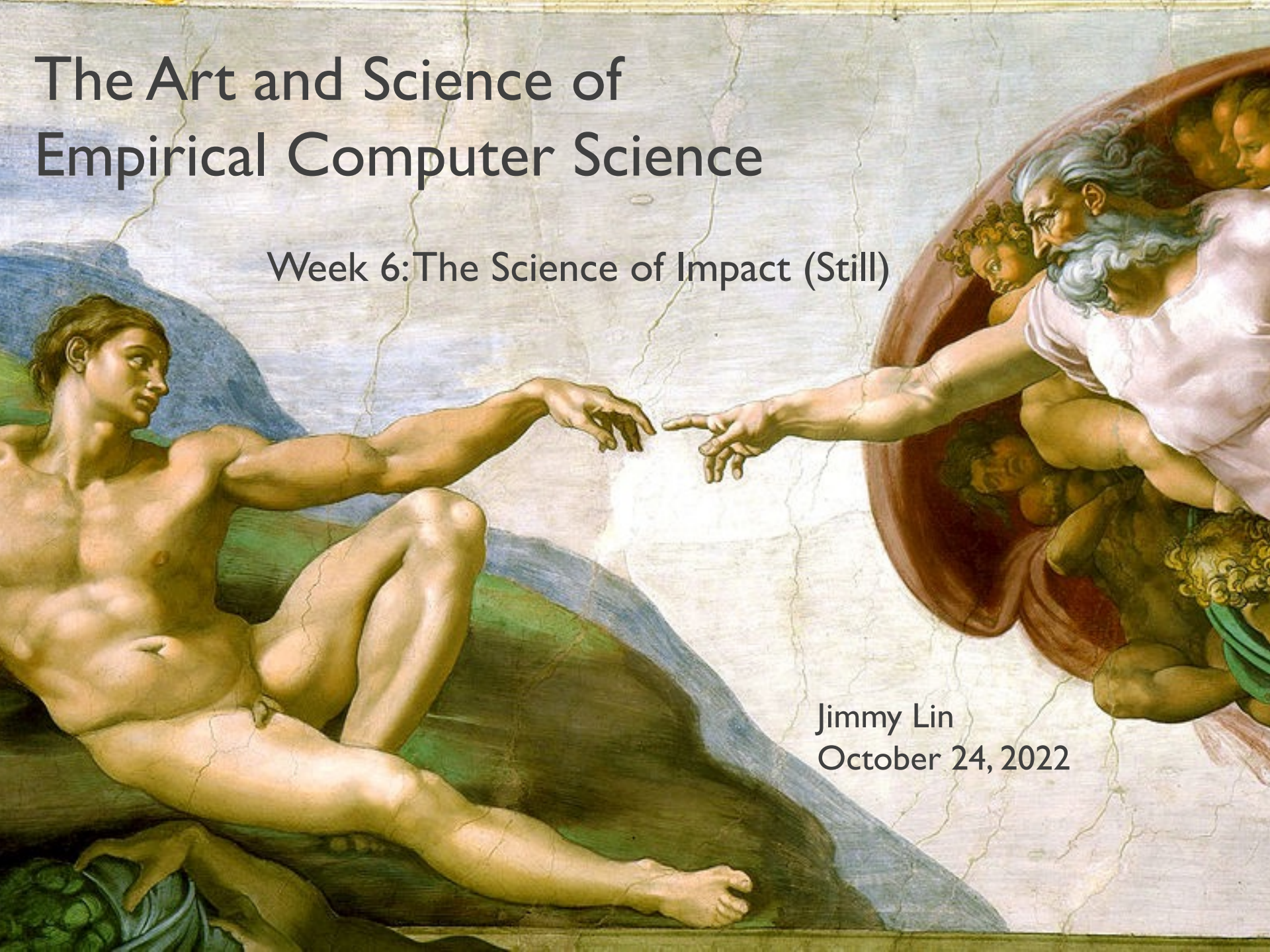
The Art and Science of Empirical Computer Science



Week 6: The Science of Impact (Still)

Jimmy Lin
October 24, 2022

Social Media: Waste of Time?

Position A: Early-stage researchers should actively incorporate social media use as a component of their career development. This means appropriate use of sites like Twitter, Facebook, and LinkedIn to build professional reputation, engage with the community, hear about recent work by others, etc.

Position B: Early-stage researchers should stay off social media. It's a complete waste of time.

Let’s debate!


https://www.flickr.com/photos/trofeunteun/33278869018/

"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

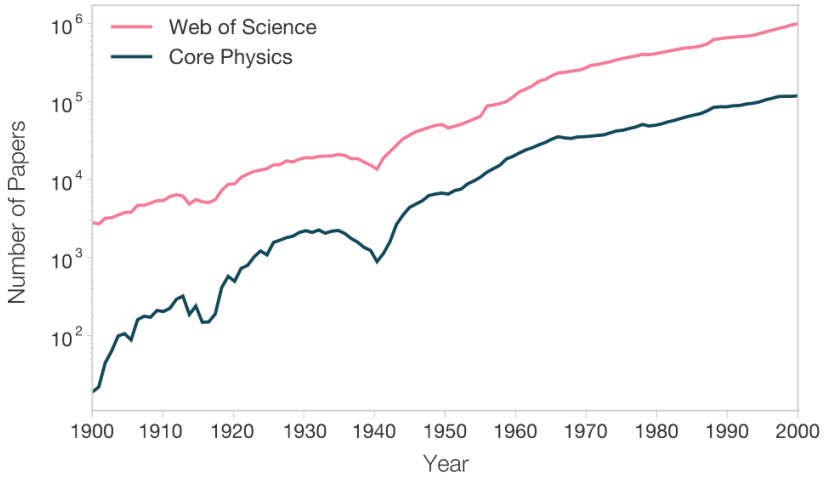
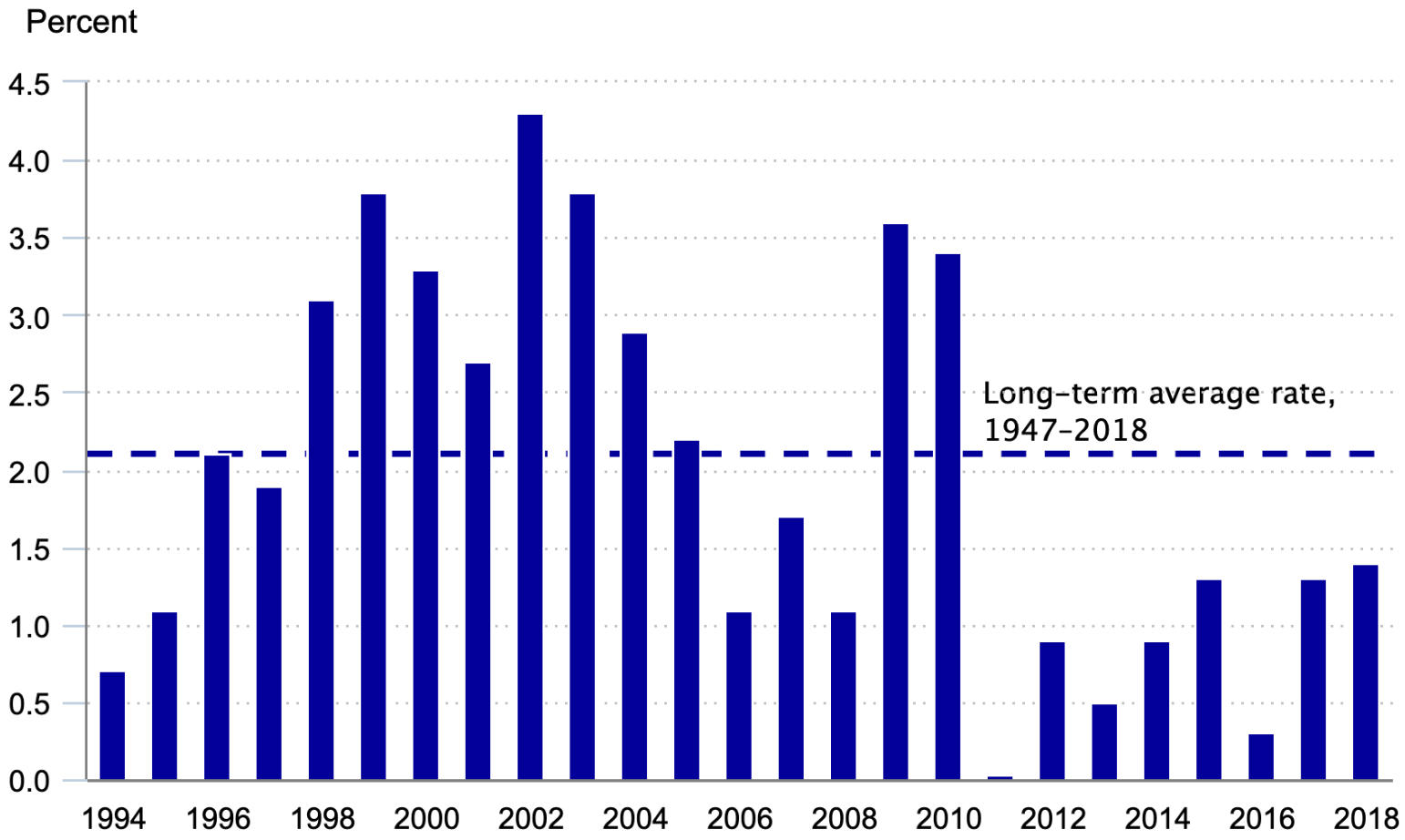


Figure 3.1.1 **The Growth of Science**. The number of papers catalogued in the Web of Science (WoS) published over the past century, illustrates the exponential growth of the scientific literature. It was only disrupted around 1915 and 1945 due to the World Wars. The figure also shows the growth of the 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



https://www.bls.gov/opub/mlr/2021/article/the-us-productivity-slowdown-the-economy-wide-and-industry-level-analysis.htm

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)

Purpose and Polarity of Citation: Towards NLP-based Bibliometrics (Abu-Jbara et al., NAACL 2013)
https://aclanthology.org/N13-1067/

Intent cateogry	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.

Structural Scaffolds for Citation Intent Classification in Scientific Publications (Cohan et al., NAACL 2019)
https://aclanthology.org/N19-1361/

"The Science of Science", Part III

The Science of Impact

Chapter 16: Citation Disparity

Discussion Points

Lognormal distribution

Importance of normalization

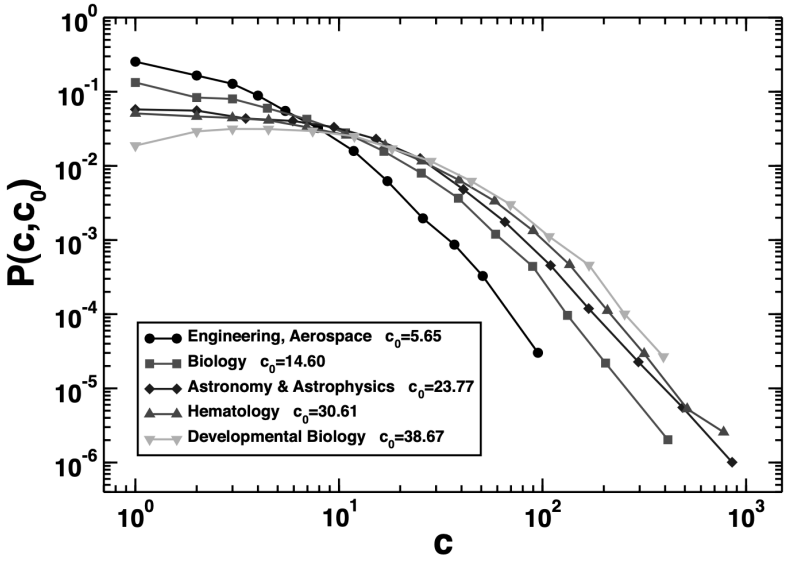


Fig. 1. Normalized histogram of the number of articles $P(c,c_0)$ published in 1999 and having received c citations. We plot $P(c,c_0)$ for several scientific disciplines with different average number c_0 of citations per article.

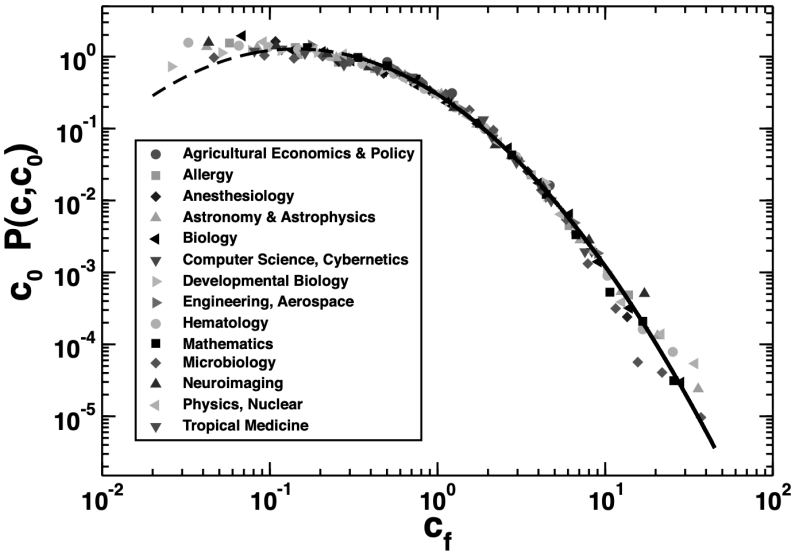


Fig. 2. Rescaled probability distribution $c_0\,P(c,c_0)$ of the relative indicator $c_f = c/c_0$, showing that the universal scaling holds for all scientific disciplines considered (see Table 1). The dashed line is a lognormal fit with $\sigma^2 = 1.3$.

Radicchi et al. Universality of citation distributions: Toward an objective measure of scientific impact. *PNAS*, 105(45):17268-17272.

“The Science of Science”, Part III

The Science of Impact

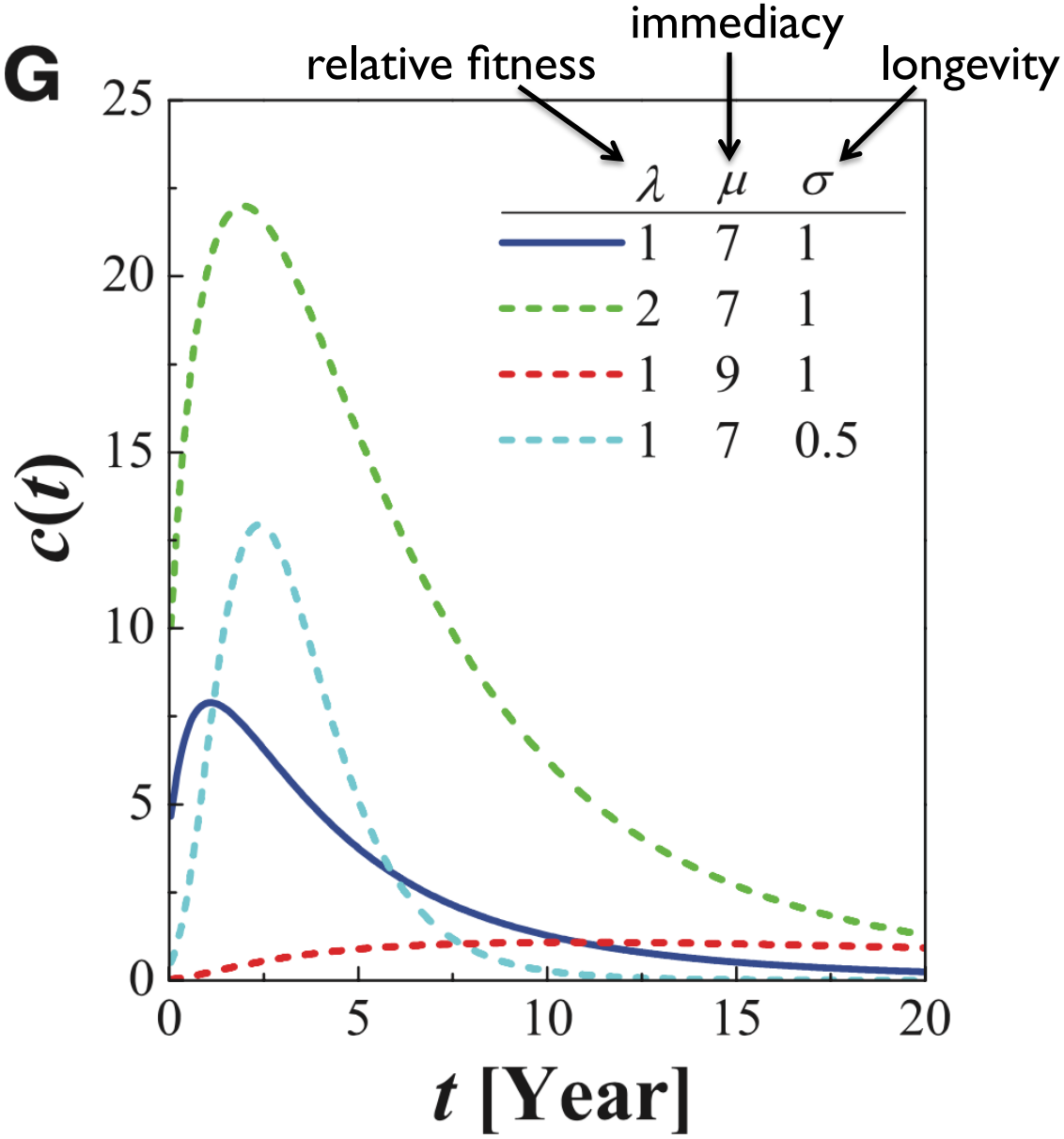
Chapter 17 – 20: Citation Models

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!