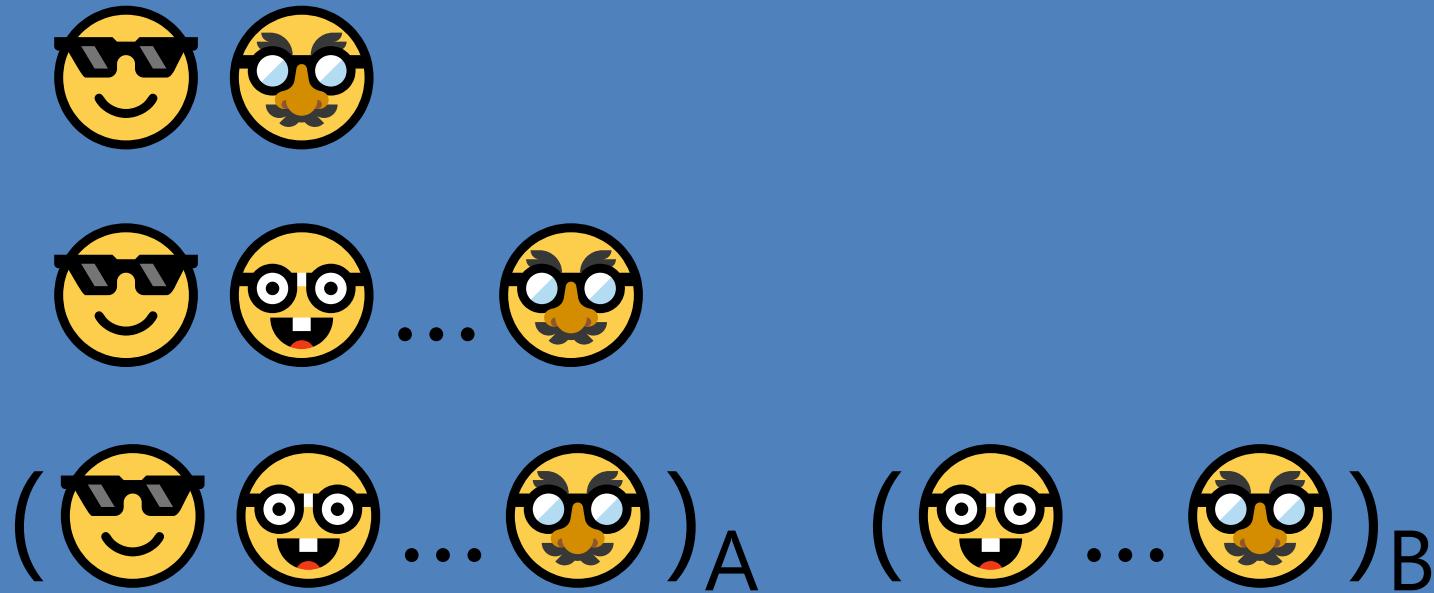
How do we use this model?

Finding: Journal impact positively correlates with p ,
negatively correlates with q .

Takeaways?

... but how does it apply to me?







(...) _A _B

()_A (...) _B

(...) _A (...) _B



_A



_B

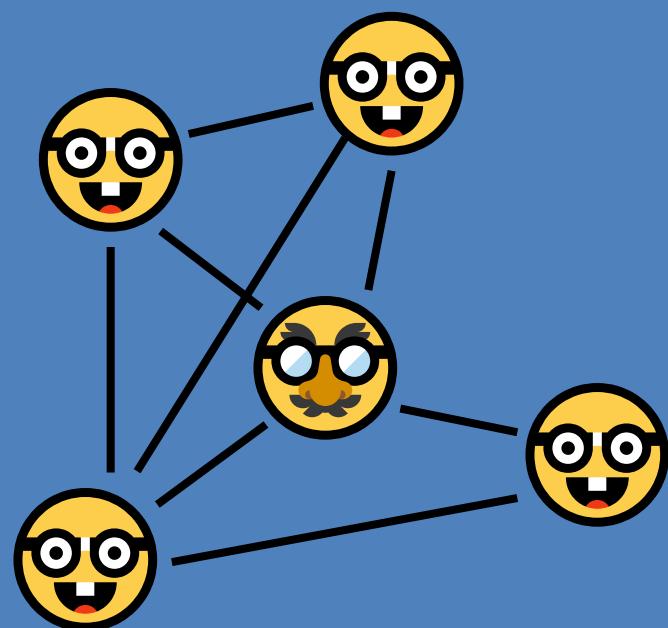
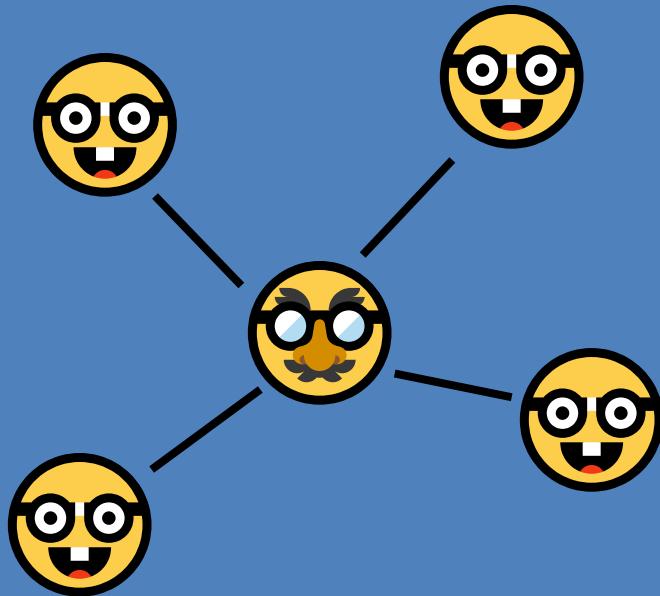


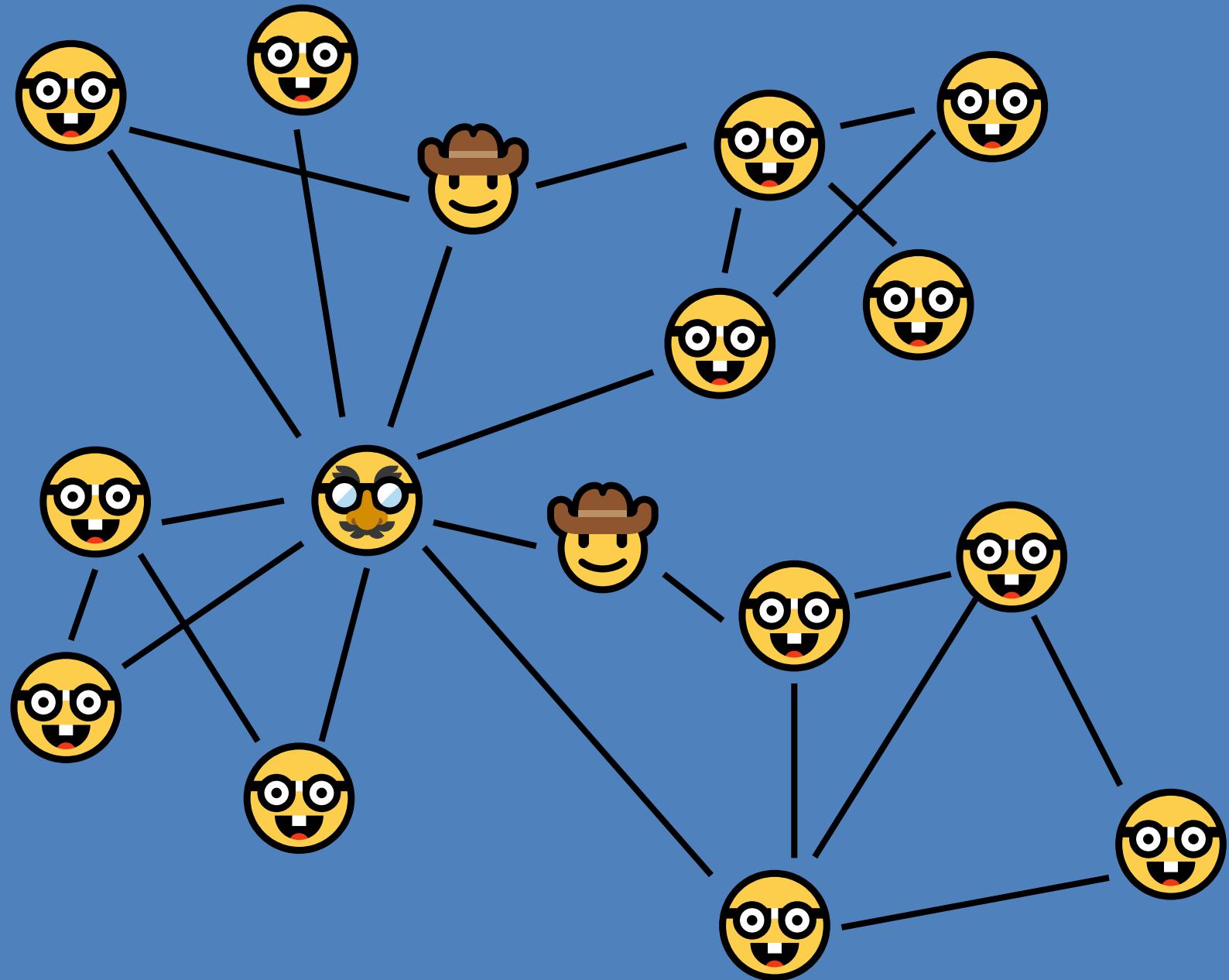
A

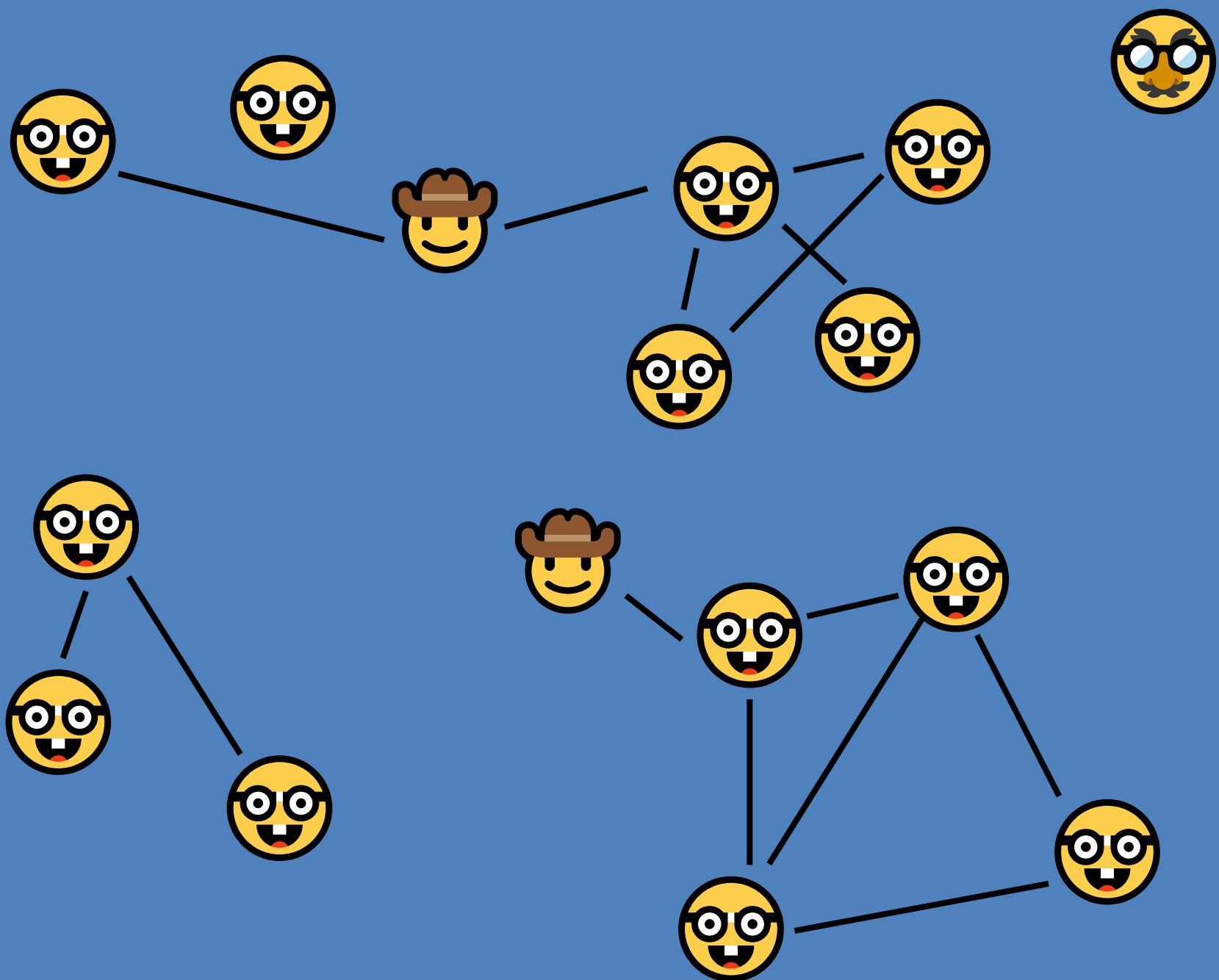


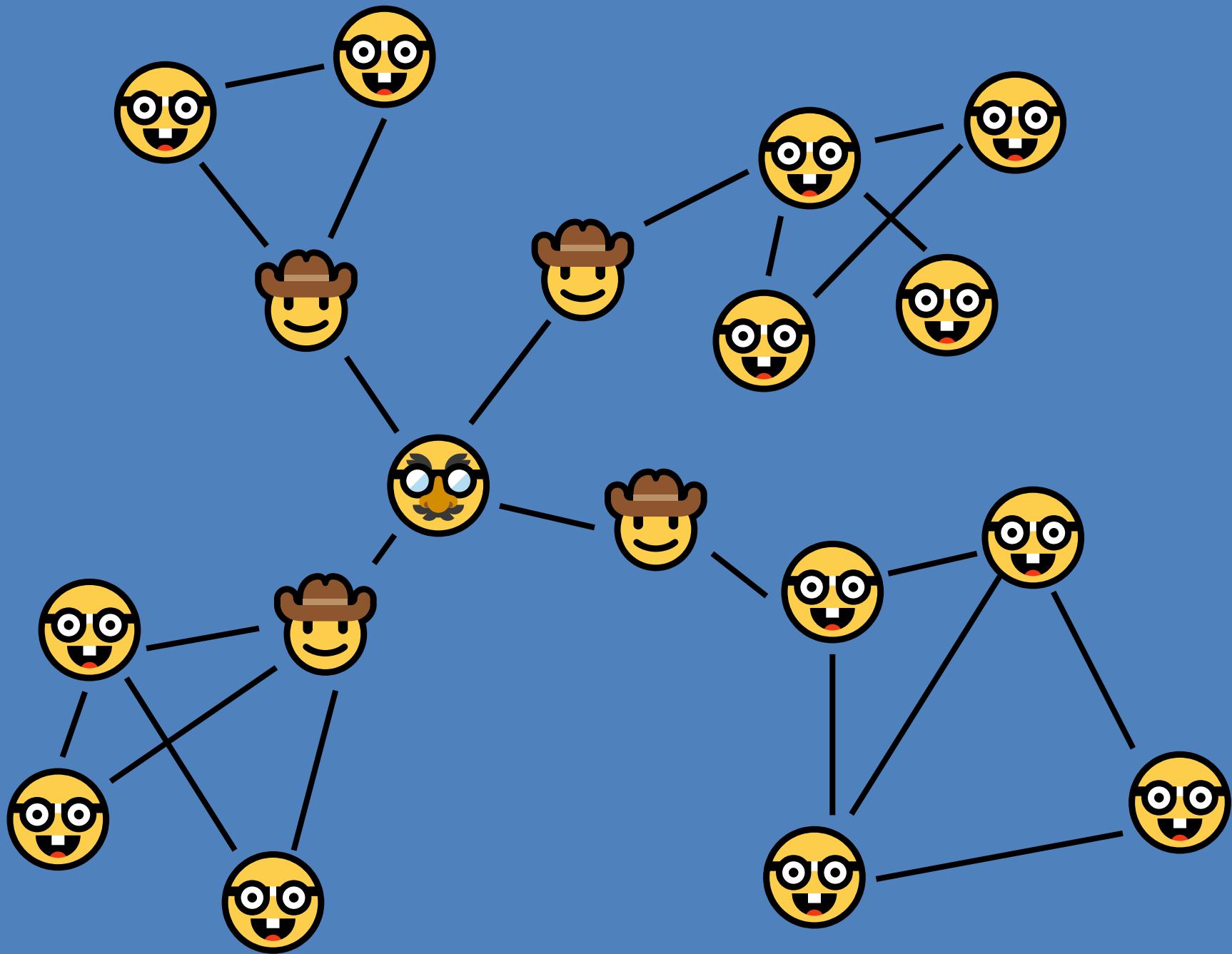
B

Same organization?
Industry vs. Academia?
Experience?
Complementary?









How do teams *really* form?

Source: student vs. advisor
Actual mechanisms?

What should you do?

Shockley's Model

$N \sim p_1 p_2 p_3 p_4 p_5 p_6 p_7 p_8$

- F₁. Identify a good problem
- F₂. Make progress with it
- F₃. Recognize a worthwhile result
- F₄. Decide when to stop the research and start writing up the results
- F₅. Write adequately
- F₆. Profit constructively from criticism
- F₇. Show determination to submit the paper for publication
- F₈. Make changes if required by the journal or the referees

tl;dr – my advice:

Seek homogeneity to share workload

Seek complementarity for everything else

Important: let your advisor know!

“The Science of Science”, Part II

The Science of Collaboration

Chapter 12: Small and Large Teams

Discussion Points

Why are teams growing in size?

Top-down vs. Bottom-up

Who determines what research gets done?

“The Science of Science”, Part II

The Science of Collaboration

Chapter 13: Scientific Credit
Chapter 14: Credit Allocation

Discussion Points

Normative vs. Positive

Understanding norms

Practical advice

Who gets credit?
Who *should* get credit?
Who *actually* gets credit?

Is it fair?

a VOLUME 76, NUMBER 11

PHYSICAL REVIEW LETTERS

11 MARCH 1996

Generation of Nonclassical Motional States of a Trapped Atom

D. M. Meekhof, C. Monroe, B. E. King, W. M. Itano, and D. J. Wineland

Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80303-3328



b VOLUME 55, NUMBER 1

PHYSICAL REVIEW LETTERS

1 JULY 1985

**Three-Dimensional Viscous Confinement and Cooling of Atoms
by Resonance Radiation Pressure**

Steven Chu, L. Hollberg, J. E. Bjorkholm, Alex Cable, and A. Ashkin

AT&T Bell Laboratories, Holmdel, New Jersey 07733



c VOLUME 61, NUMBER 21

PHYSICAL REVIEW LETTERS

21 NOVEMBER 1988

Giant Magnetoresistance of (001) Fe/(001) Cr Magnetic Superlattices

M. N. Baibich, ^(a) J. M. Broto, A. Fert, F. Nguyen Van Dau, and F. Petroff

Laboratoire de Physique des Solides, Université Paris-Sud, F-91405 Orsay, France

P. Eitenne, G. Creuzet, A. Friederich, and J. Chazelas

Laboratoire Central de Recherches, Thomson CSF, B.P. 10, F-91401 Orsay, France



Figure 2.6.1 **Who gets the Nobel?** (A) The last author, David J. Wineland was awarded the 2012 Nobel Prize in Physics for his contribution to quantum computing. (B) Steven Chu, the first author, won the 1997 Nobel in Physics for the paper focusing on the cooling and trapping of atoms with laser light. (C) In 2007, Albert Fert, the middle author of the paper, received the Nobel Prize in Physics for the discovery of the giant magnetoresistance effect (GMR). All three examples are prize-winning papers published in the same journal, *Physical Review Letters*, demonstrating the ambiguity of allocating credit by simply reading the byline of a paper.

AlphaFold developers win US\$3-million Breakthrough Prize

DeepMind's system for predicting the 3D structure of proteins is among five recipients of science's most lucrative awards.

[Zeeya Merali](#)



Demis Hassabis (left) and John Jumper (right) from DeepMind developed AlphaFold, an AI that can predict the structure of proteins. Credit: Breakthrough Prize

The researchers behind the AlphaFold artificial-intelligence (AI) system have won one of this year's US\$3-million Breakthrough prizes – the most lucrative awards in science. Demis Hassabis and John Jumper, both at DeepMind in London, were recognized for creating the tool that has predicted the 3D structures of almost every known protein on the planet.



Dan Roy
@roydanroy

...

I was curious how AlphaFold was honored by the breakthrough prize to two individuals. I went back to one of the Nature papers to look at the author contributions. Curious what people see that could justify two people being named.

9:35 PM · Sep 23, 2022 · Twitter Web App

<https://twitter.com/roydanroy/status/1573486328502689827>

Improved protein structure prediction using potentials from deep learning

[Andrew W. Senior](#) [Richard Evans](#), [John Jumper](#), [James Kirkpatrick](#), [Laurent Sifre](#), [Tim Green](#), [Chongli Qin](#), [Augustin Žídek](#), [Alexander W. R. Nelson](#), [Alex Bridgland](#), [Hugo Penedones](#), [Stig Petersen](#), [Karen Simonyan](#), [Steve Crossan](#), [Pushmeet Kohli](#), [David T. Jones](#), [David Silver](#), [Koray Kavukcuoglu](#) & [Demis Hassabis](#)

[Nature](#) **577**, 706–710 (2020) | [Cite this article](#)

Contributions

R.E., J.J., J.K., L.S., A.W.S., C.Q., T.G., A.Ž., A.B., H.P. and K.S. designed and built the AlphaFold system with advice from D.S., K.K. and D.H. D.T.J. provided advice and guidance on protein structure prediction methodology. S.P. contributed to software engineering. S.C., A.W.R.N., K.K. and D.H. managed the project. J.K., A.W.S., T.G., A.Ž., A.B., R.E., P.K. and J.J. analysed the CASP results for the paper. A.W.S. and J.K. wrote the paper with contributions from J.J., R.E., L.S., T.G., A.B., A.Ž., D.T.J., P.K., K.K. and D.H. A.W.S. led the team.

Corresponding author

Correspondence to [Andrew W. Senior](#).

THE AUTHOR LIST: GIVING CREDIT WHERE CREDIT IS DUE

The first author

Senior grad student on the project. Made the figures.

Michaels, C., Lee, E. F., Sap, P. S., Nichols, S. T., Oliveira, L., Smith, B. S.

The third author

First year student who actually did the experiments, performed the analysis and wrote the whole paper. Thinks being third author is "fair".

The second-to-last author

Ambitious assistant professor or post-doc who instigated the paper.

The second author

Grad student in the lab that has nothing to do with this project, but was included because he/she hung around the group meetings (usually for the food).

The middle authors

Author names nobody really reads. Reserved for undergrads and technical staff.

The last author

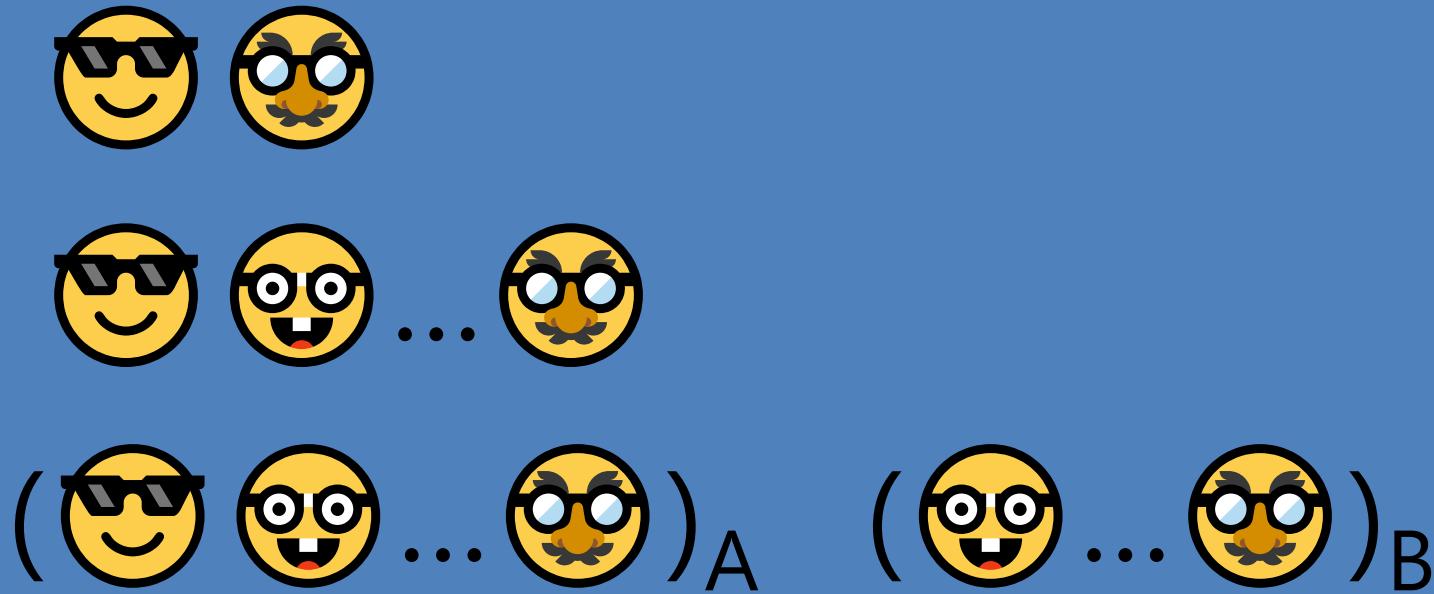
The head honcho. Hasn't even read the paper but, hey, he/she got the funding, and their famous name will get the paper accepted.

... but how does it apply to me?

Ordering of Authors

Field-specific norms (alphabetical vs. contribution-based)
You (as a student) have very limited control

Important first step: understand what the norms are!





(...)_A _B

()_A (...)_B

(...)_A (...)_B



A



B

Practical Advice

Determine coauthorship (and ordering) early.

Don't forget to communicate with non-coauthors.

(Personal view) Be generous with the middle author positions.

Who actually gets the credit? For the most part... your advisor 😞

Why?

How can you overcome it?



“The Science of Science”, Part III

The Science of Impact

Chapter 15: Big Science

Discussion Points

Exponential growth of science

Slowdown in labor productivity growth

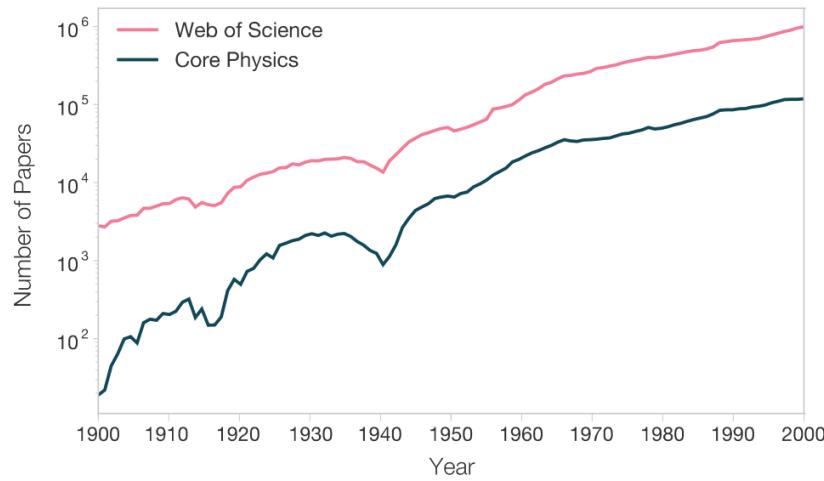
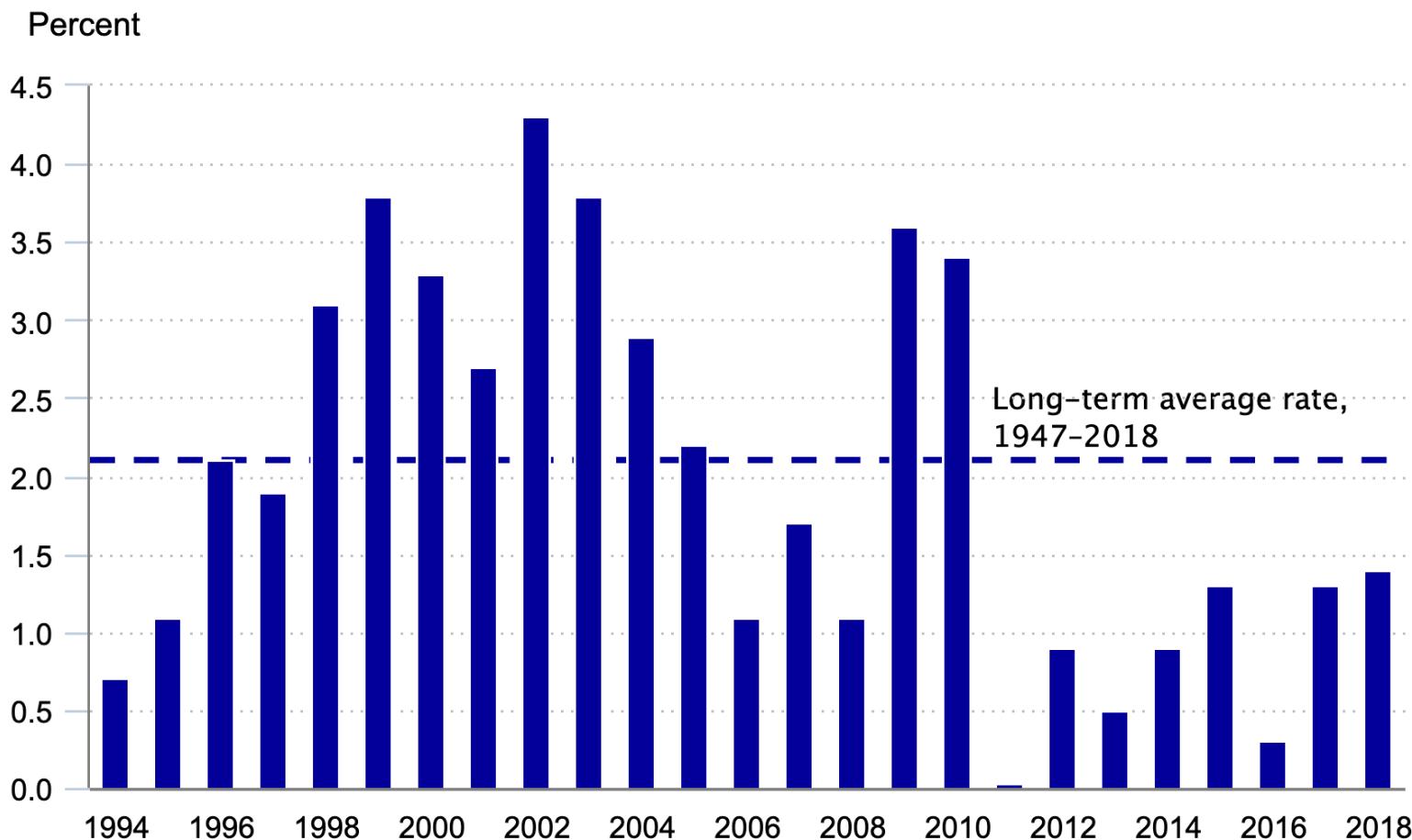


Figure 3.1.1 The Growth of Science. The number of papers catalogued in the Web of Science (WoS) published over the past century, illustrates the exponential growth of the scientific literature. It was only disrupted around 1915 and 1945 due to the World Wars. The figure also shows the growth of the physics literature, which follows an exponential growth similar to that followed by science as a whole. After Sinatra *et al.* [5].

Figure 1. Labor productivity growth: annual percent changes, nonfarm business sector, 1994–2018



From impact... to citations?

Category	Description	Example
Criticizing	Criticism can be positive or negative. A citing sentence is classified as "criticizing" when it mentions the weakness/strengths of the cited approach, negatively/positively criticizes the cited approach, negatively/positively evaluates the cited source.	Chiang (2005) introduced a constituent feature to reward phrases that match a syntactic tree but did not yield significant improvement.
Comparison	A citing sentence is classified as "comparison" when it compares or contrasts the work in the cited paper to the author's work. It overlaps with the first category when the citing sentence says one approach is not as good as the other approach. In this case we use the first category.	Our approach permits an alternative to minimum error-rate training (MERT; Och, 2003);
Use	A citing sentence is classified as "use" when the citing paper uses the method, idea or tool of the cited paper.	We perform the MERT training (Och, 2003) to tune the optimal feature weights on the development set.
Substantiating	A citing sentence is classified as "substantiating" when the results, claims of the citing work substantiate, verify the cited paper and support each other.	It was found to produce automated scores, which strongly correlate with human judgements about translation fluency (Papineni et al., 2002).
Basis	A citing sentence is classified as "basis" when the author uses the cited work as starting point or motivation and extends on the cited work.	Our model is derived from the hidden-markov model for word alignment (Vogel et al., 1996; Och and Ney, 2000).
Neutral (Other)	A citing sentence is classified as "neutral" when it is a neutral description of the cited work or if it doesn't come under any of the above categories.	The solutions of these problems depend heavily on the quality of the word alignment (Och and Ney, 2000).

Table 2: Annotation scheme for citation purpose. Motivated by the work of (Spiegel-Rösing, 1977) and (Teufel et al., 2006)

Intent category	Definition	Example
Background information	The citation states, mentions, or points to the background information giving more context about a problem, concept, approach, topic, or importance of the problem in the field.	Recent evidence suggests that co-occurring alexithymia may explain deficits [12]. Locally high-temperature melting regions can act as permanent termination sites [6-9]. One line of work is focused on changing the objective function (Mao et al., 2016).
Method	Making use of a method, tool, approach or dataset	Fold differences were calculated by a mathematical model described in [4]. We use Orthogonal Initialization (Saxe et al., 2014)
Result comparison	Comparison of the paper’s results/findings with the results/findings of other work	Weighted measurements were superior to T2-weighted contrast imaging which was in accordance with former studies [25-27] Similar results to our study were reported in the study of Lee et al (2010).

Table 1: The definition and examples of citation intent categories in our SciCite.

“The Science of Science”, Part III

The Science of Impact

Chapter 16: Citation Disparity

Discussion Points

Lognormal distribution

Importance of normalization

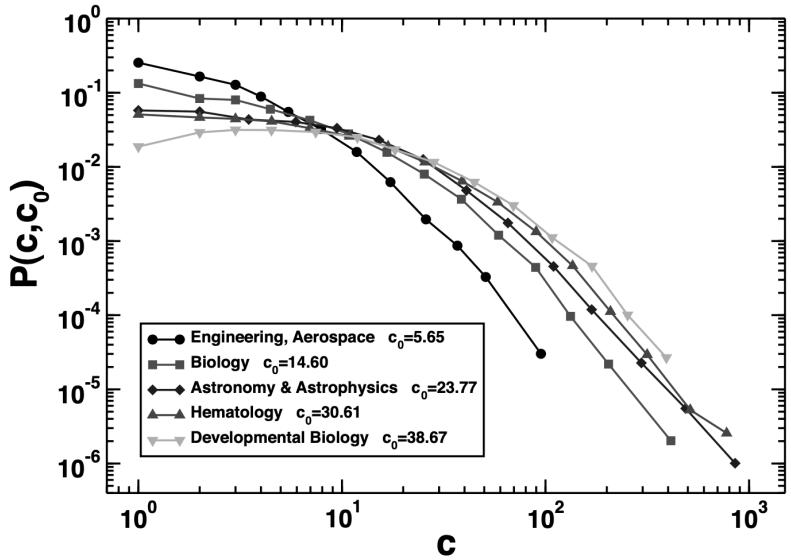


Fig. 1. Normalized histogram of the number of articles $P(c, c_0)$ published in 1999 and having received c citations. We plot $P(c, c_0)$ for several scientific disciplines with different average number c_0 of citations per article.

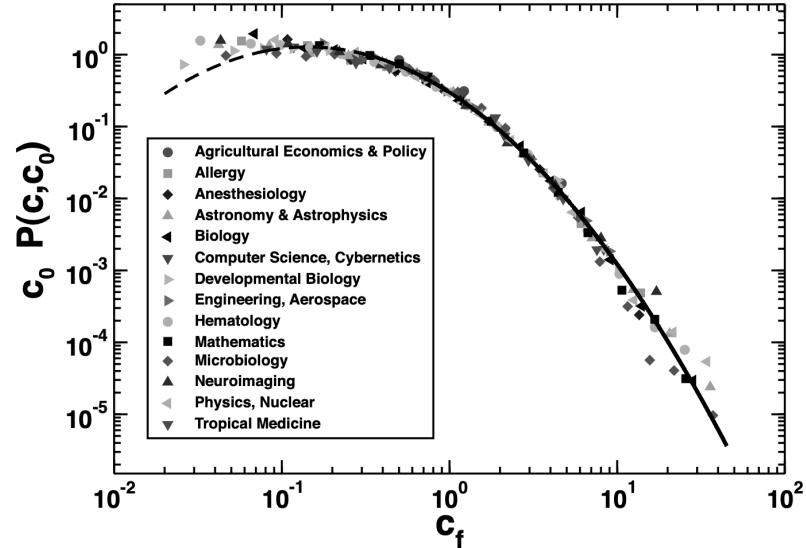


Fig. 2. Rescaled probability distribution $c_0 P(c, c_0)$ of the relative indicator $c_f = c/c_0$, showing that the universal scaling holds for all scientific disciplines considered (see Table 1). The dashed line is a lognormal fit with $\sigma^2 = 1.3$.

“The Science of Science”, Part III

The Science of Impact

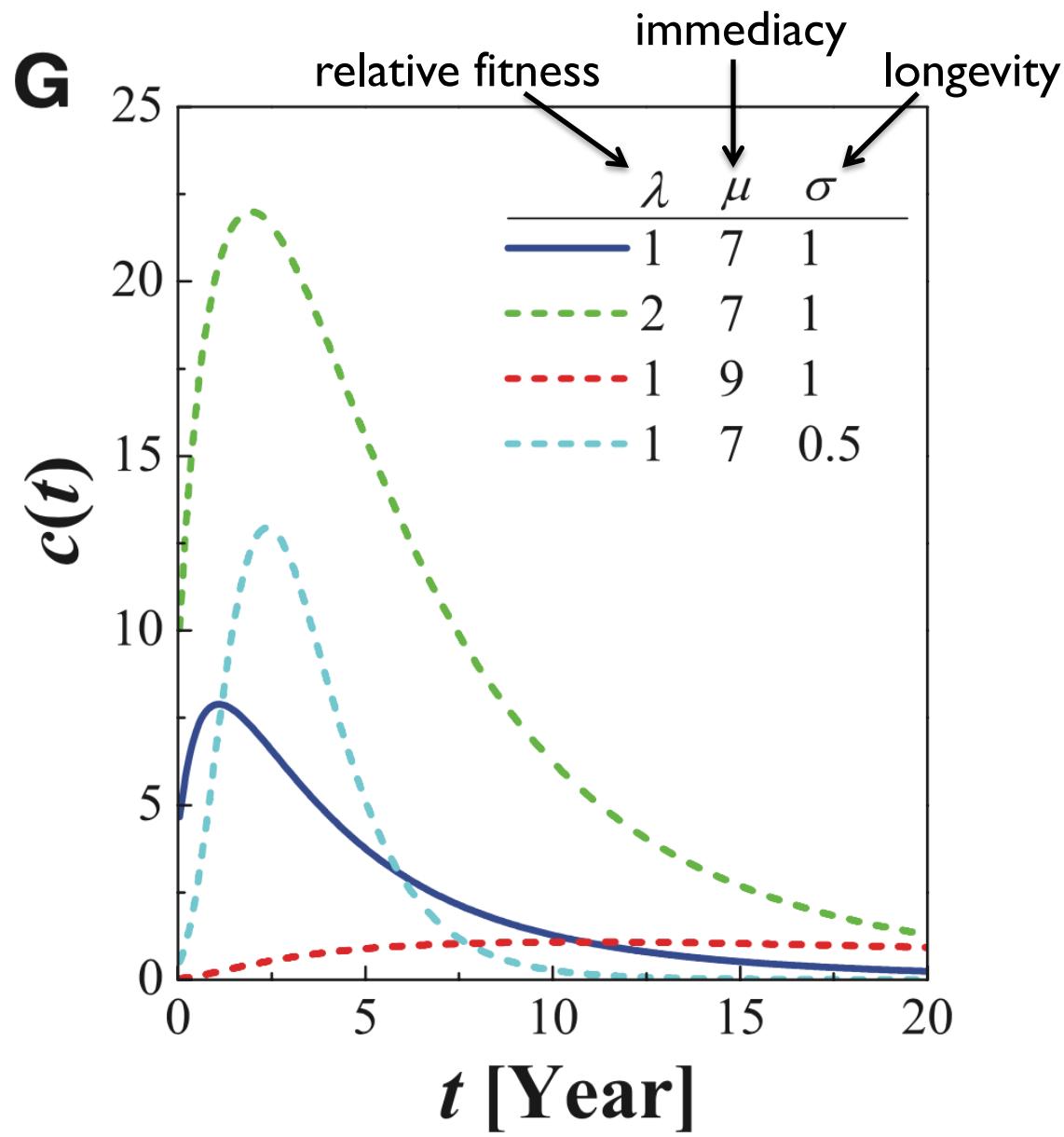
Chapter 17 – 20: Citation Models

Discussion Points

Price Model

Bianconi–Barabási Model

Wang–Song–Barabási Model



What about ultimate impact?

What *really* about ultimate impact?

That's all for this week!