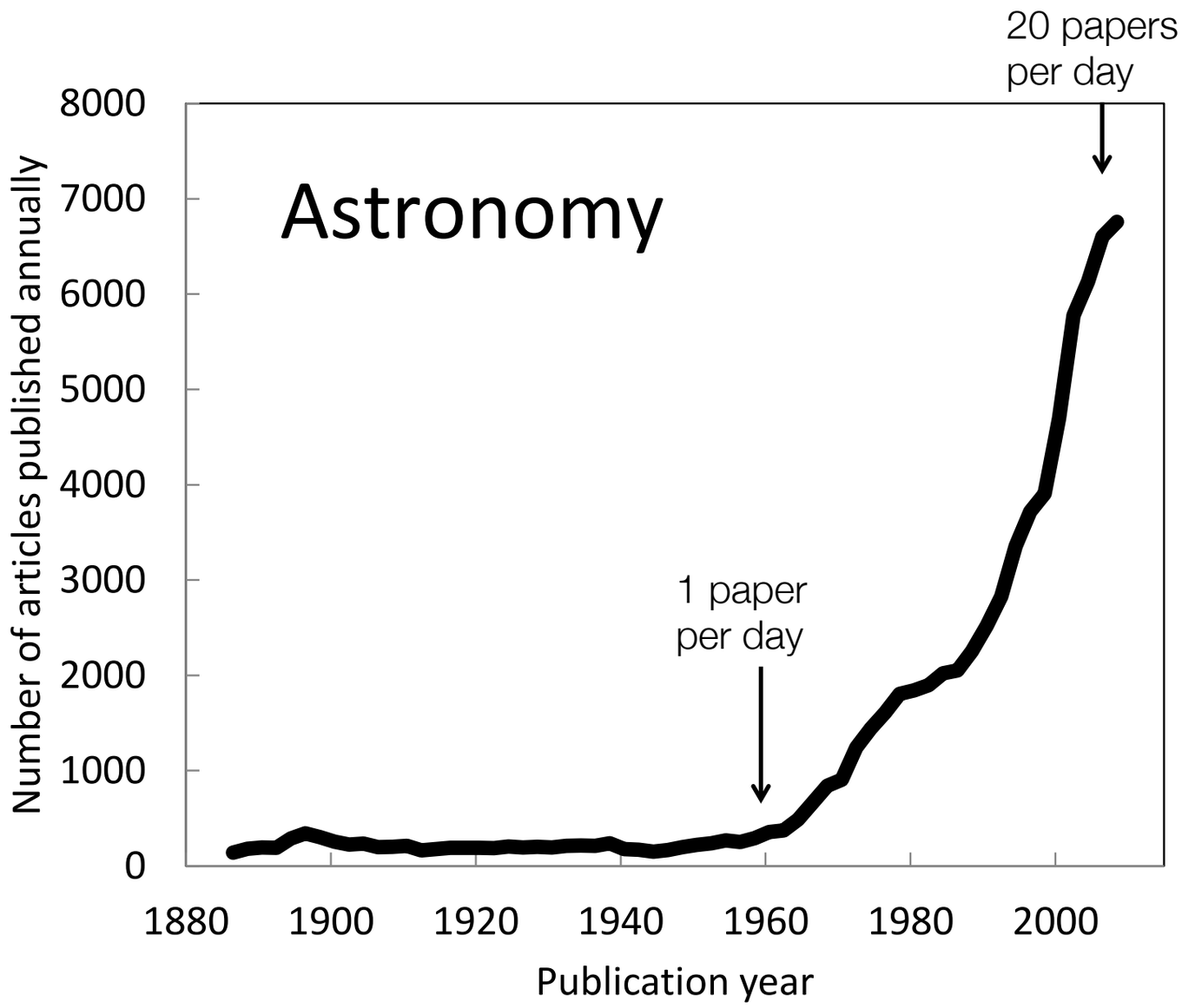
Intellectual diversity of scientific disciplines and research teams

Staša Milojević

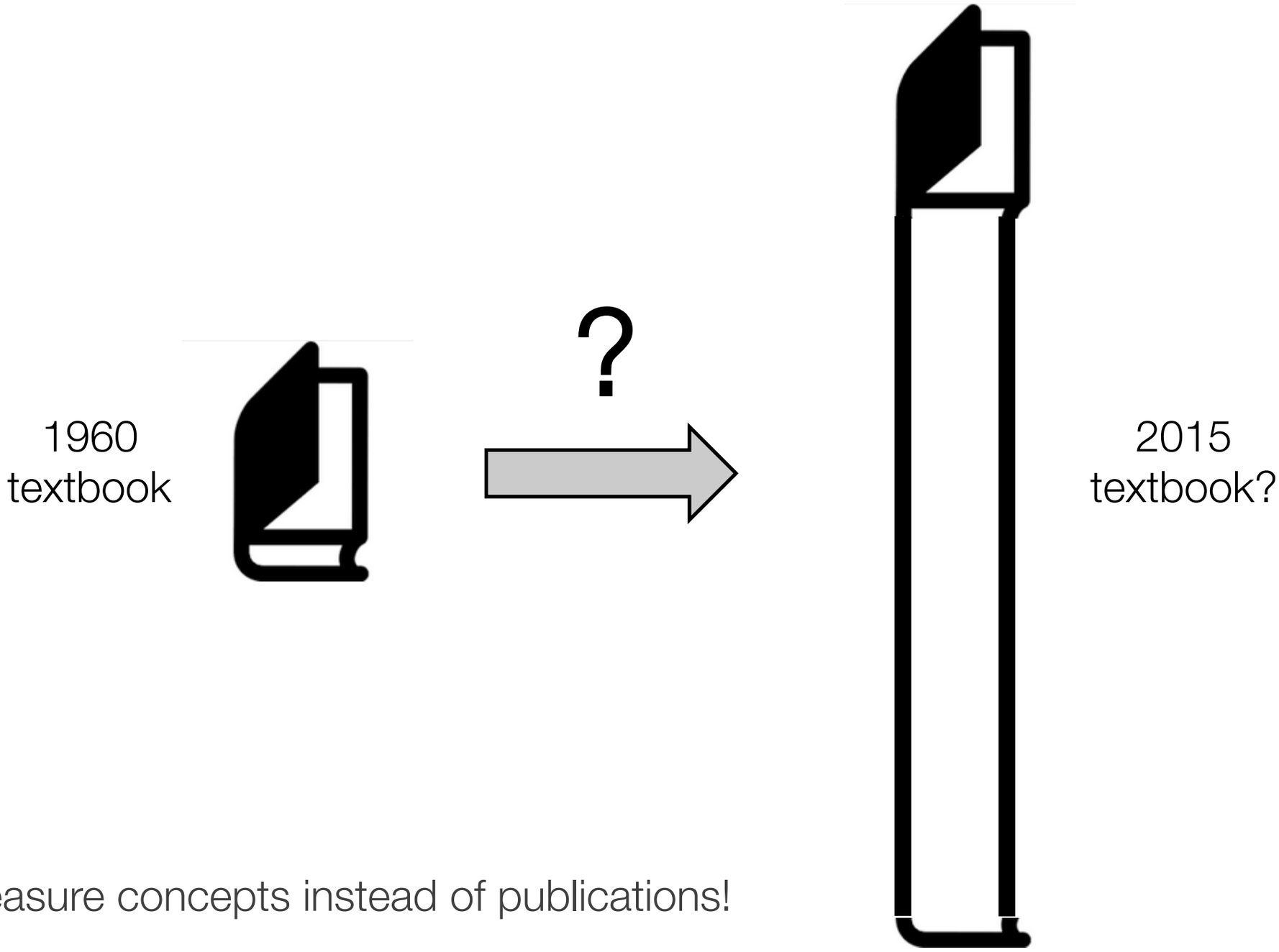
School of Informatics and Computing, Indiana University Bloomington

Exponential rise in scientific literature

- Doubling time 10-20 years



Is publication volume a good proxy for the growth of science?



Measure concepts instead of publications!

Cognitive concepts from document titles

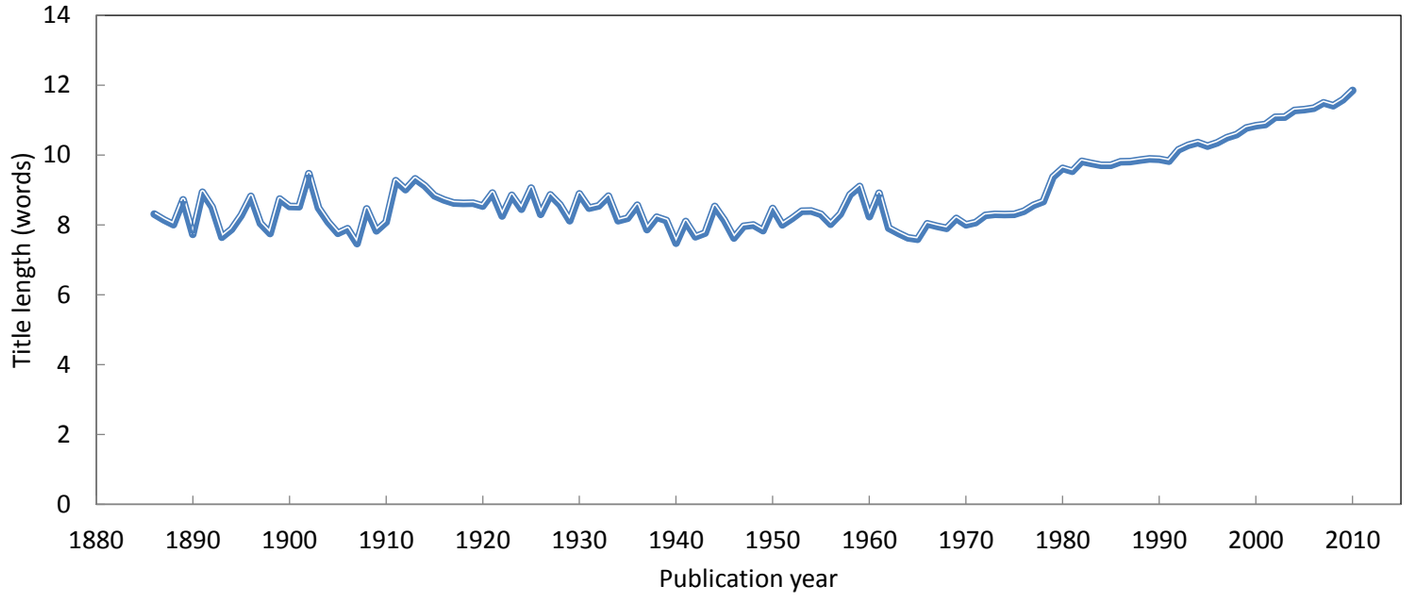
- Full text not normally available
- Abstracts: availability incomplete
- Keywords: very restricted
- Titles function as “attention triggers” (Bazerman 1985,1989)
- Titles have undergone a change during the 20th century, becoming
    - More informative
    - More specific
    - Containing a larger number of words that indicate article content
- Any given article title captures the concepts only crudely and incompletely

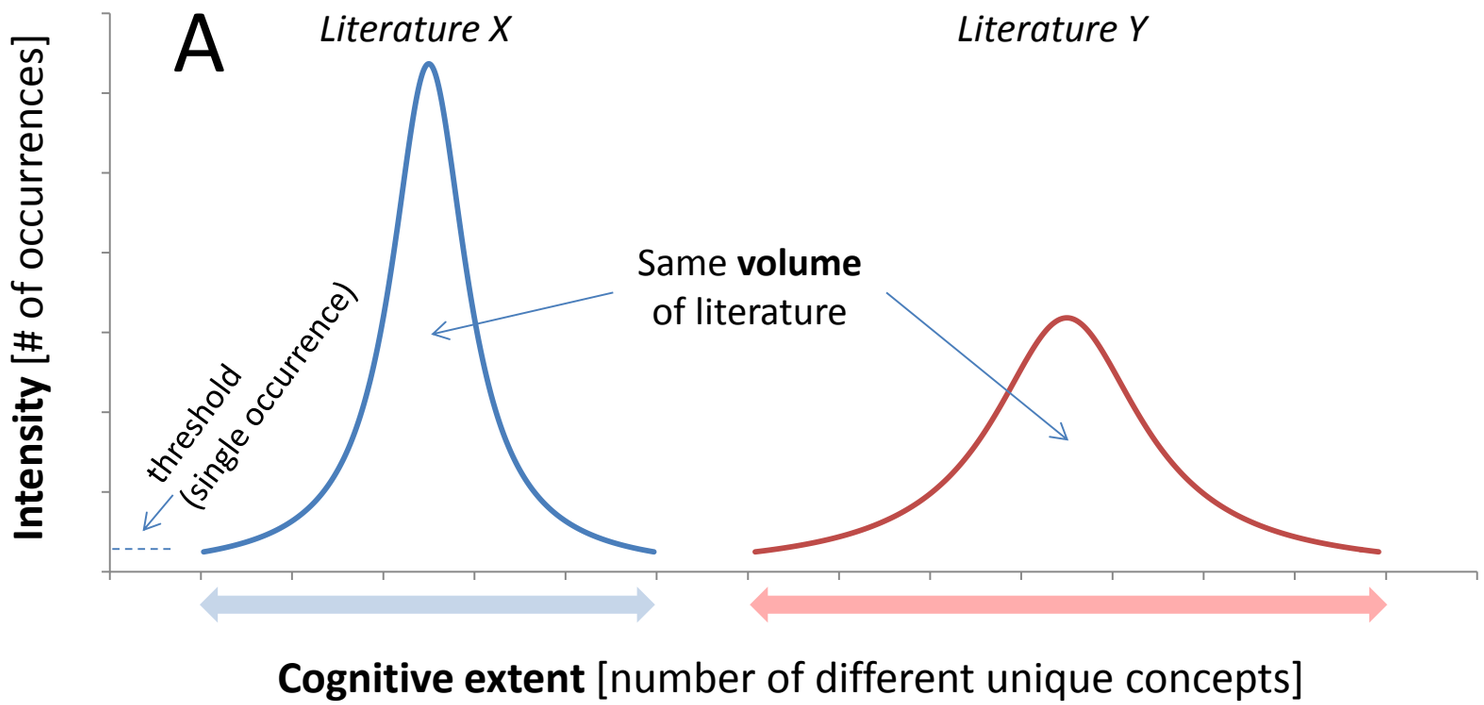
  => Use very large number of articles!

Measuring cognitive content from titles

- Idea: count number of unique phrases in titles
- Must normalize for unequal volume of literature

=> **Cognitive content: number of unique phrases in 10,000 titl words (~1000 articles)**





Concepts from titles

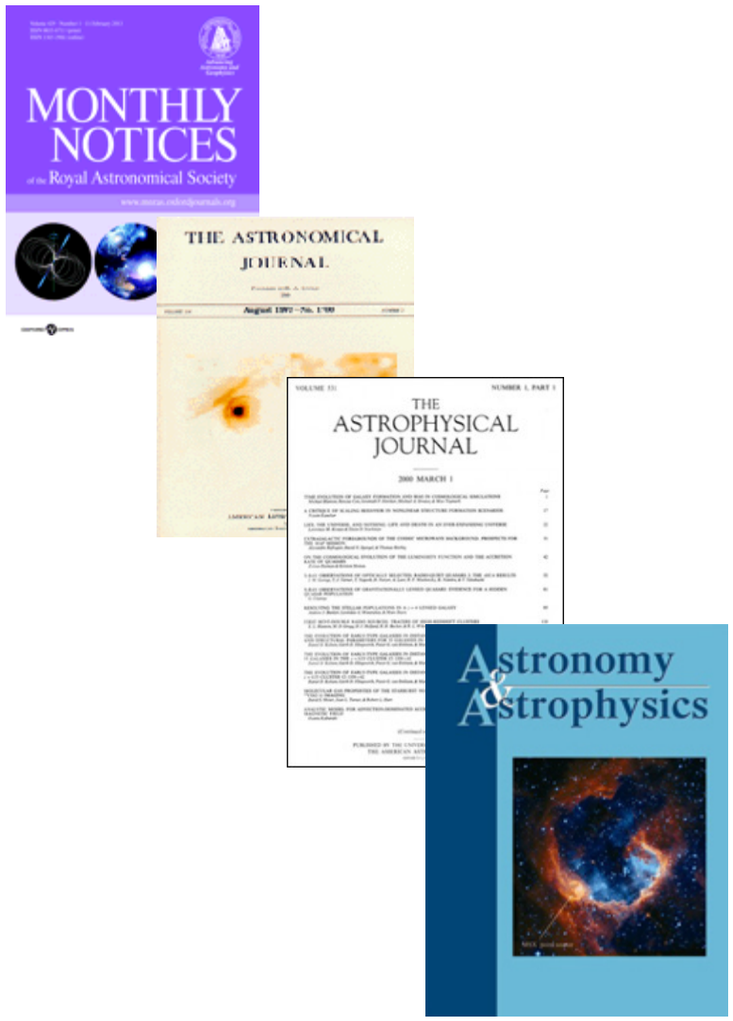
- Automatically identify phrases
- Use general words to separate phrases
- General words: prepositions, articles + non-specific words, such as:

  study, analysis, result, determination, comparison, discovery
- Examples of titles with phrases capitalized:
  - HALOS and VOIDS in MULTIFRACTAL+MODEL of COSMIC+STRUCTURE
  - COLLISION+STRENGTHS for ELECTRON+IMPACT+EXCITATION of FINE+STRUCTURE+LEVELS in FE+XIII

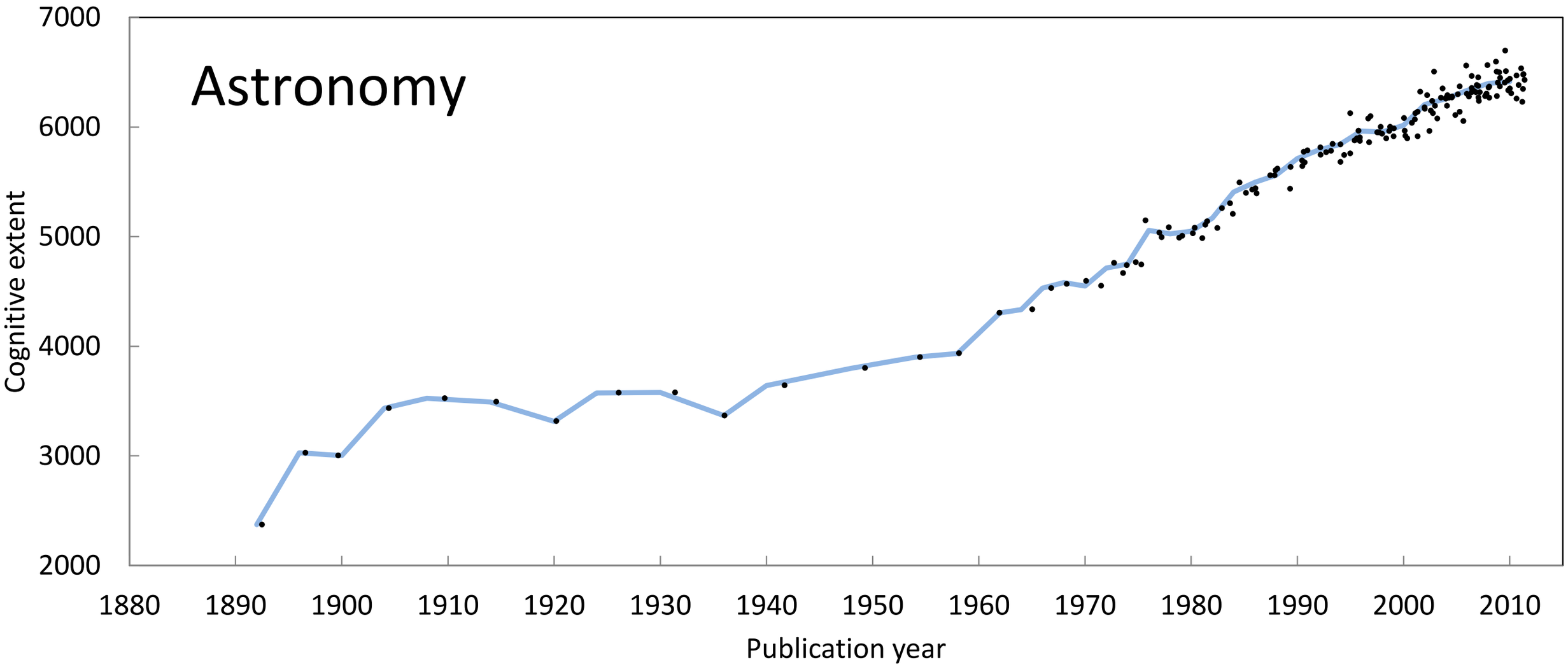
- Old words in new context:
  - MASS, ENERGY -> MASS-ENERGY EQUIVALENCE
  - considered a new phrase (introduced in relativity theory)

Data

- Research articles for:
  - Astronomy (160,000 articles since 1892)
  - Physics (440,000 articles since 1895)
  - Biomedicine (20 million articles since 1946)
- Sources:
  - Thompson-Reuters Web of Science
  - NASA ADS
  - American Physical Society
  - PubMed

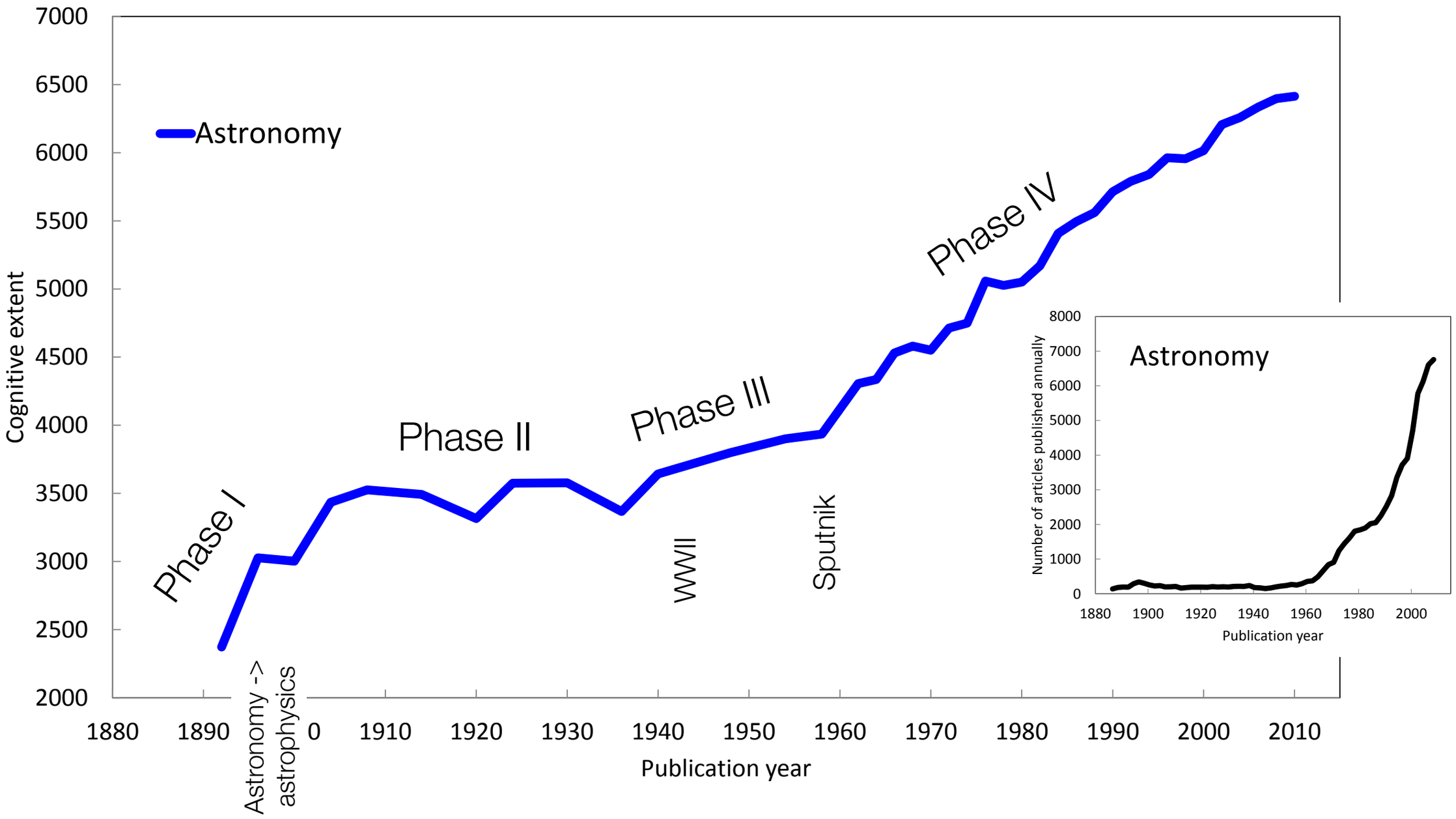


Evolution of the cognitive extent of astronomy



Dot = ca. 1000 articles
Blue line = 2-year average
Standard deviation from independent batches of articles: ~ 1.5%

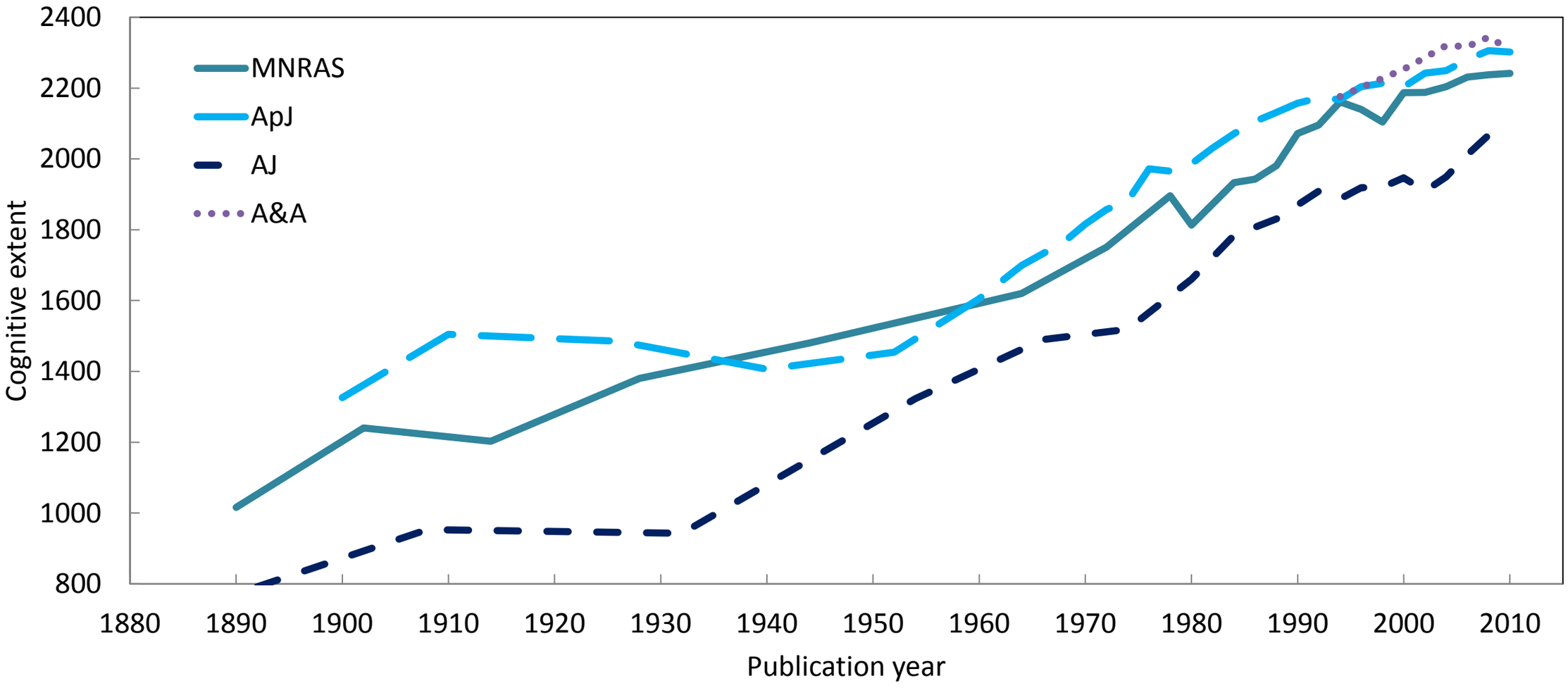
Evolution of the cognitive extent of astronomy



Milojević, S. (in prep.)

Cognitive extent by journal

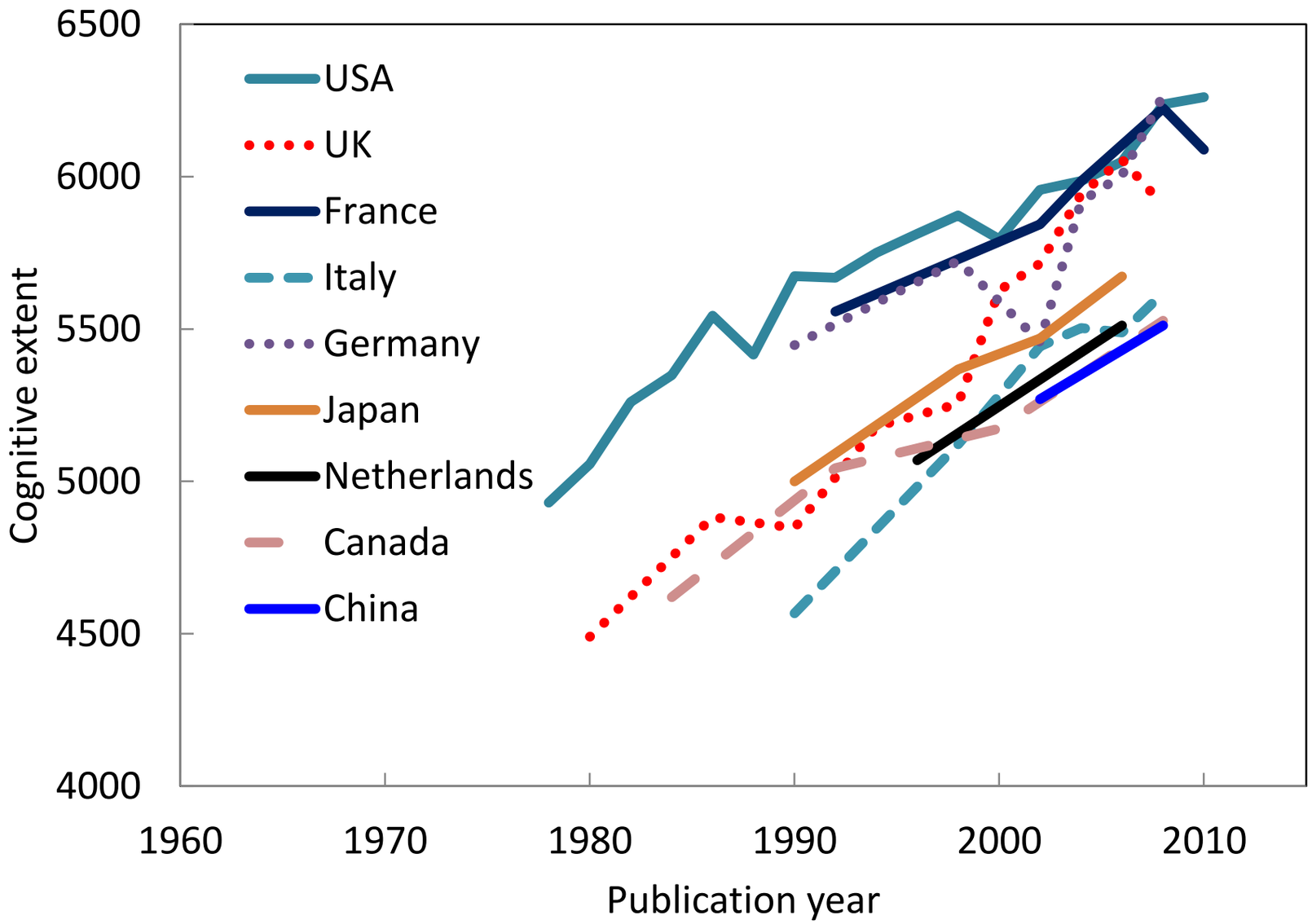
- Cognitive extent should not be equaled with quality!



Milojević, S. (in prep.)

Cognitive content by country

- Country of lead author
- Again, this is all normalized to the same publication volume

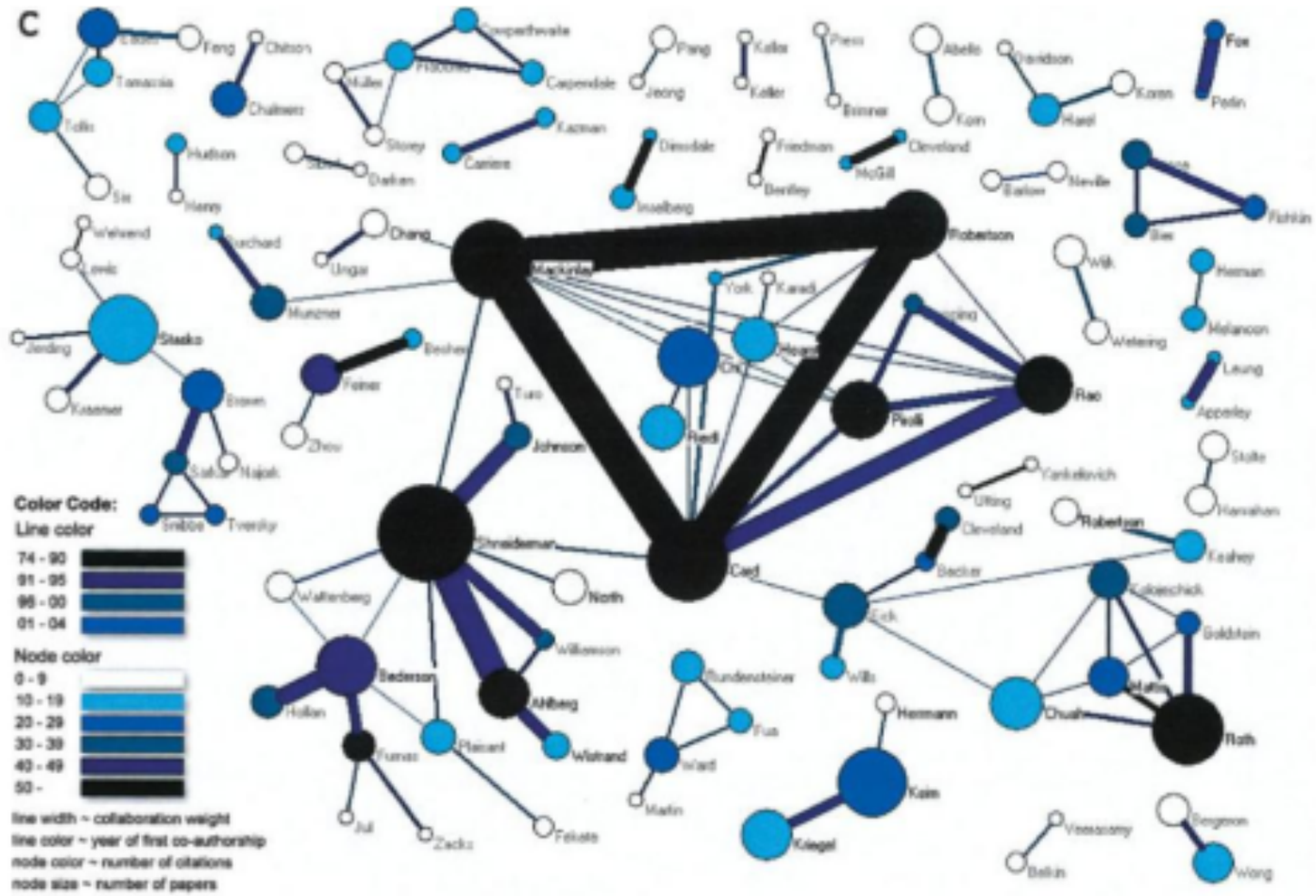


Milojević, S. (in prep.)

Science as a collaborative effort

- Collaboration is one of the defining features of modern science

- Most common way of studying collaboration: through **coauthorship networks**
  - Connect authors on a paper
  - Add to the network



Teams as foundation for collaboration

- Teams are becoming dominant in the production of knowledge (Wuchty, Jones, & Uzzi 2007)

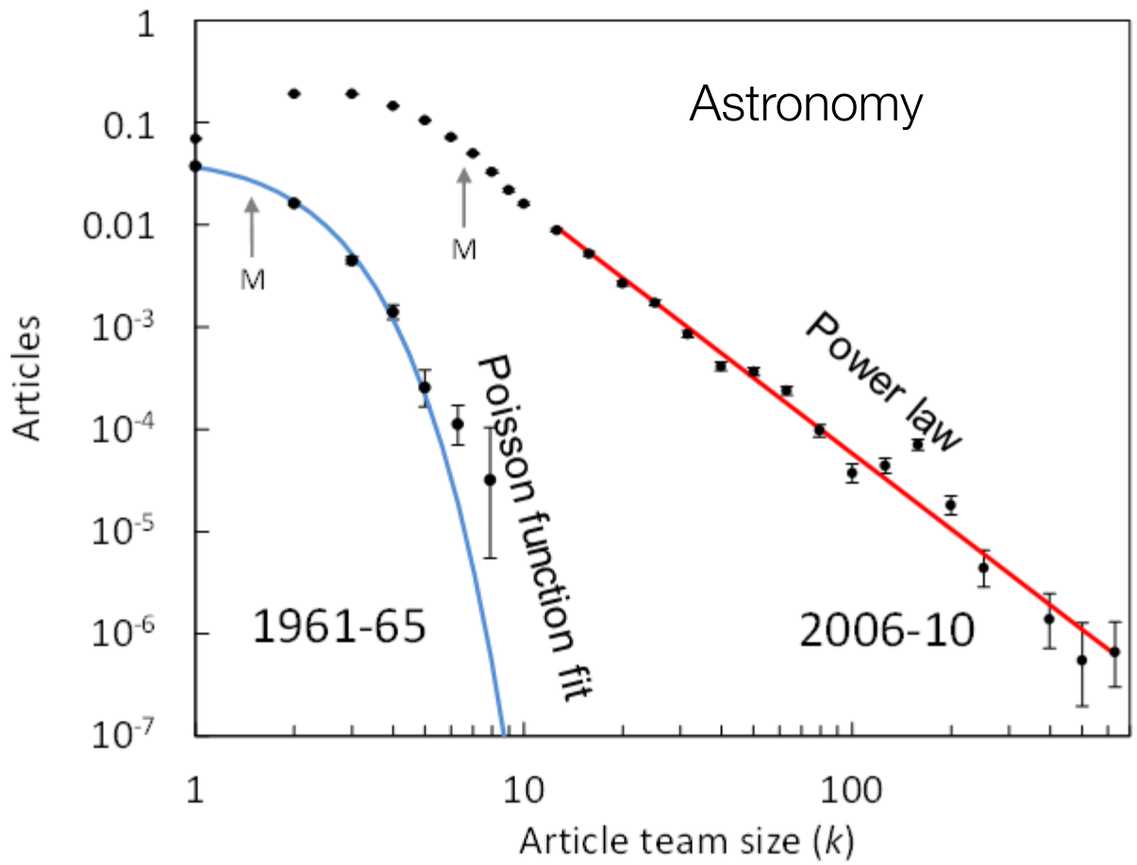
- Teams are growing in size (Bordons & Gomez 2000)

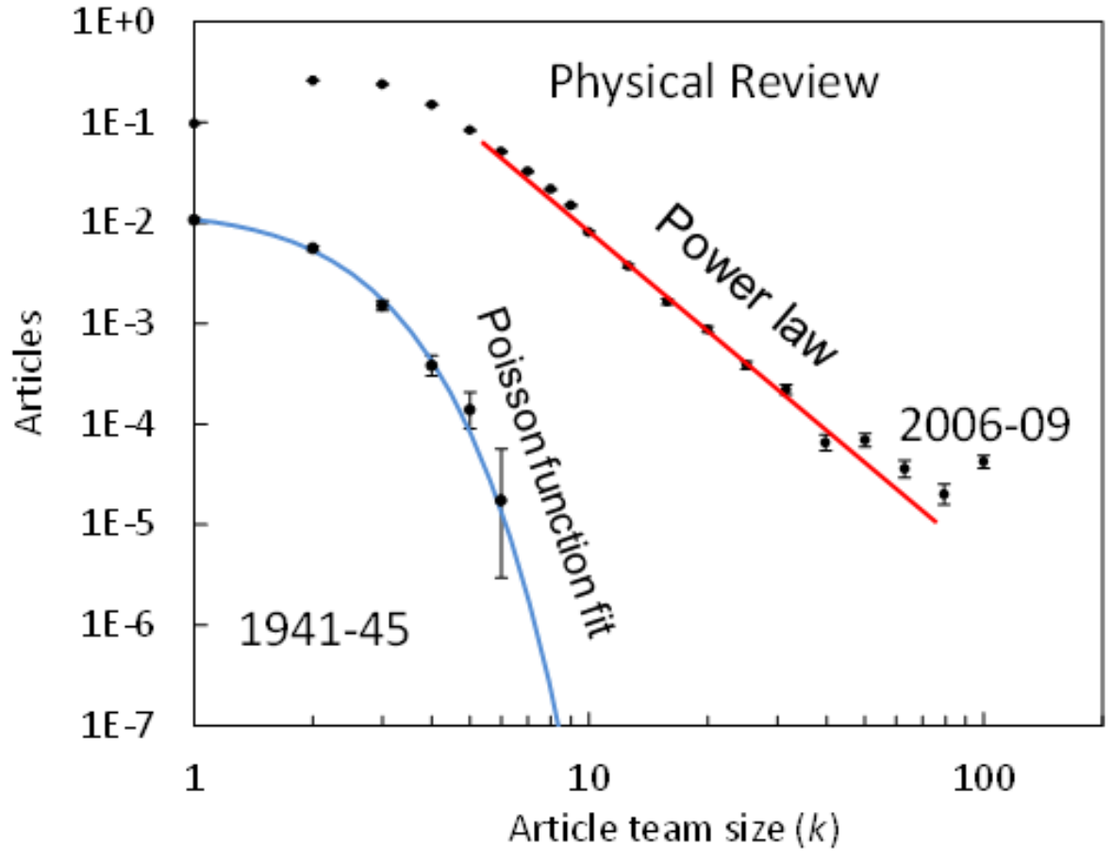
- High-impact research is increasingly attributed to large teams (Wuchty, Jones, & Uzzi 2007; Börner *et al.* 2010)

- Research that features novel ideas is attributed to large teams (Uzzi, Mukherjee, Stringer, & Jones 2013)

Change in the character of team size distribution

- Poisson function used to be a good fit, but not any more!





Team formation essentially a Poisson process, which somehow transformed

Milojević, S. (2014). *PNAS, 111*(11), 3984

Functional description of different team types

Distribution of each team type, as determined in the modeling, can be described by a corresponding analytical function

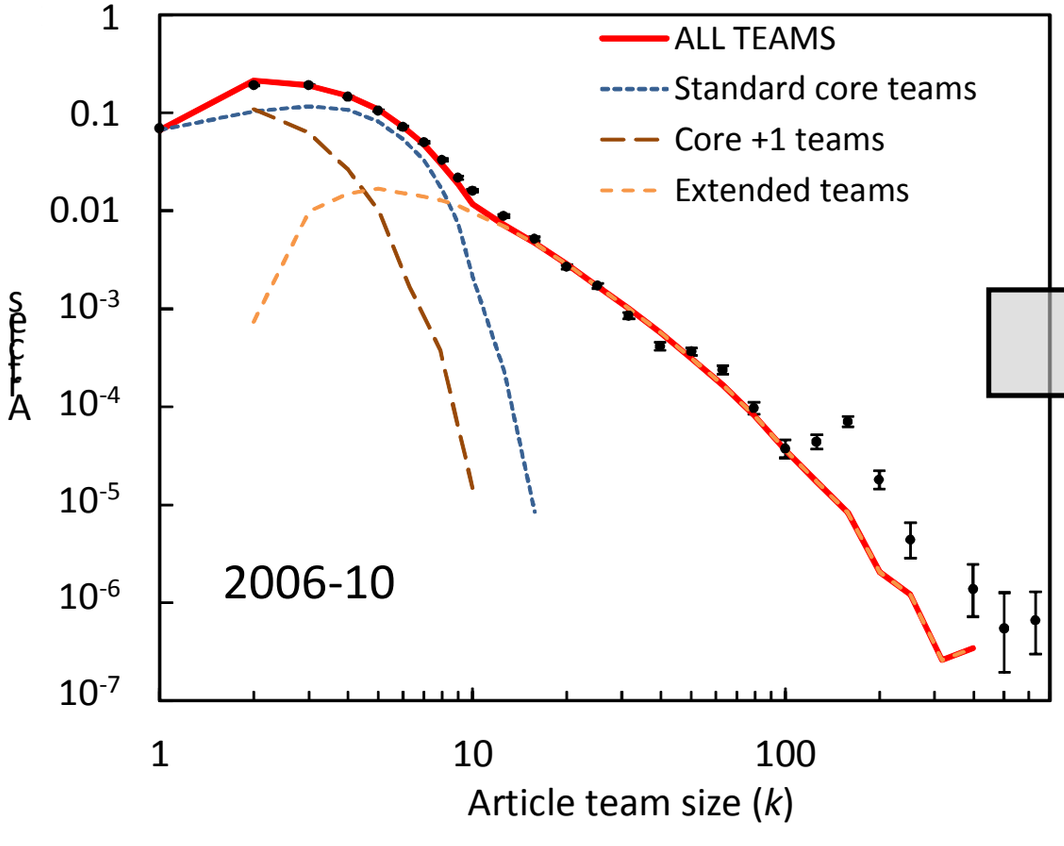
Functional forms allow the empirical distribution to be analyzed without running a model simulation

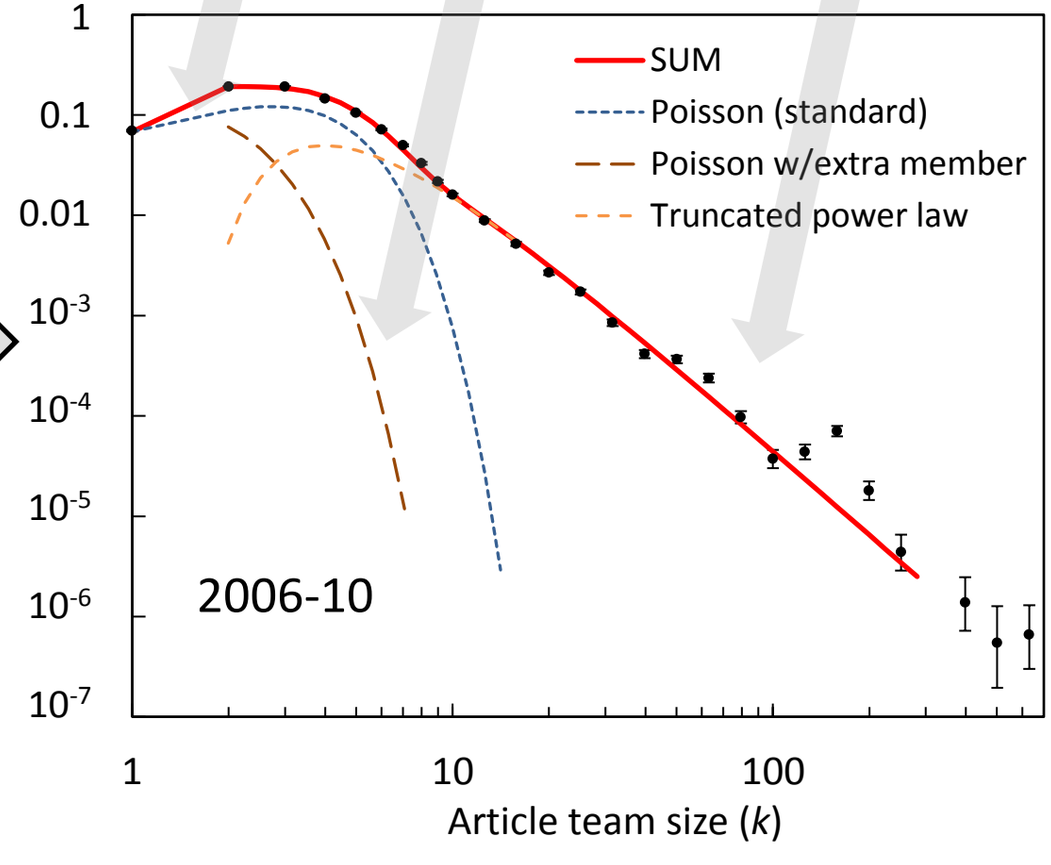
$$F(k) = n_1 \frac{\lambda_1^k e^{-\lambda_1}}{k!} + n_2 \frac{\lambda_2^{k-1} e^{-\lambda_2}}{(k-1)!} + n_3 e^{-\beta/(k-1)} k^{-\alpha}$$

POISSON

POISSON+1 (shifted)

TRUNCATED POWER LAW

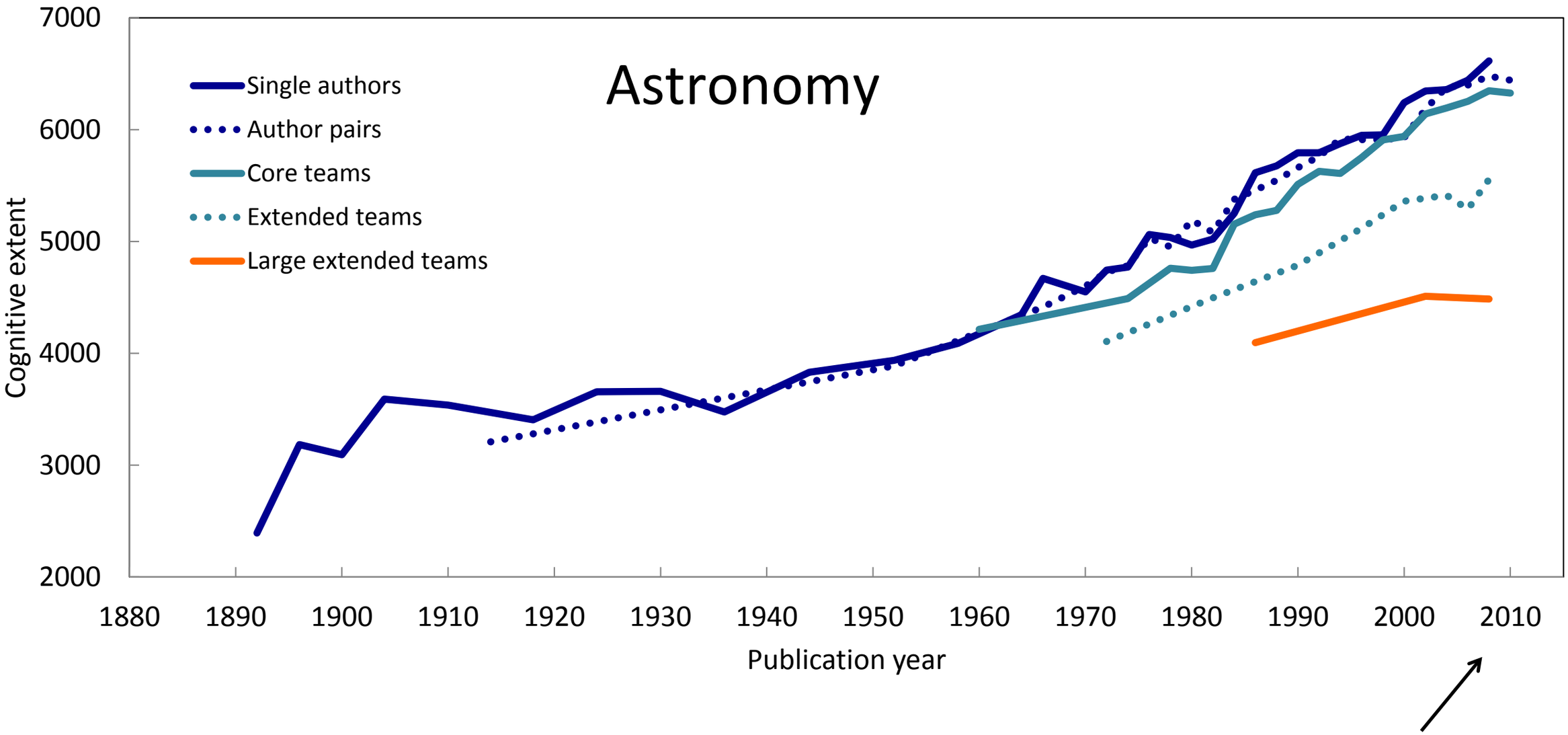




Milojević, S. (2014). *PNAS, 111*(11), 3984

Cognitive extent covered by different types of teams

- Large teams do net (yet) cover the entire field

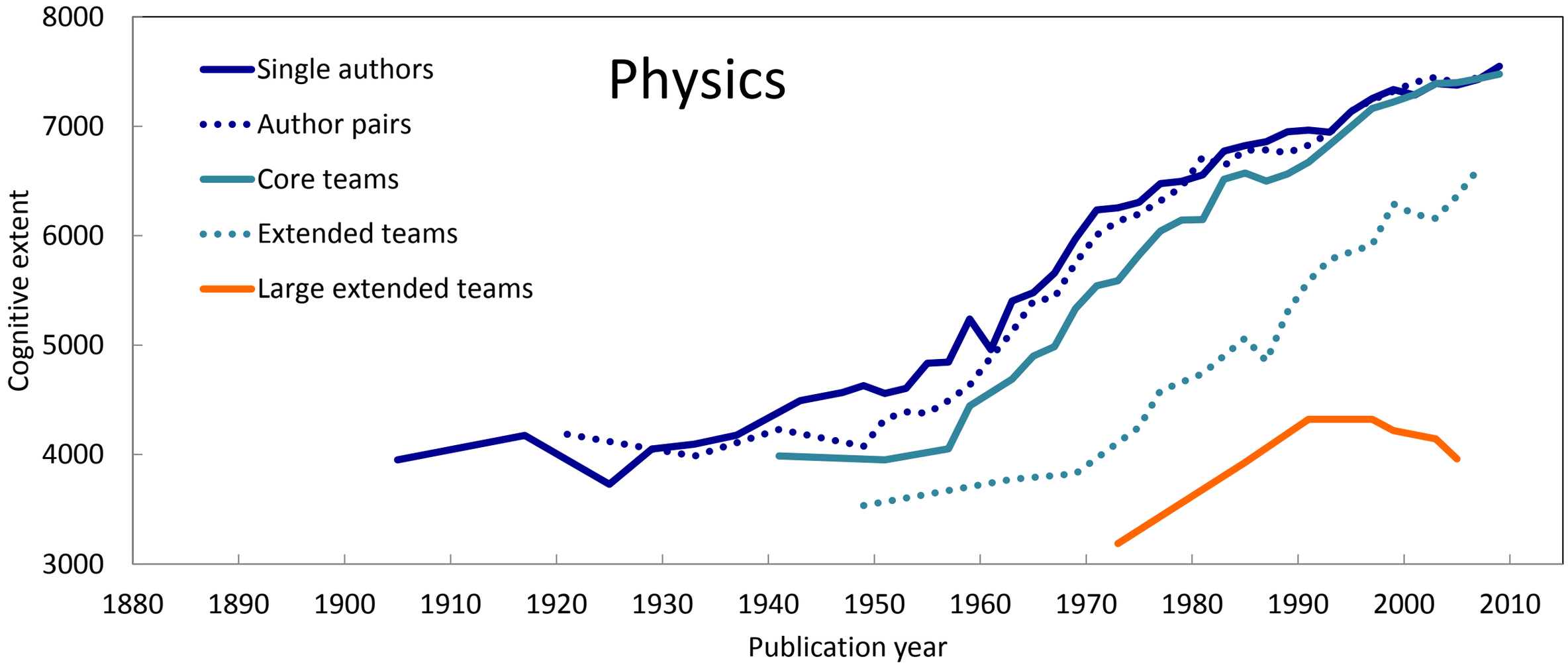


Currently: core (3-9 authors), extended (10-20), large extended (>20)

Milojević, S. (in prep.)

Cognitive extent covered by different types of teams

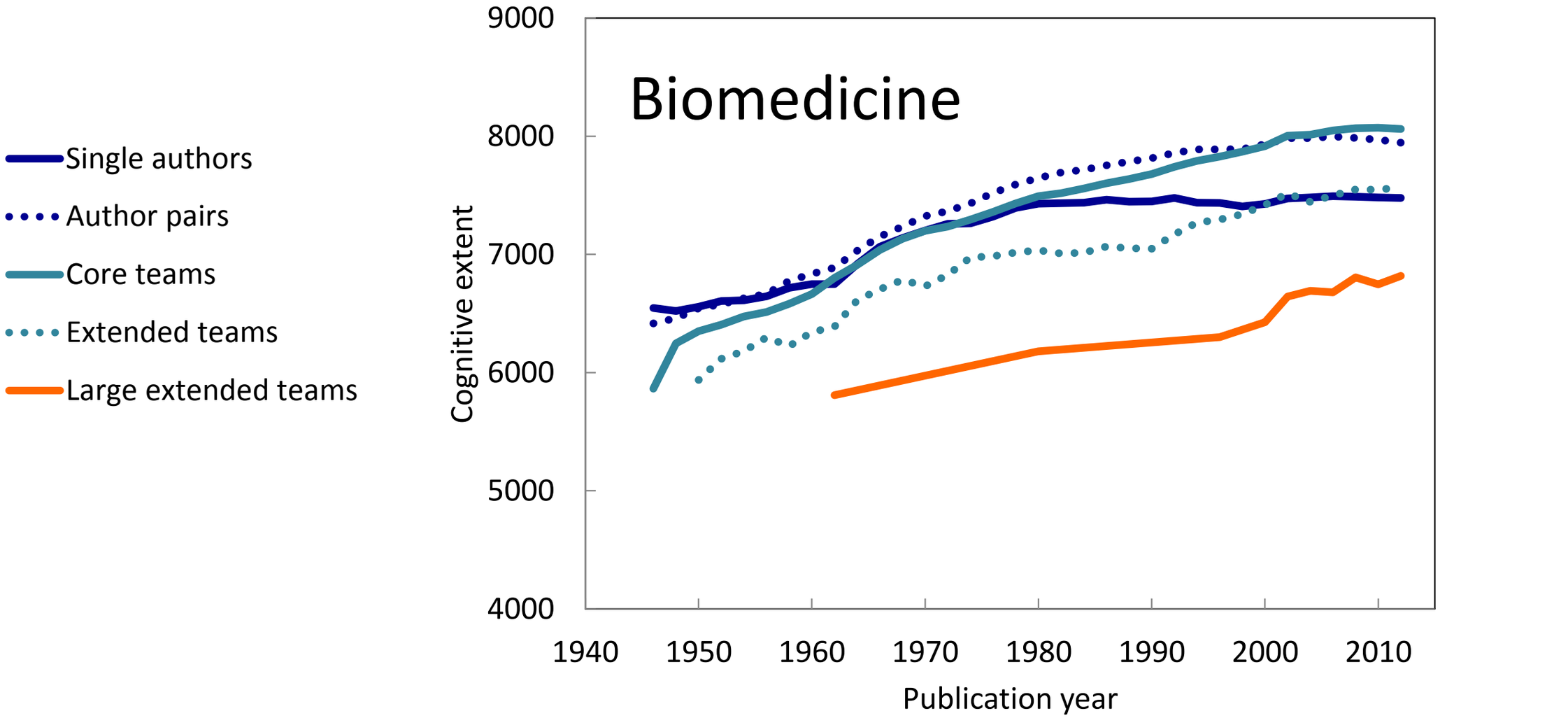
- Large teams do net (yet) cover the entire field



Milojević, S. (in prep.)

Cognitive extent covered by different types of teams

- Single authors cover less than small teams



Milojević, S. (in prep.)

Thank you!

Staša Milojević
School of Informatics and Computing
Indiana University Bloomington
smilojev@indiana.edu