

Slides at:
<https://...>

@dakotasmurray

Measuring disagreement in science

Dakota Murray, Wout Lamers, Kevin Boyack,
Vincent Larivière, Cassidy R. Sugimoto,
Nees Jan van Eck, Ludo Waltman



Disagreement in science

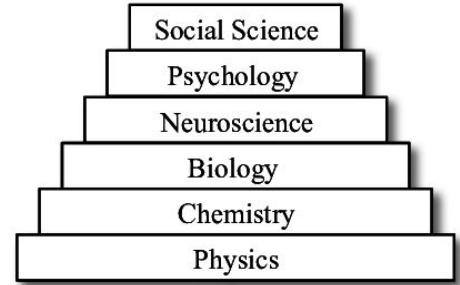
Disagreement in science

- Dialectic debate in early history



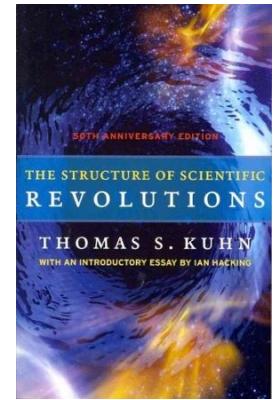
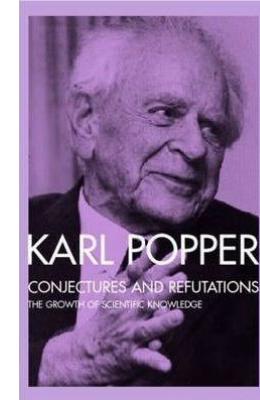
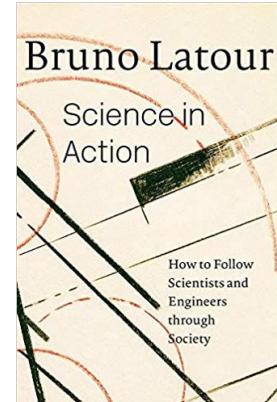
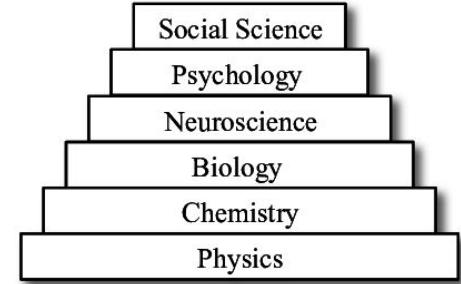
Disagreement in science

- Dialectic debate in early history
- Consensus in Comte's hierarchy of sciences



Disagreement in science

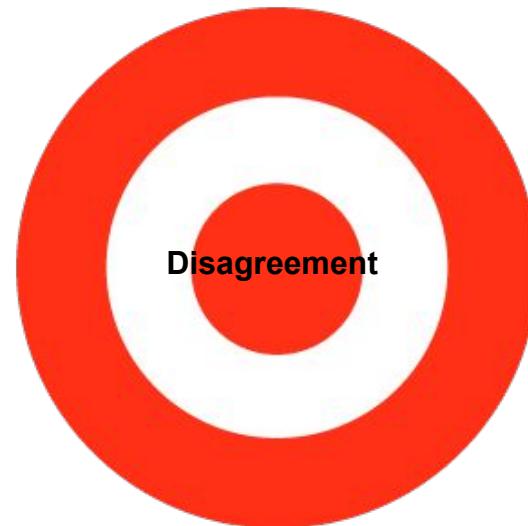
- Dialectic debate in early history
- Consensus in Comte's hierarchy of sciences
- Popper's falsification,
- Kuhn's anomalies
- Latour's controversies
- more



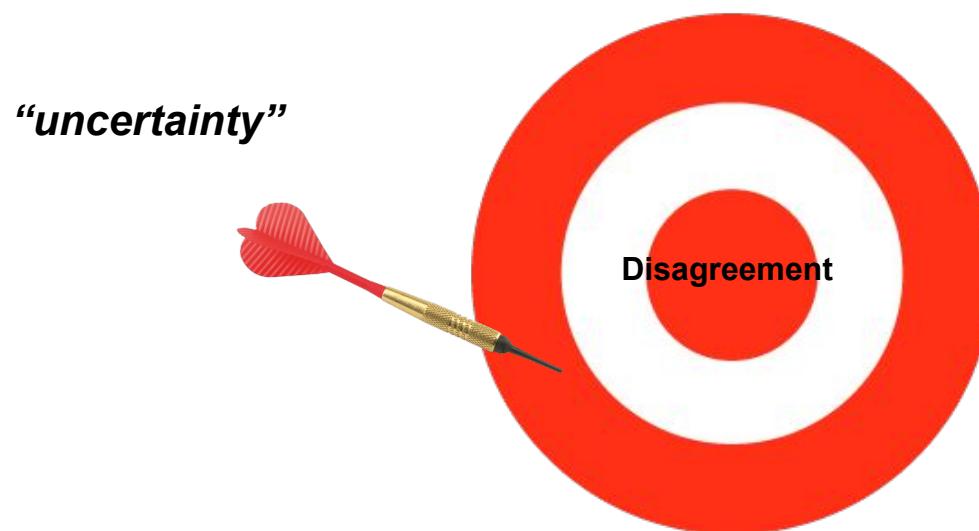
Disagreement is integral
to science.

Disagreement is integral
to science...but where
and when does it
happen?

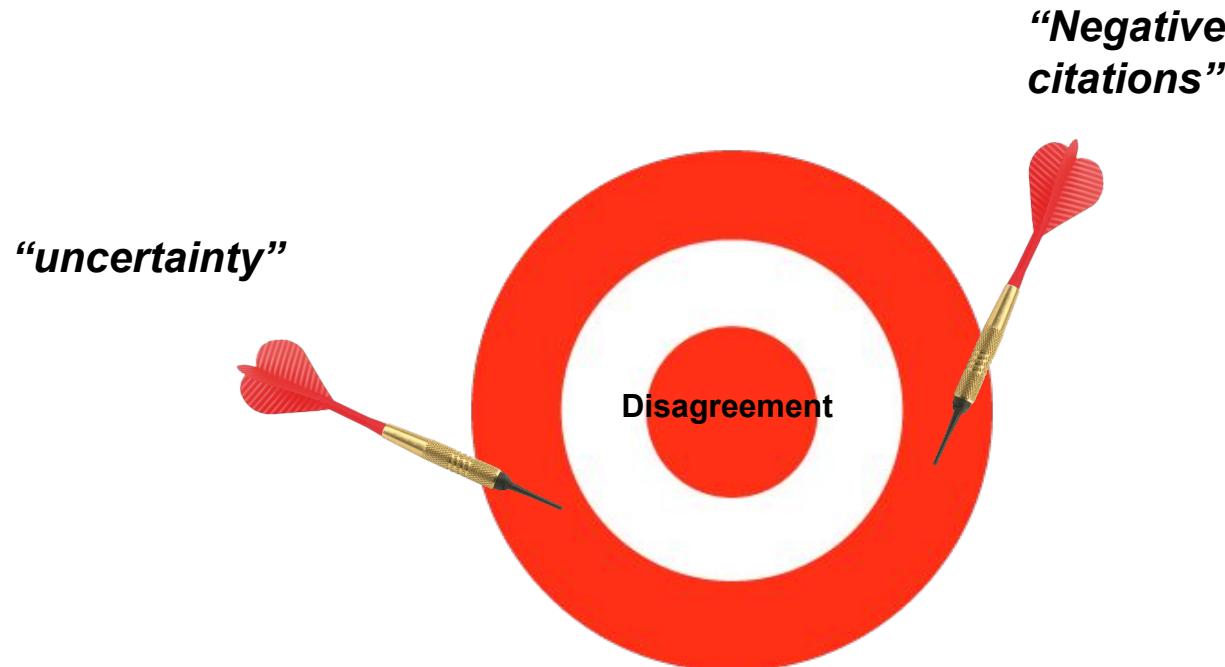
Past work?



Past work?



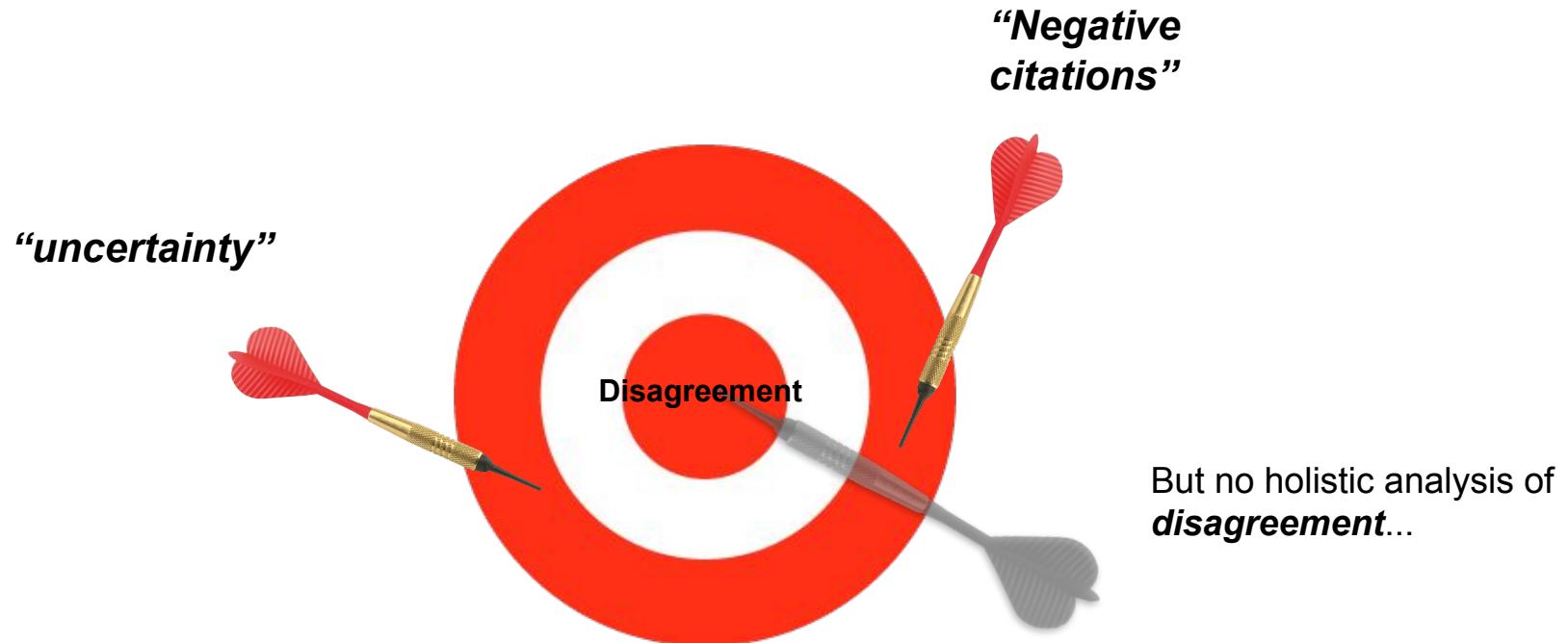
Past work?



Chen, C., Song, M., & Heo, G. E. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *JOI*

Catalini, C., Lacetera, N., & Oettl, A. (2015). The incidence and role of negative citations in science. *PNAS*

Past work?



Chen, C., Song, M., & Heo, G. E. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *JOI*

Catalini, C., Lacetera, N., & Oettl, A. (2015). The incidence and role of negative citations in science. *PNAS*

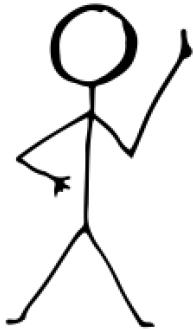
Goal:

Develop a methodological framework to quantify the degree of *disagreement* (*controversy*, *disconsensus*, and *dissonance...*) in scientific literature

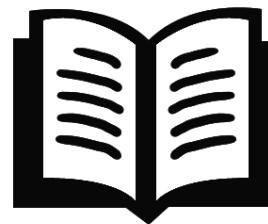
Operationalizing disagreement

Our focus is *explicit, interpersonal* disagreement

*Author of
cited paper*



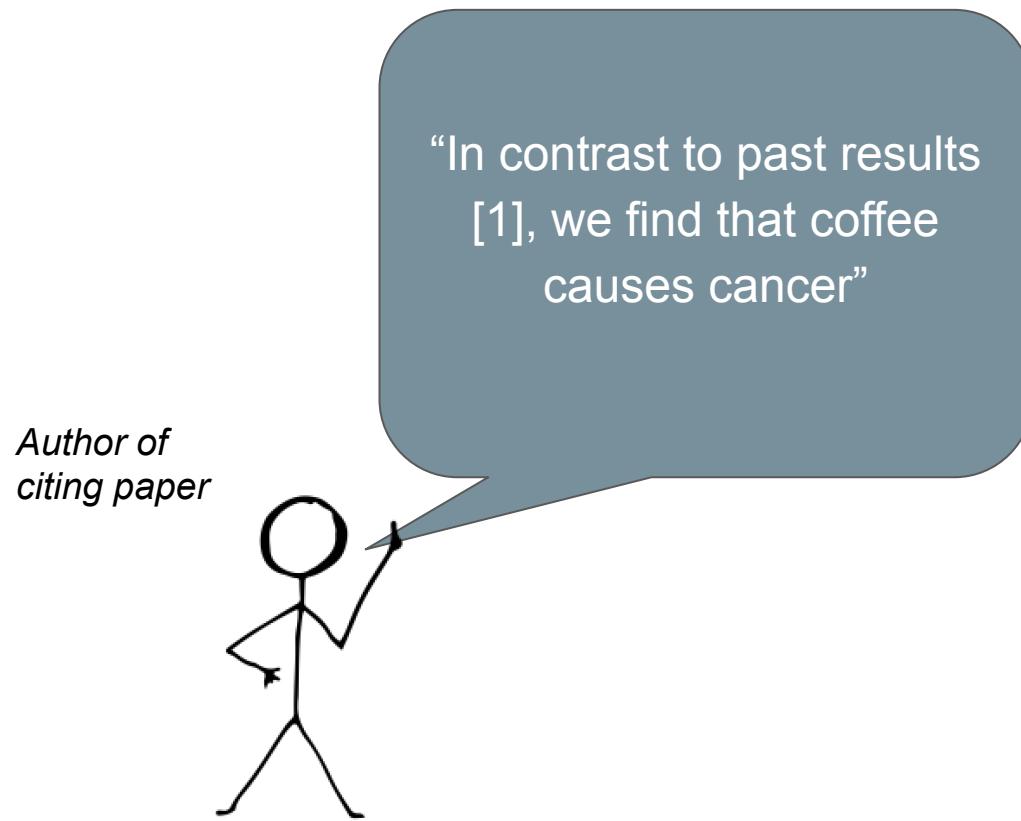
*Some cited
paper*



Cites and explicitly disagrees with



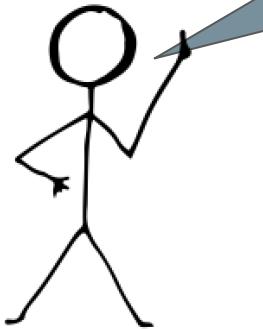
General disagreement: direct contradiction



General disagreement: different model assumptions

“The model used by Kim et al [2] predicts coffee drinker’s life expectancy to be 80 years, whereas our model predicts 85 years

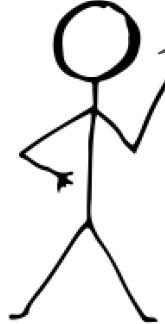
*Author of
citing paper*



Disagreement in the literature

“There is disagreement in past studies over whether coffee does or does not causes cancer [3-10]”

*Author of
citing paper*



$$X = y$$

$$X = 2y$$

$$X \neq y$$

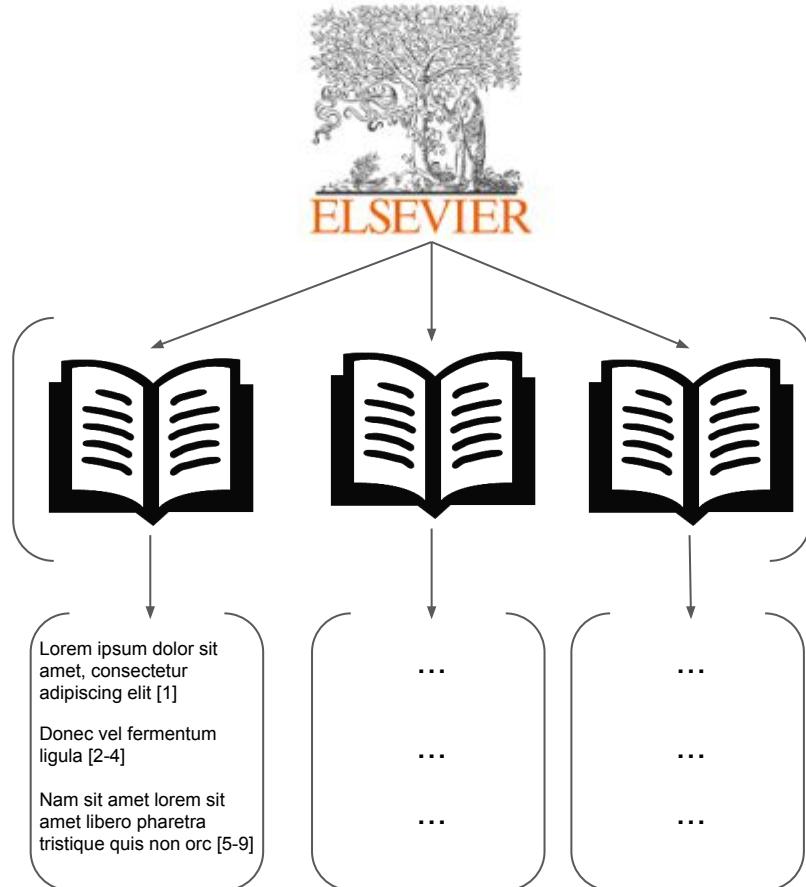
$$X \sim y$$



Approach

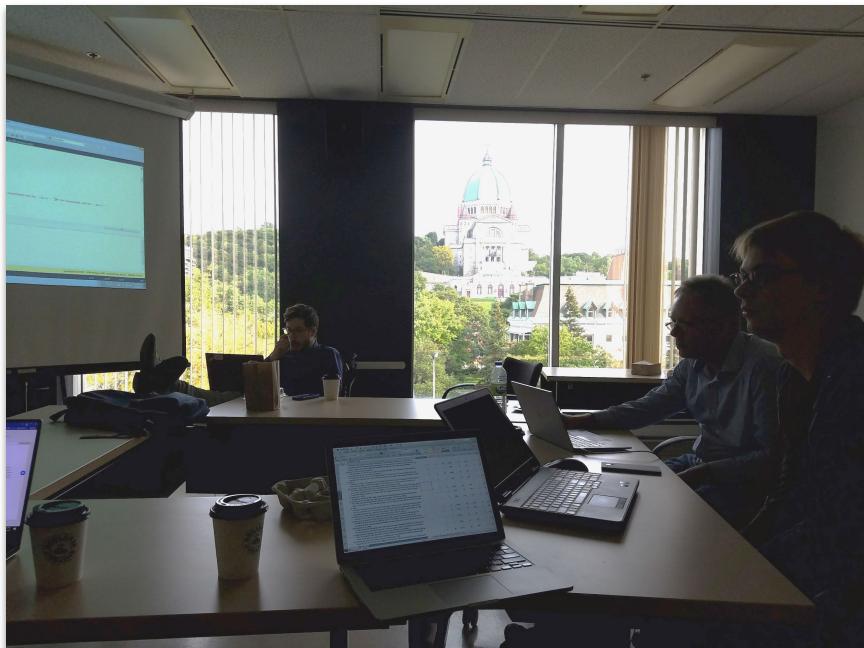
Data

- Our interest is *interpersonal* disagreement, and so we focus on in-text citations
- Over 3 million full text articles from english articles published in ScienceDirect between 1998 to 2015
- Citances extracted



How can we precisely and
reliably identify a citation
sentence as disagreement?

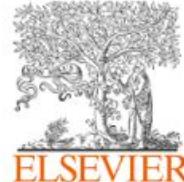
A long, iterative process of identifying, validating, and refining queries



---- Signal terms ----

Challenge*
Conflict*
Contradict*
Contrary
Contrast*
Contravers*
Debat*
Differ*
Disagree*
Disprov*
No consensus
Questionable*
Refut*

ScienceDirect



---- Signal terms ----

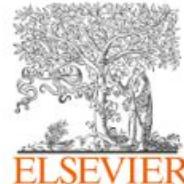
Challenge*
Conflict*
Contradict*
Contrary
Contrast*
Contravers*
Debat*
Differ*
Disagree*
Disprov*
No consensus
Questionable*
Refut*

query

retrieve
sentences

"...recruiting participants was challenging..."

ScienceDirect



---- Signal terms ----

Challenge*
Conflict*
Contradict*
Contrary
Contrast*
Contravers*
Debat*
Differ*
Disagree*
Disprov*
No consensus
Questionable*
Refut*

query

retrieve
sentences

"...recruiting participants was challenging..." **X Invalid**

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

studies; previous
works; earlier works;
literature, analyses;
reports

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

“...recruiting participants was
challenging...”

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

“...recruiting participants was **challenging**...”

“However, **recent studies** have **disagreed** with this approach”

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

“...recruiting participants was challenging...”

“However, recent studies have disagreed with this approach”

“This theory have since been disproven”

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

“...recruiting participants was challenging...”

“However, recent studies have disagreed with this approach”

“This theory have since been disproven”

“These ideas were influenced by Popper’s *Conjectures and Refutations*”

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

“...recruiting participants was challenging...” **X Invalid**

“However, recent studies have disagreed with this approach”
✓ Valid

“This theory have since been disproven” **✓ Valid**

“These ideas were influenced by Popper’s *Conjectures and Refutations*” **X Invalid**

---- Filter terms ----

---- Signal terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*	?	?	?	?	?
Conflict*	?	?	?	?	?
Contradict*	?	?	?	?	?
Contrary	?	?	?	?	?
Contrast*	?	?	?	?	?
Contravers*	?	?	?	?	?
Debat*	?	?	?	?	?
Differ*	?	?	?	?	?
Disagree*	?	?	?	?	?
Disprov*	?	?	?	?	?
No consensus	?	?	?	?	?
Questionable*	?	?	?	?	?
Refut*	?	?	?	?	?

Which queries are the most valid?

---- Filter terms ----

---- Signal terms ----

	<u>_standalone_</u>	+studies	+ideas	+methods	+results
Challenge*	50 citances	50 citances	50 citances	50 citances	50 citances
Conflict*	50 citances	50 citances	50 citances	50 citances	50 citances
Contradict*	50 citances	50 citances	50 citances	50 citances	50 citances
Contrary	50 citances	50 citances	50 citances	50 citances	50 citances
Contrast*	50 citances	50 citances	50 citances	50 citances	50 citances
Contravers*	50 citances	50 citances	50 citances	50 citances	50 citances
Debat*	50 citances	50 citances	50 citances	50 citances	50 citances
Differ*	50 citances	50 citances	50 citances	50 citances	50 citances
Disagree*	50 citances	50 citances	50 citances	50 citances	50 citances
Disprov*	50 citances	50 citances	50 citances	50 citances	50 citances
No consensus	50 citances	50 citances	50 citances	50 citances	50 citances
Questionable*	50 citances	50 citances	50 citances	50 citances	50 citances
Refut*	50 citances	50 citances	50 citances	50 citances	50 citances

Which queries are the most valid?

For each combination, query a random sample of 50 citances

Two coders manually identify each citance as

✓ Valid OR ✗ Invalid

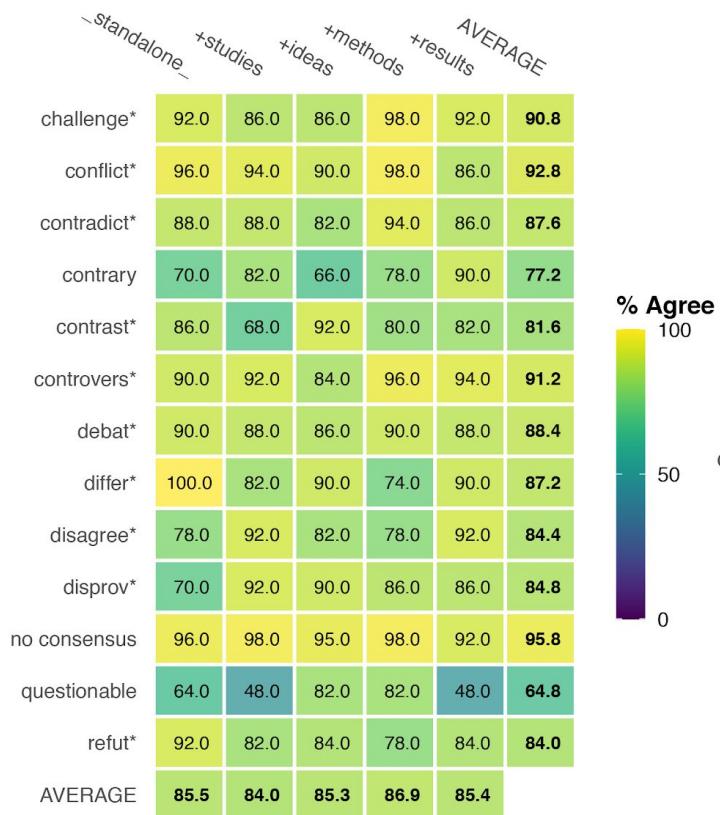
Calculate % agreement and % validity

% agreement

Coder 1: ✓ Valid Coder 2: ✓ Valid	Coder 1: ✓ Valid Coder 2: ✗ Invalid
Coder 1: ✗ Invalid Coder 2: ✓ Valid	Coder 1: ✗ Invalid Coder 2: ✗ Invalid

% agreement

Coder 1: ✓ Valid	Coder 1: ✓ Valid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid
Coder 1: ✗ Invalid	Coder 1: ✗ Invalid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid

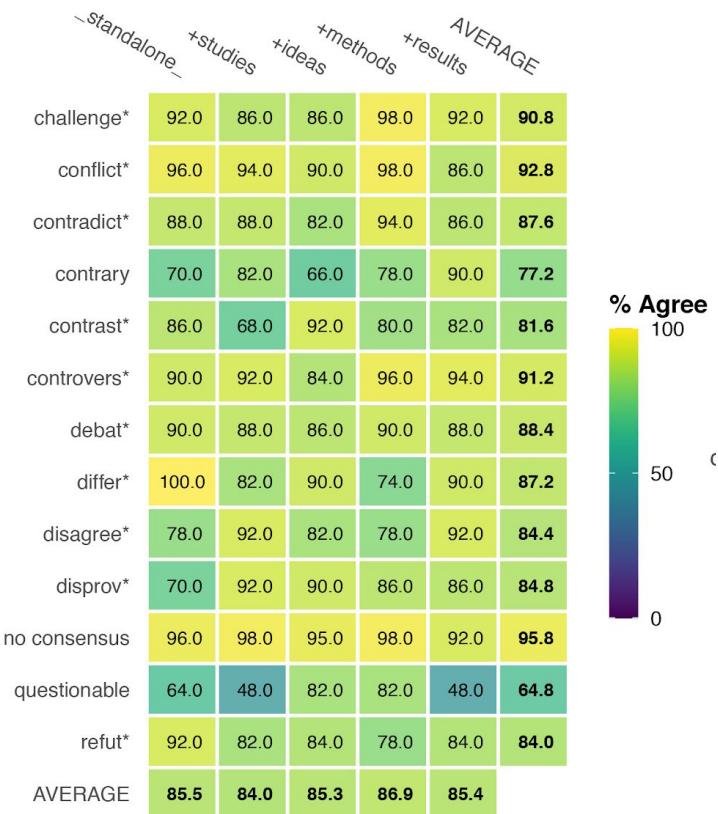


% agreement

Coder 1: ✓ Valid	Coder 1: ✓ Valid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid
Coder 1: ✗ Invalid	Coder 1: ✗ Invalid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid

Overall, 85% agreement

This demonstrates that our coding system was robust

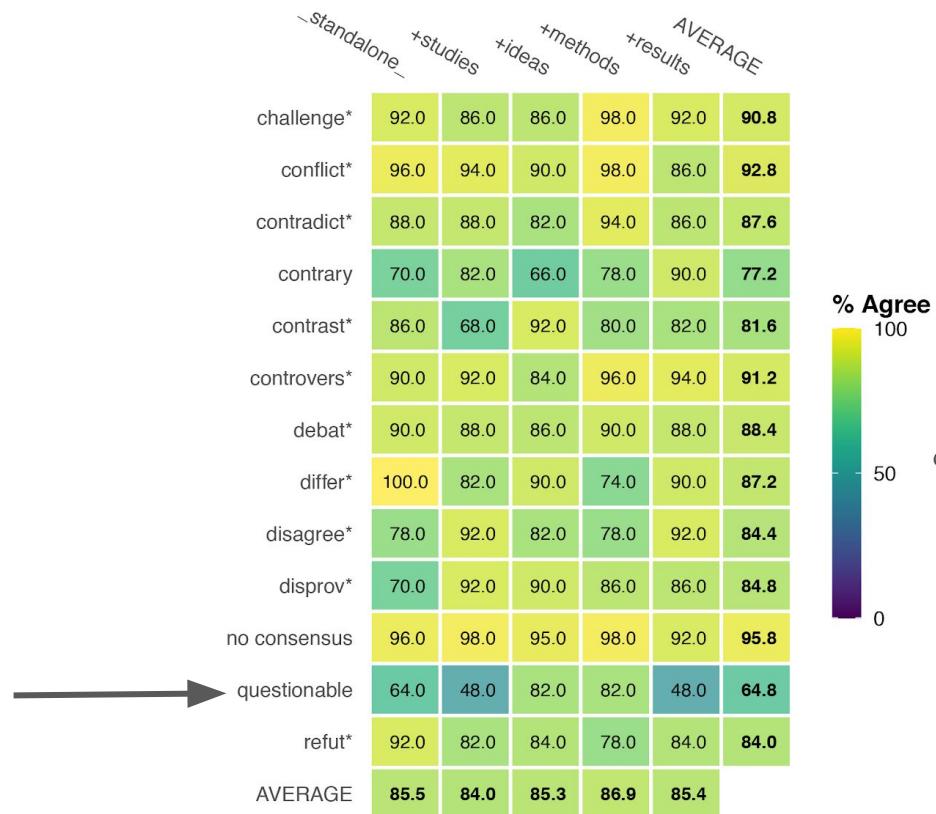


% agreement

Coder 1: ✓ Valid Coder 2: ✓ Valid	Coder 1: ✓ Valid Coder 2: ✗ Invalid
Coder 1: ✗ Invalid Coder 2: ✓ Valid	Coder 1: ✗ Invalid Coder 2: ✗ Invalid

Overall, 85% agreement

This demonstrates that our coding system was robust

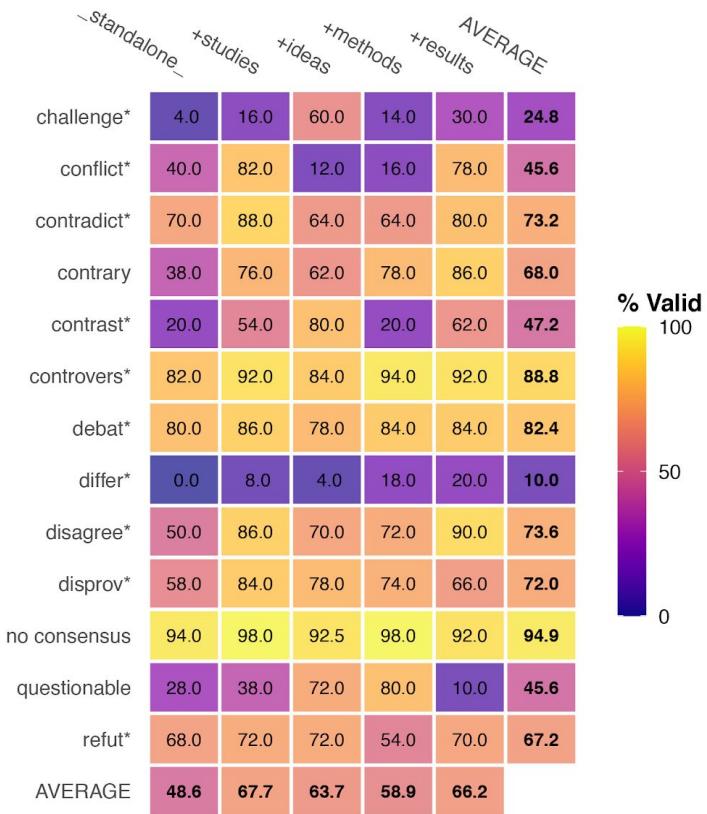


% Validity

<u>Coder 1: ✓ Valid</u> <u>Coder 2: ✓ Valid</u>	Coder 1: ✓ Valid Coder 2: ✗ Invalid
Coder 1: ✗ Invalid Coder 2: ✓ Valid	Coder 1: ✗ Invalid Coder 2: ✗ Invalid

% Validity

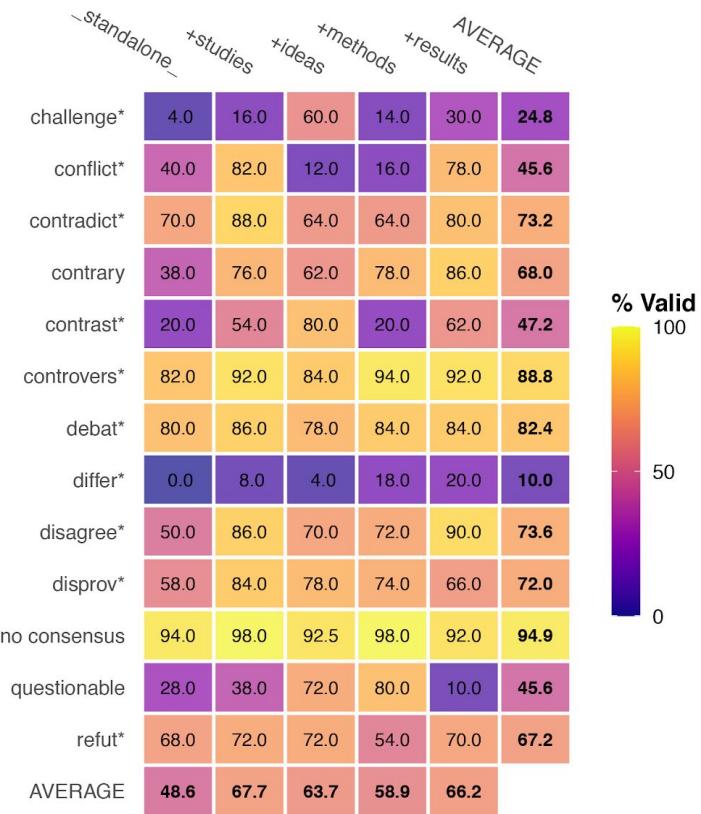
Coder 1: ✓ Valid	Coder 1: ✓ Valid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid
Coder 1: ✗ Invalid	Coder 1: ✗ Invalid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid



% Validity

<u>Coder 1: ✓ Valid</u>	Coder 1: ✓ Valid
<u>Coder 2: ✓ Valid</u>	Coder 2: ✗ Invalid
Coder 1: ✗ Invalid	Coder 1: ✗ Invalid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid

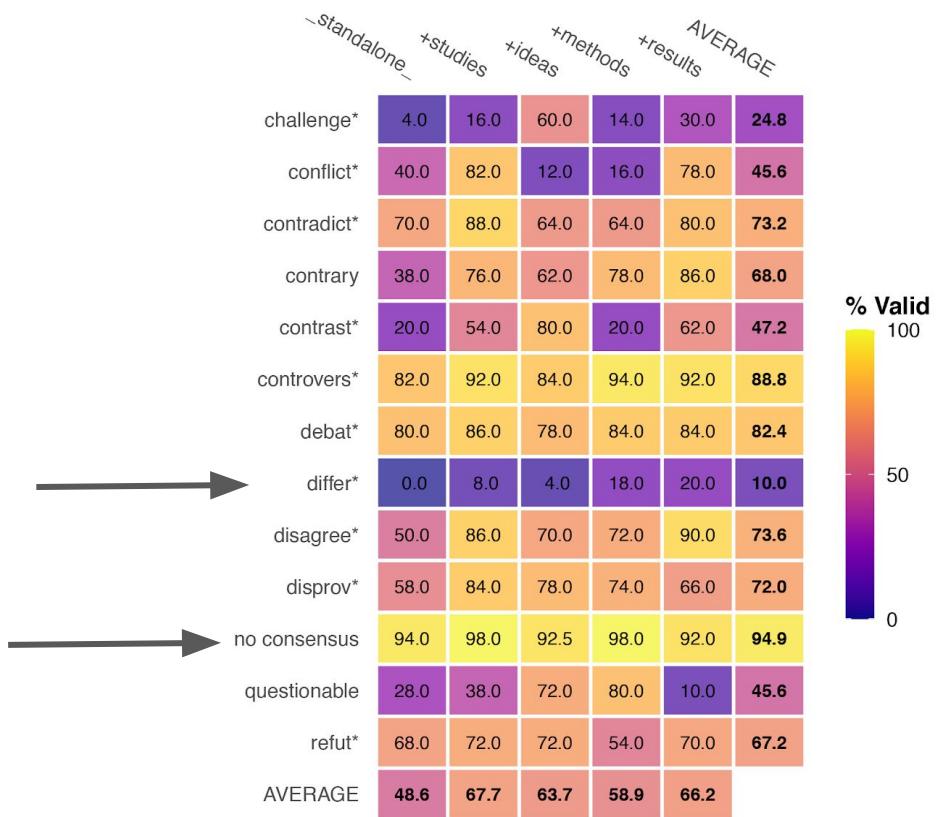
Lots of heterogeneity



% Validity

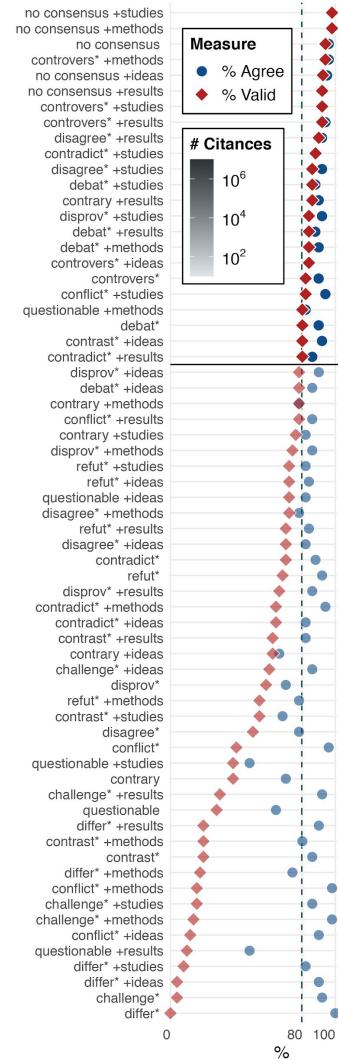
Coder 1: ✓ Valid	Coder 1: ✓ Valid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid
Coder 1: ✗ Invalid	Coder 1: ✗ Invalid
Coder 2: ✓ Valid	Coder 2: ✗ Invalid

Lots of heterogeneity



Both

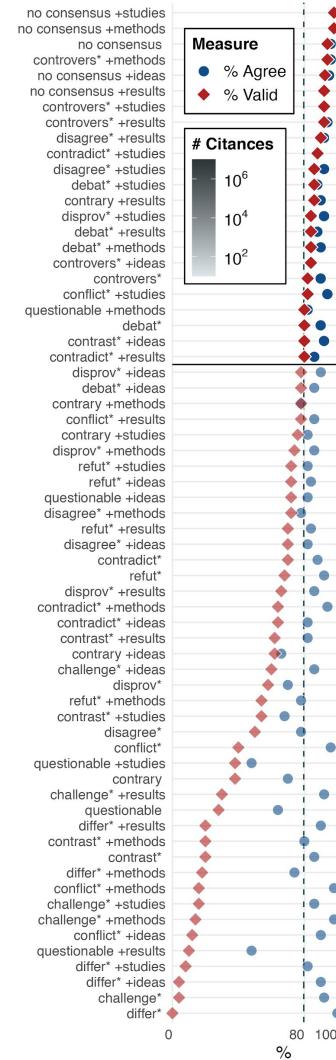
Ordered by Validity



Both

Ordered
by Validity

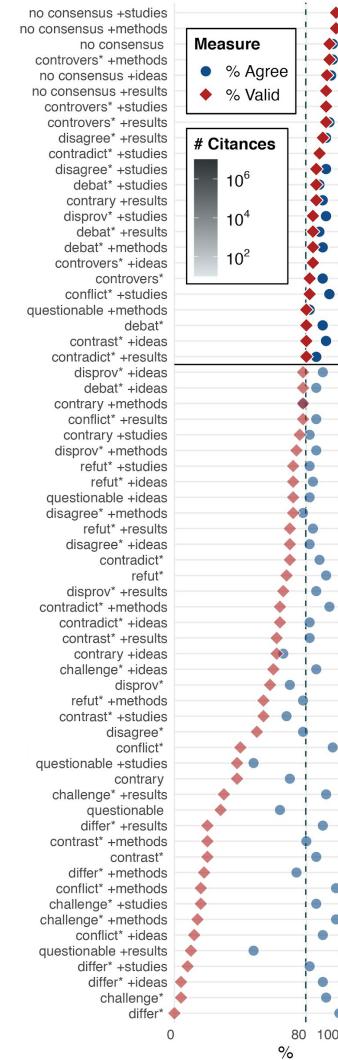
- % Agreement stable, %valid varied widely



Both

Ordered
by Validity

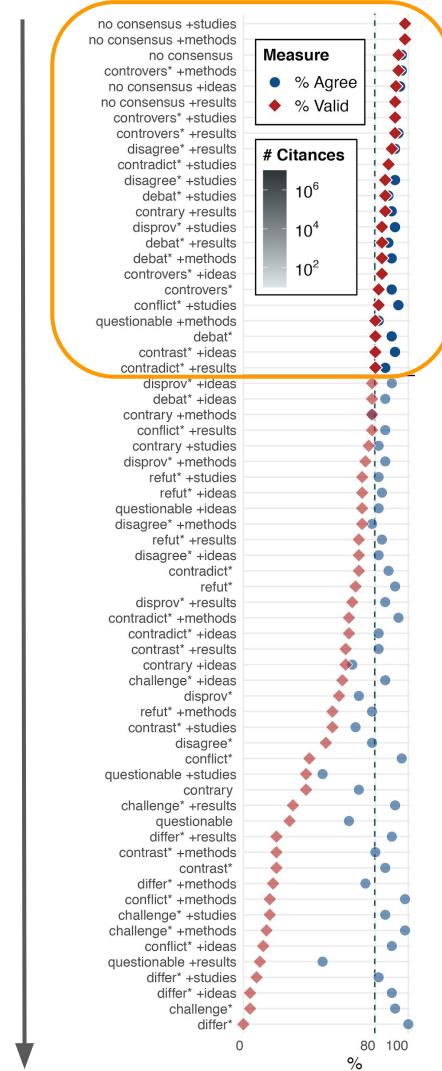
- % Agreement stable, %valid varied widely
- We don't want obviously bad queries



Both

Ordered
by Validity

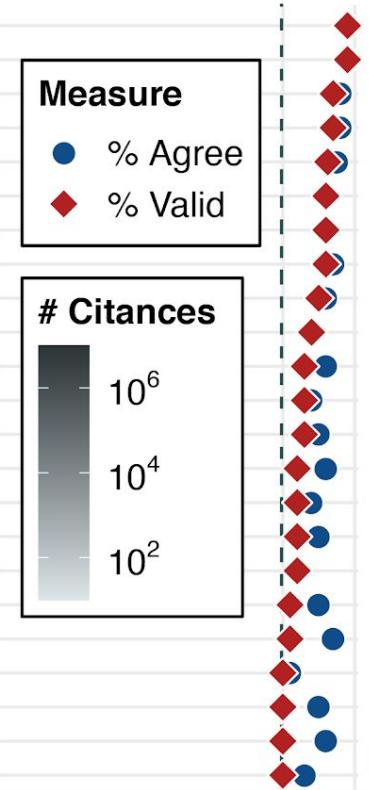
- % Agreement stable, %valid varied widely
- We don't want obviously bad queries
- Set a validity threshold—80%



Both

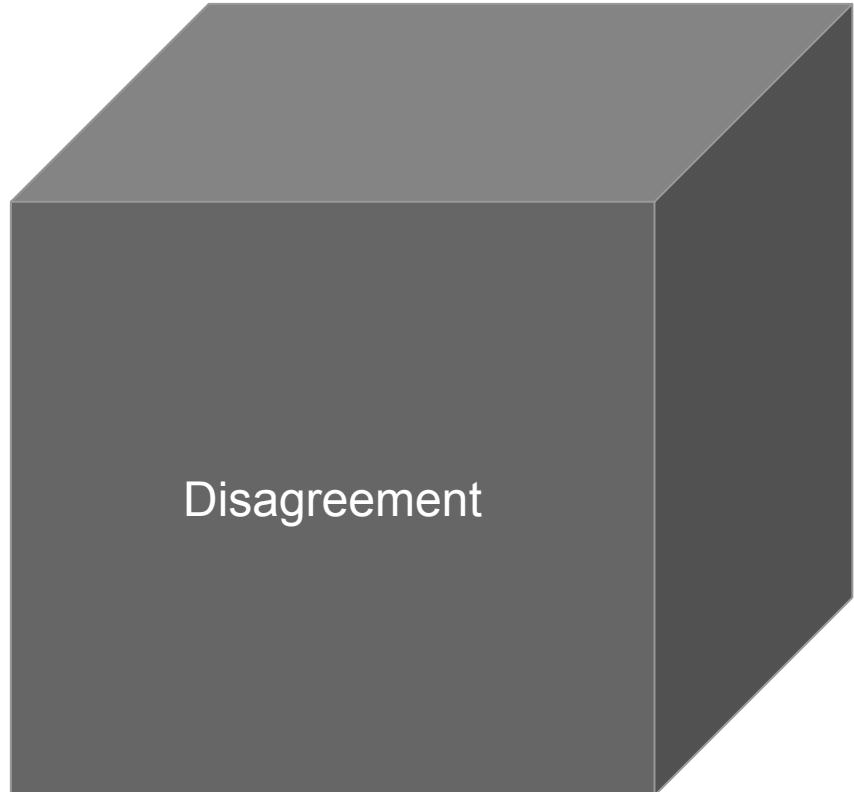
- % Agreement stable, %valid varied widely
- We don't want obviously bad queries
- Set a validity threshold—80%
- 23 queries accounting for ~450,000 citation sentences.

no consensus +studies
no consensus +methods
no consensus
controvers* +methods
no consensus +ideas
no consensus +results
controvers* +studies
controvers* +results
disagree* +results
contradict* +studies
disagree* +studies
debat* +studies
contrary +results
disprov* +studies
debat* +results
debat* +methods
controvers* +ideas
controvers*
conflict* +studies
questionable +methods
debat*
contrast* +ideas
contradict* +results



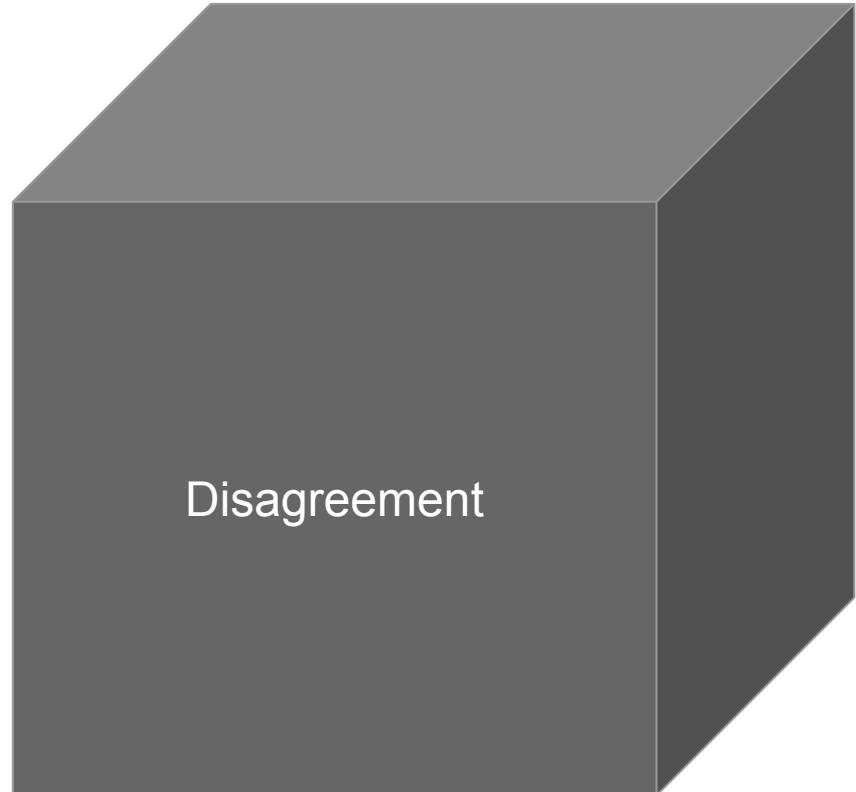
Both

- % Agreement stable, %valid varied widely
- We don't want obviously bad queries
- Set a validity threshold—80%
- 23 queries accounting for ~450,000 citation sentences. **Non-exhaustive, but precise!**

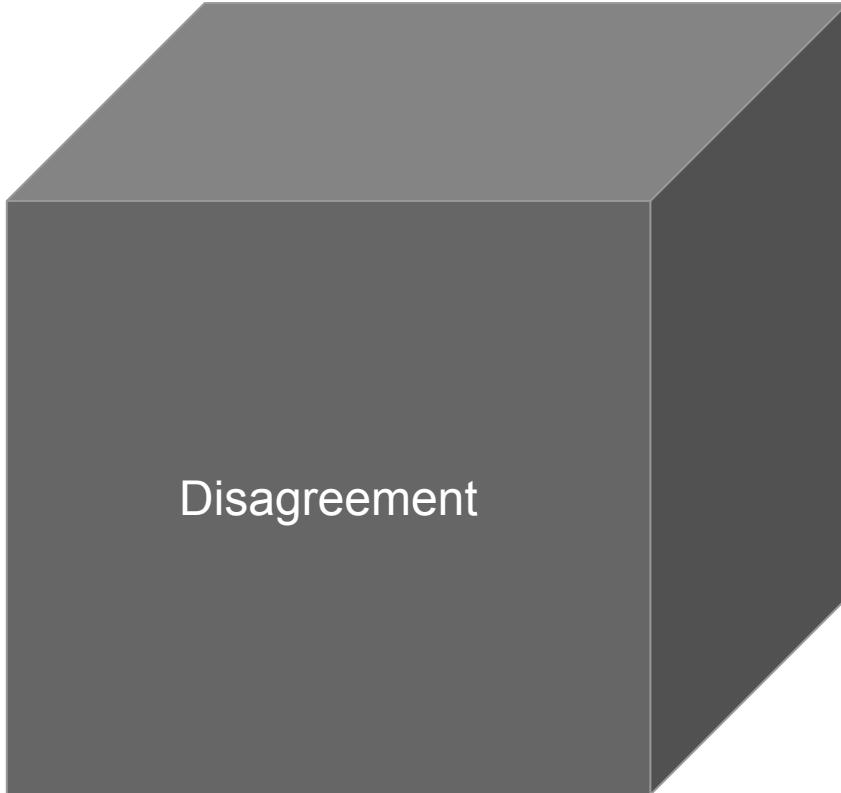


Both

- % Agreement stable, %valid varied widely
- We don't want obviously bad queries
- Set a validity threshold—80%
- 23 queries accounting for ~450,000 citation sentences. **Non-exhaustive, but precise!**



Assess disagreement based on...

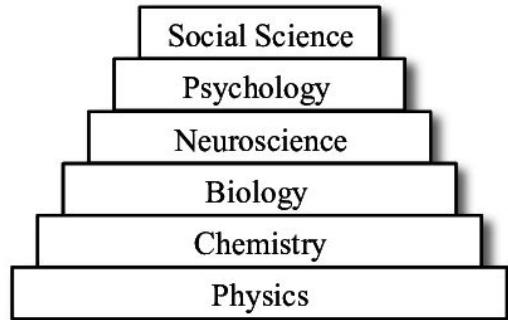


- Discipline
- Over time
- Age of cited paper
- Self Citation

Measuring disagreement *by discipline*

Disagreement by field

Auguste Comte's
*Hierarchy of
sciences*

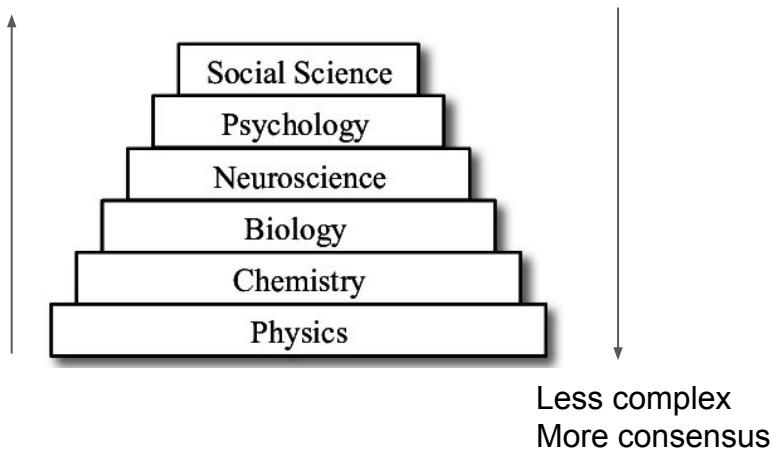


Disagreement by field

Auguste Comte's
*Hierarchy of
sciences*



More complex,
Less consensus

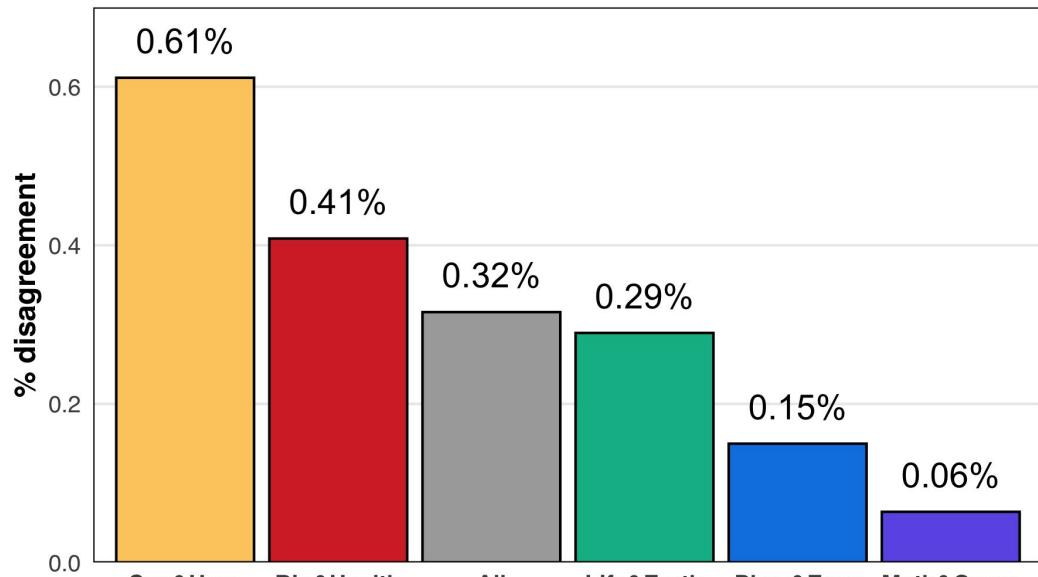
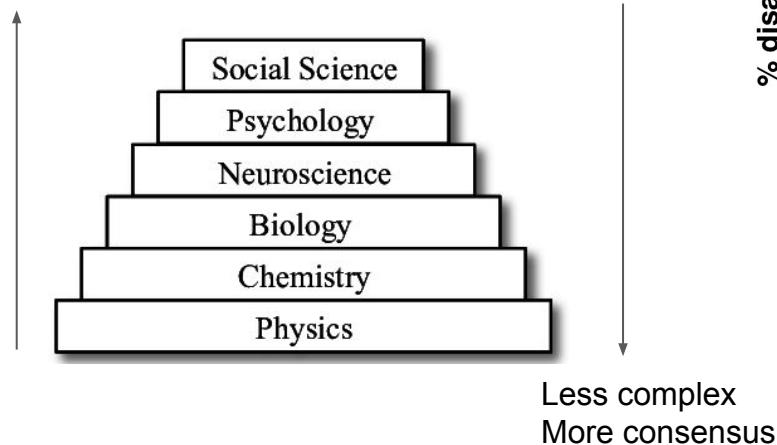


Disagreement by field

Auguste Comte's
*Hierarchy of
sciences*



More complex,
Less consensus

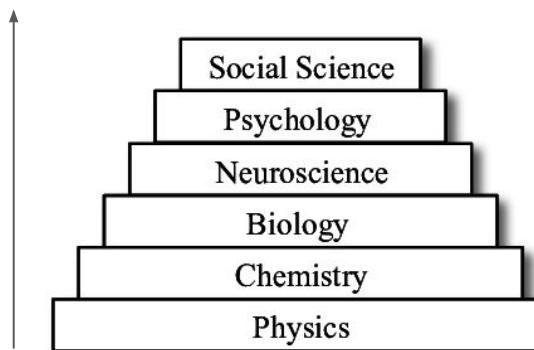


Disagreement by field

Auguste Comte's
*Hierarchy of
sciences*

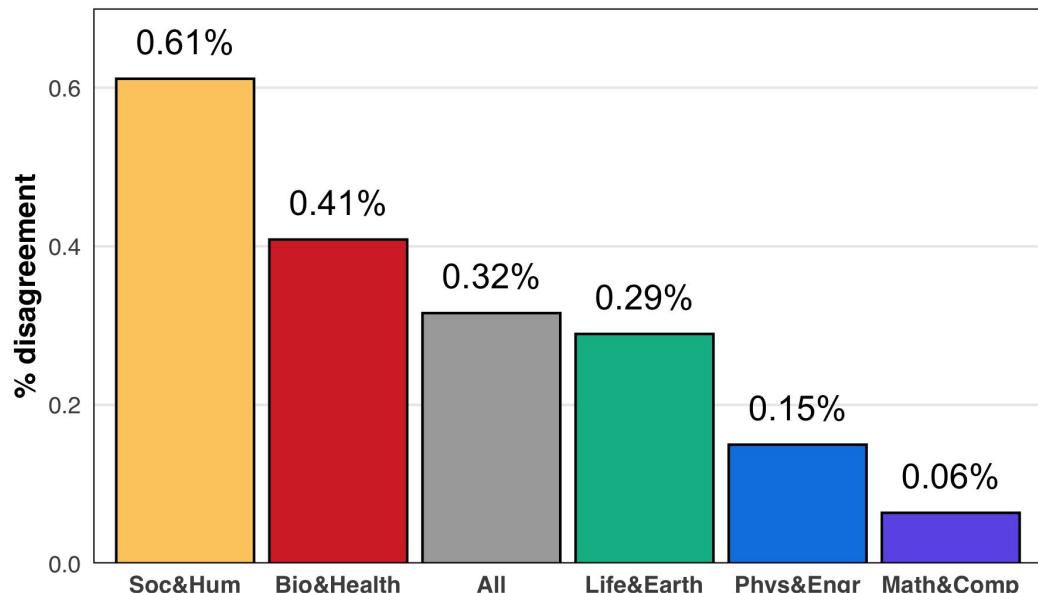


More complex,
Less consensus



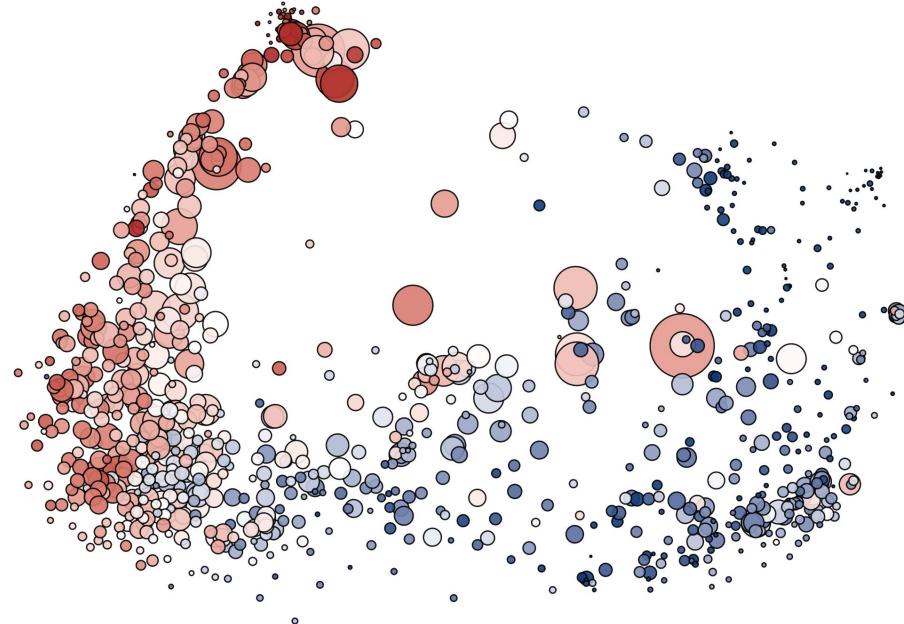
Less complex
More consensus

Too coarse-grained?



Digging deeper

Meso-level fields (clusters of papers)

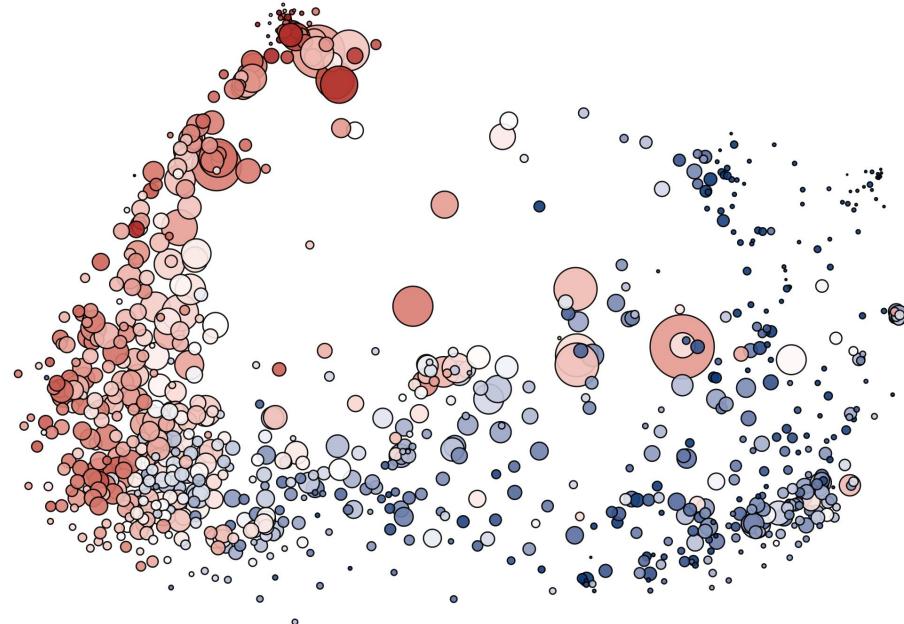


Digging deeper

Meso-level fields (clusters of papers)

Area maps to total citation sentences

Distance reflects relatedness



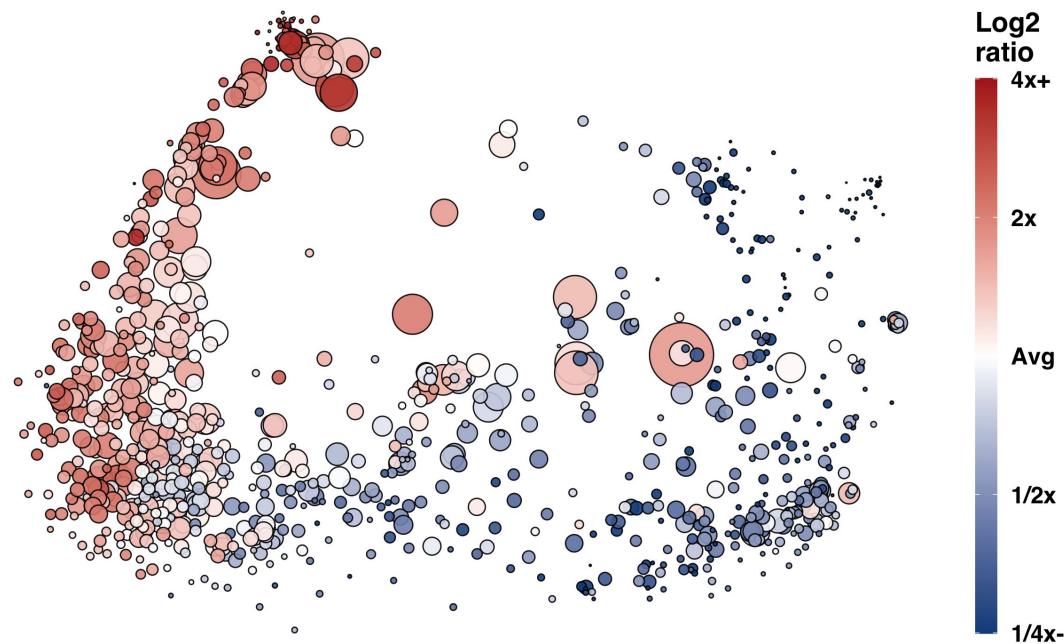
Digging deeper

Meso-level fields (clusters of papers)

Area maps to total citation sentences

Distance reflects relatedness

Color indicates ratio of %
disagreement to the average



Digging deeper

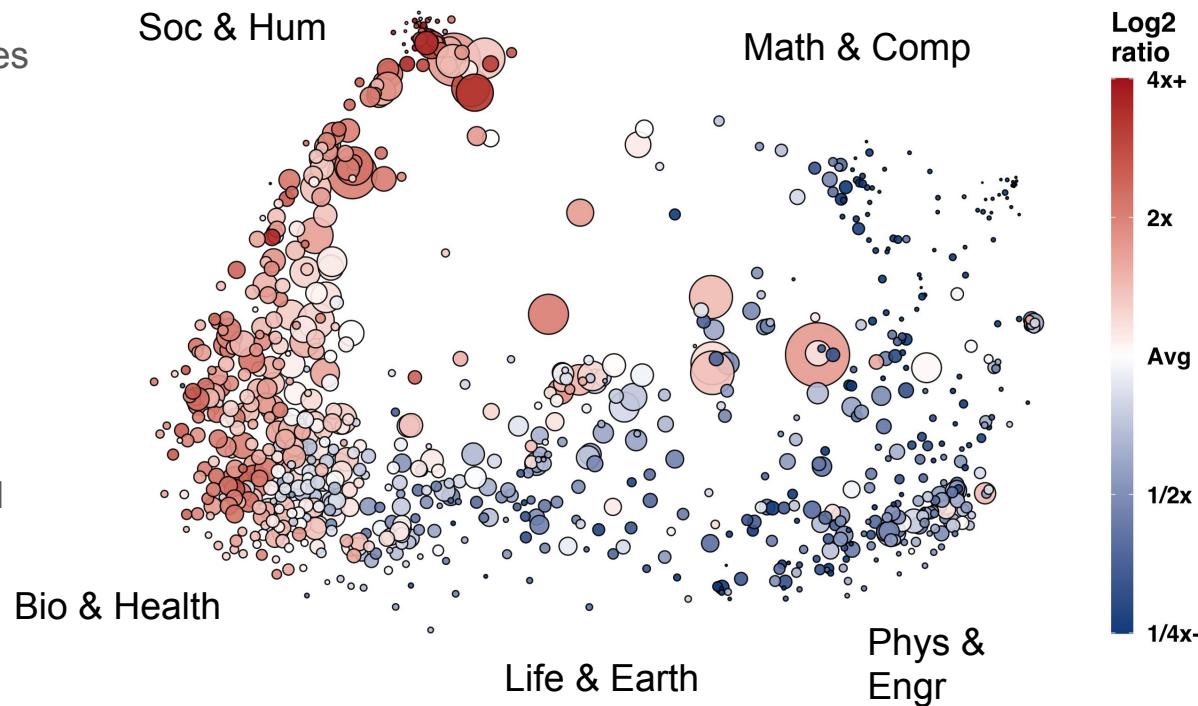
Meso-level fields (clusters of papers)

Area maps to total citation sentences

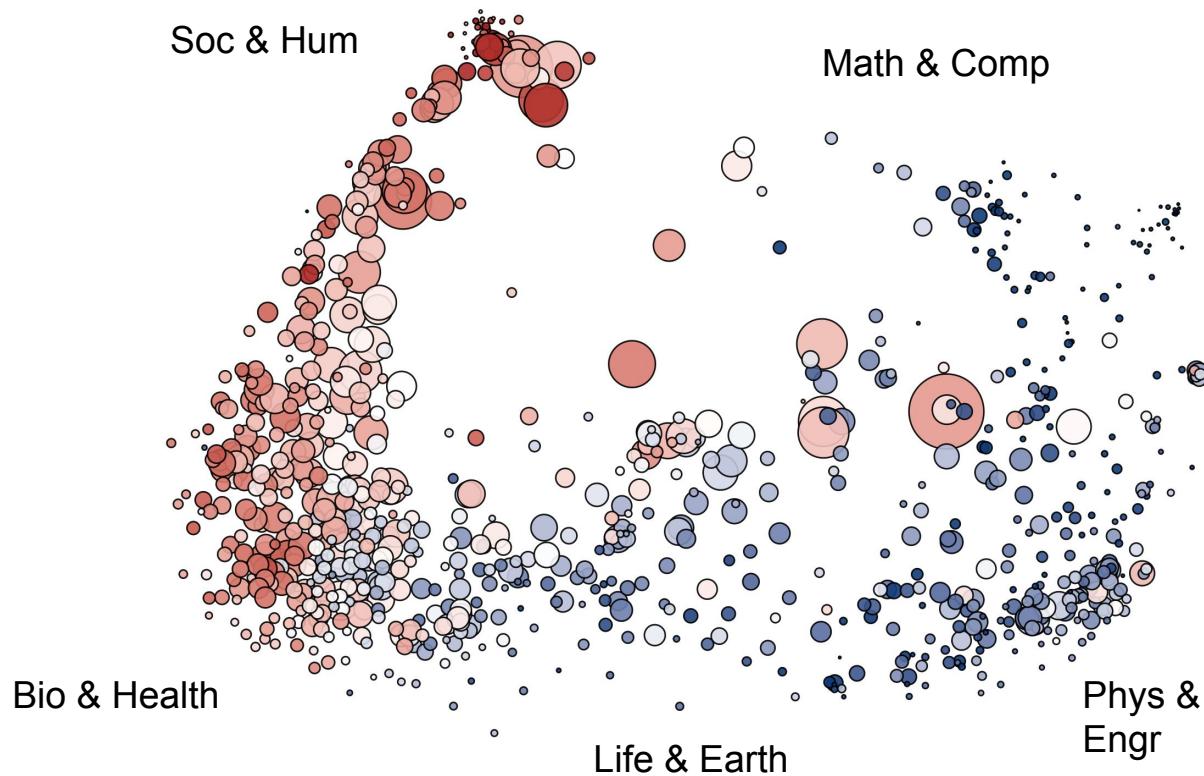
Distance reflects relatedness

Color indicates ratio of %
disagreement to the average

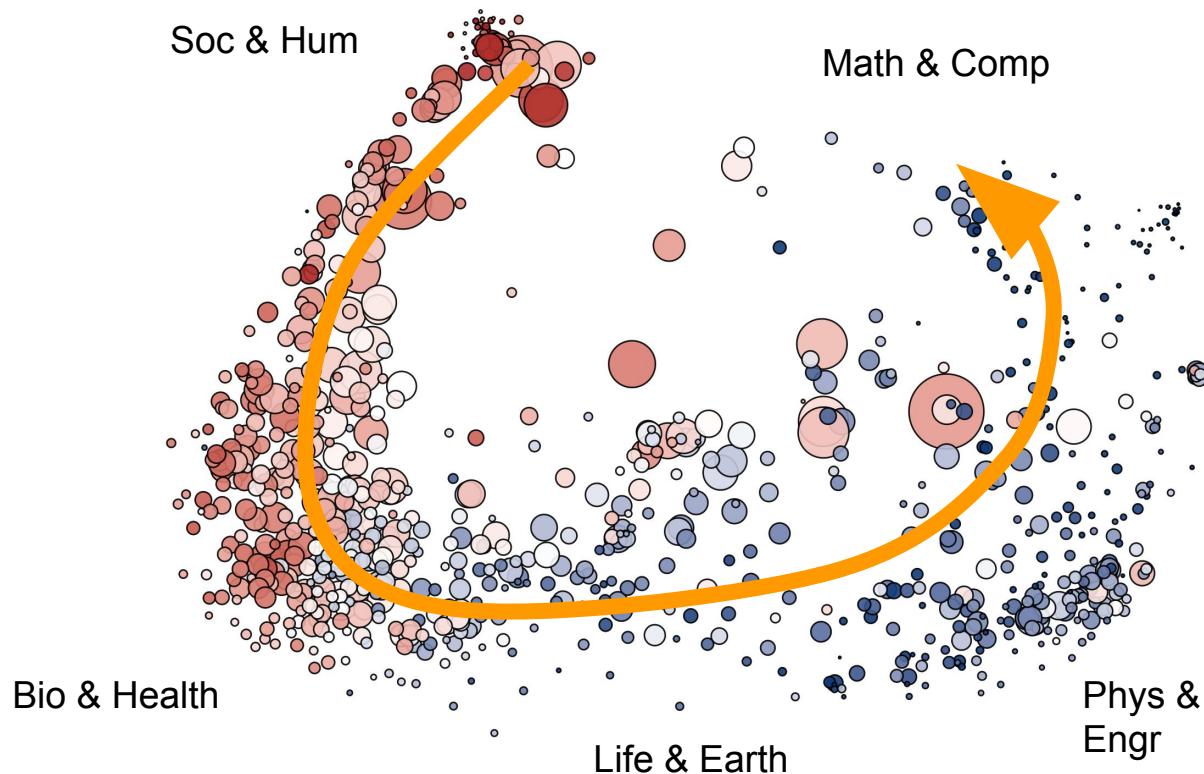
“Horseshoe” shape moves from
social sciences to Mathematics and
computer science



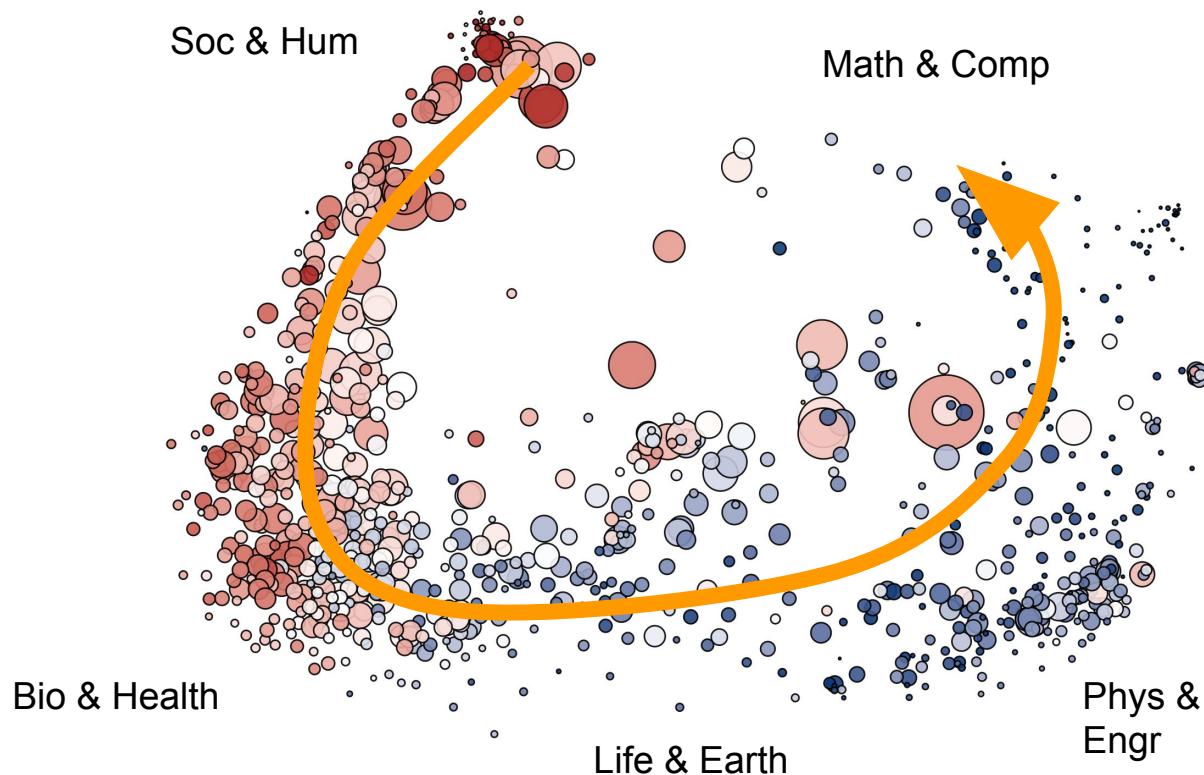
Hierarchy repeated



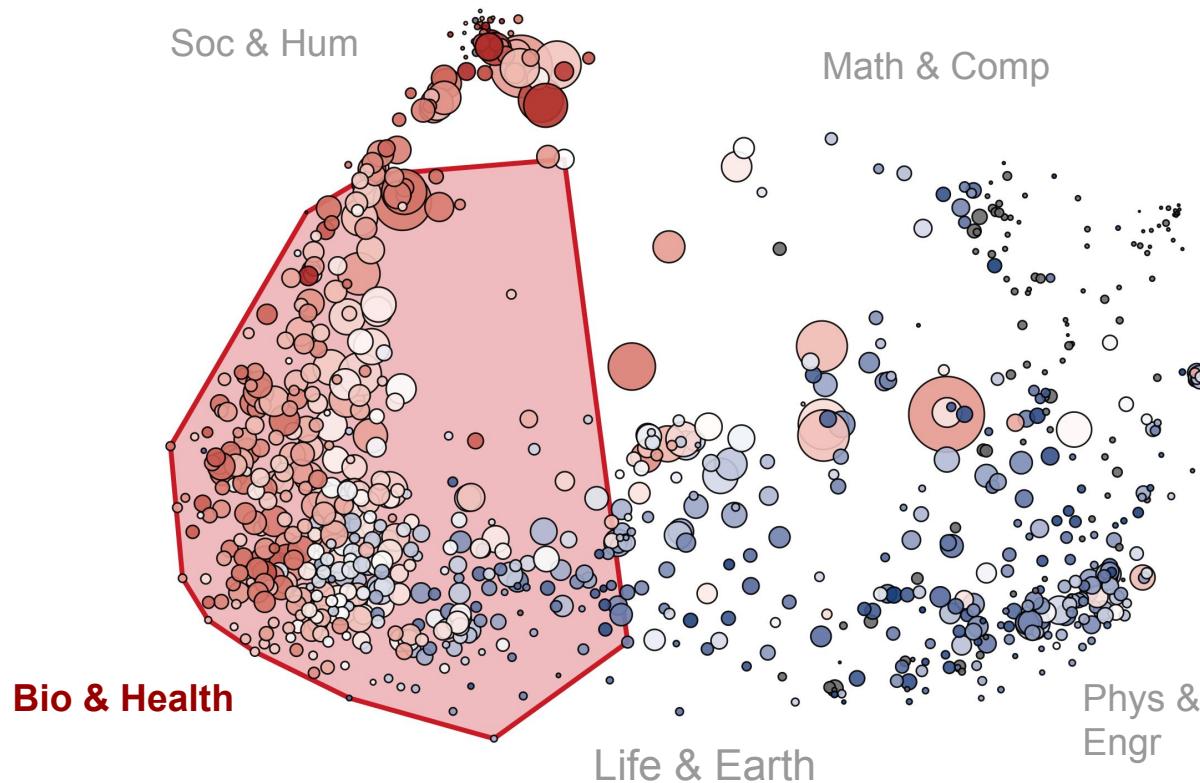
Hierarchy repeated...



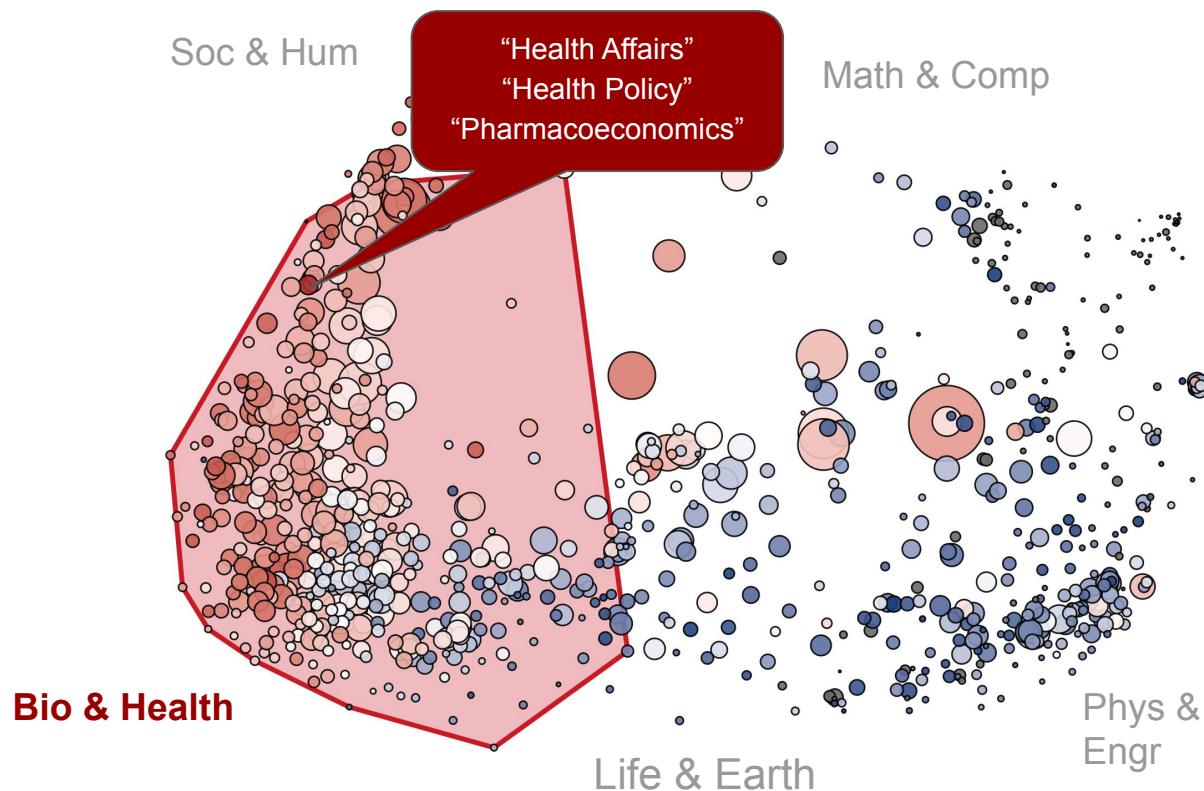
Hierarchy repeated...with heterogeneity



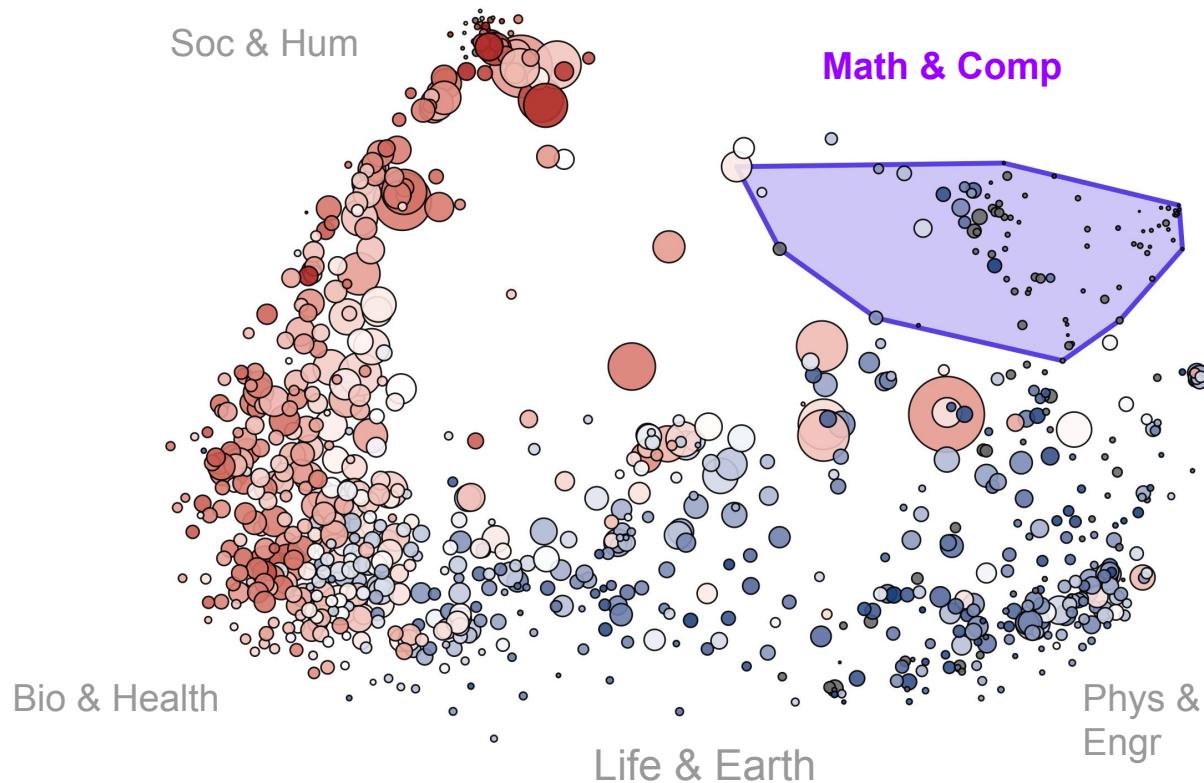
Biomedical and Health Sciences



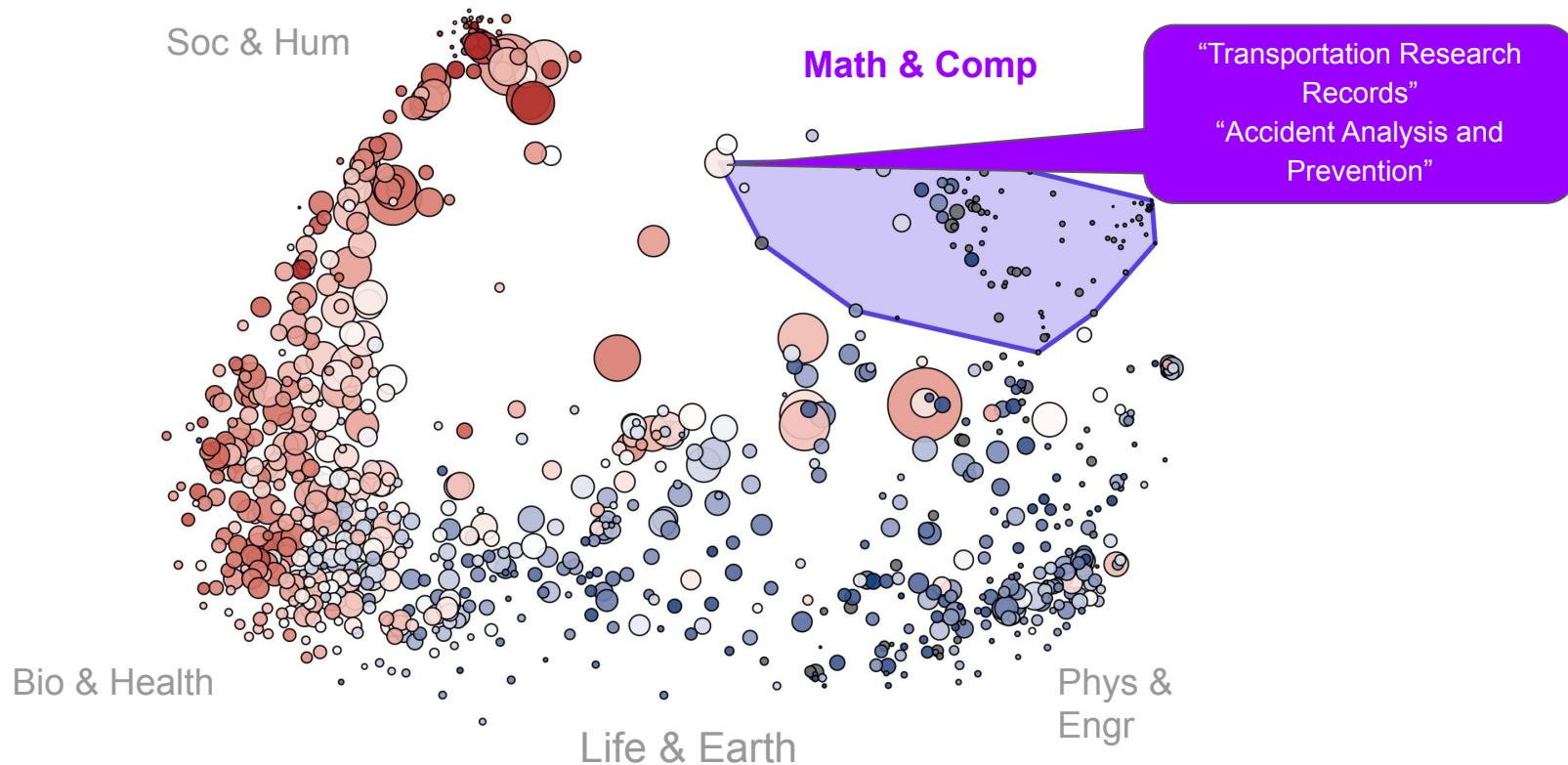
Biomedical and Health Sciences



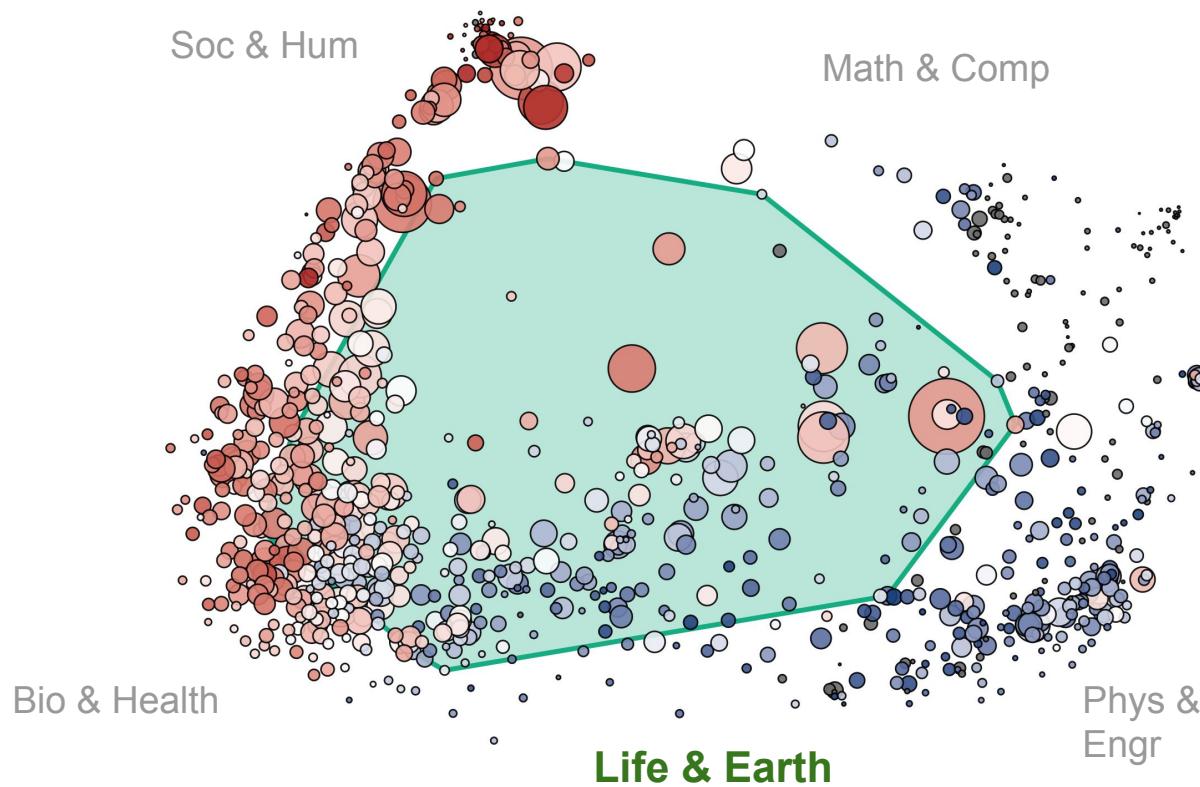
Similar in Mathematics and Computer Science



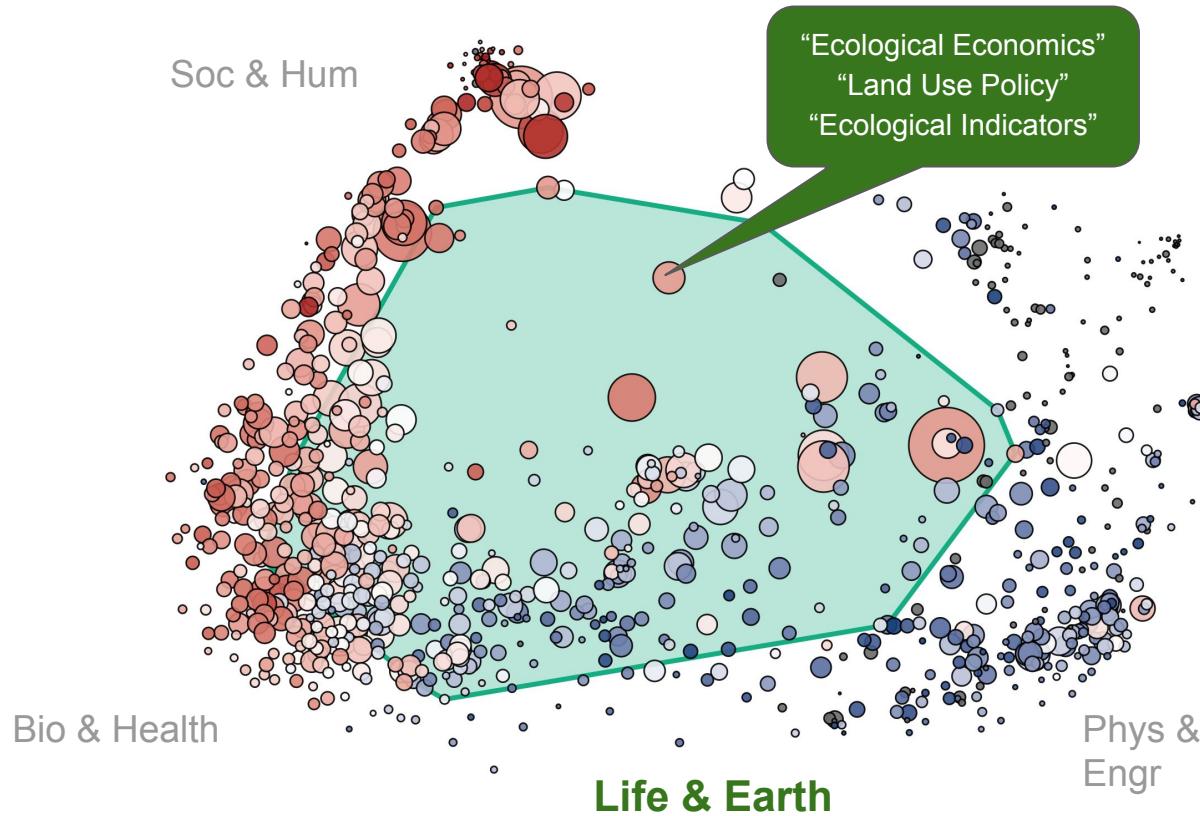
Similar in Mathematics and Computer Science



Similar in Life & Earth Science



Similar in Life & Earth Science



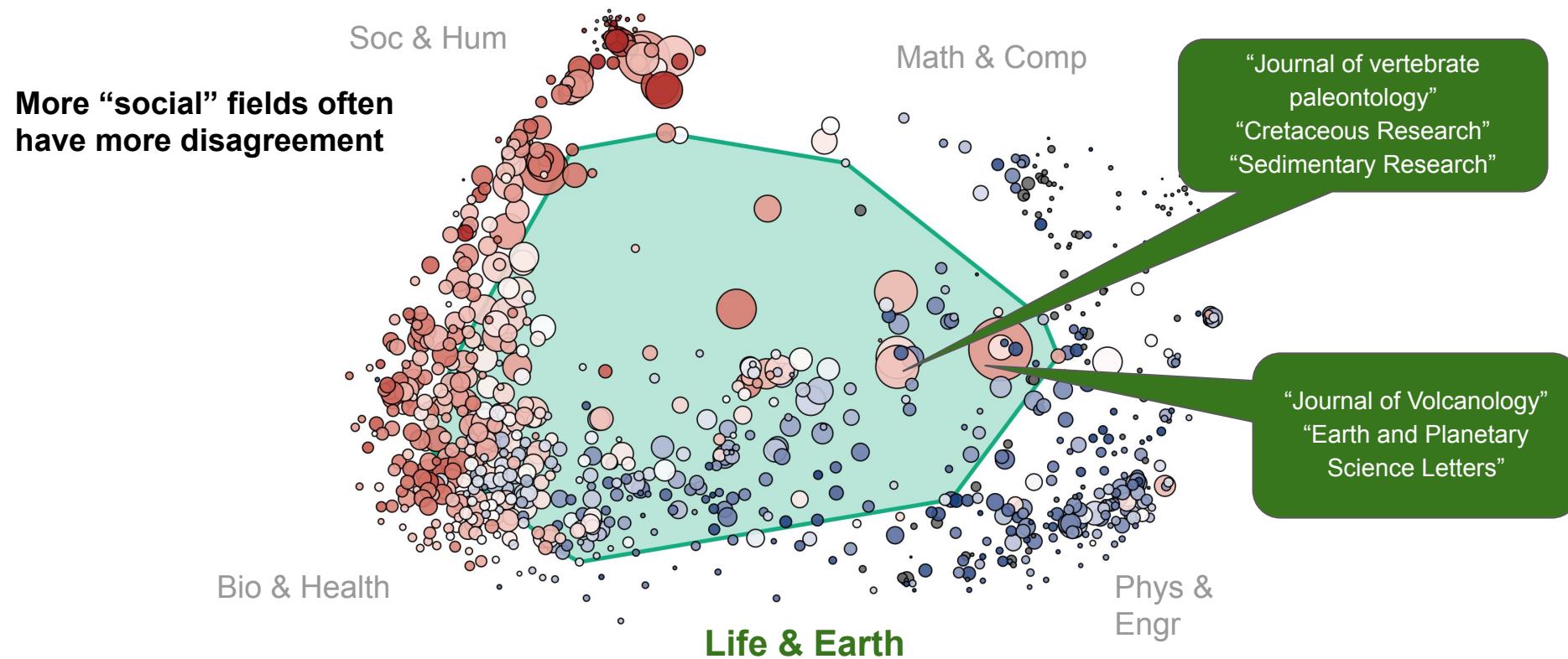
Similar in Life & Earth Science

Soc & Hum
More “social” fields often have more disagreement

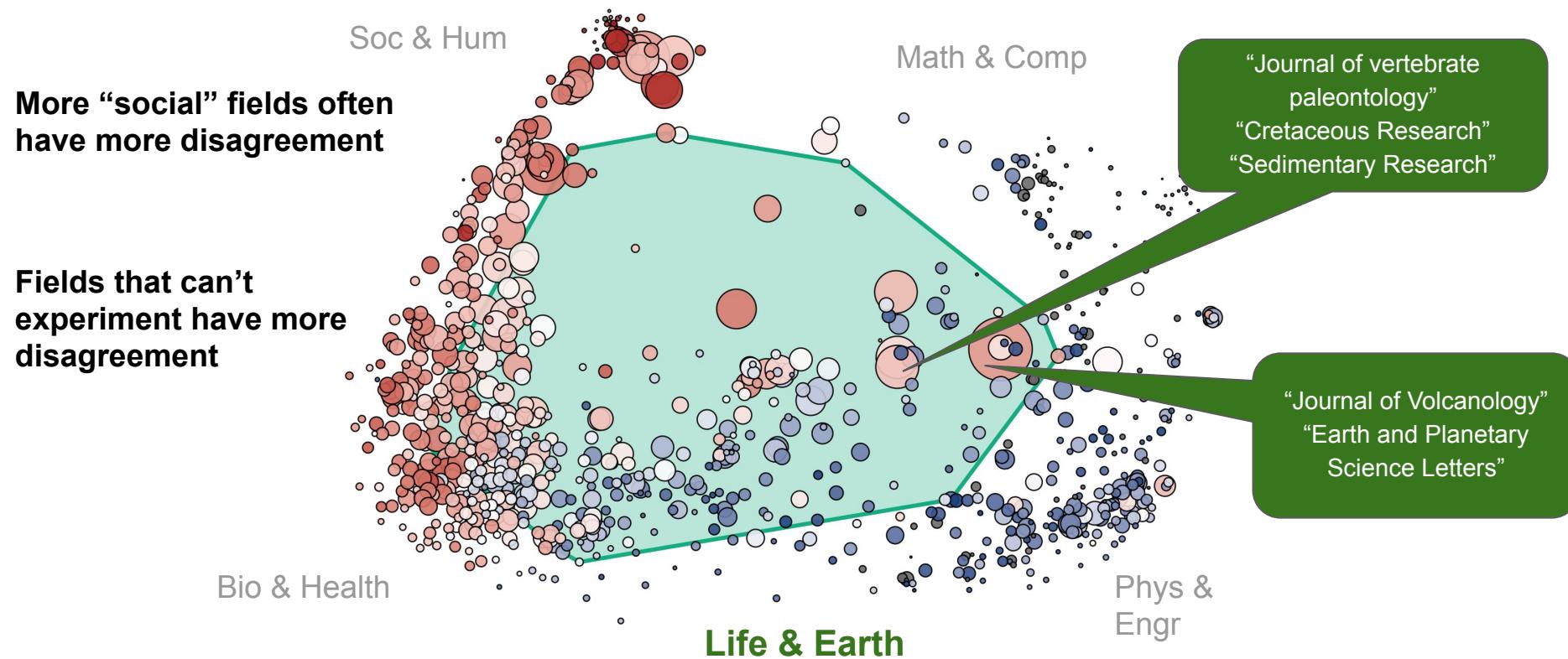
“Ecological Economics”
“Land Use Policy”
“Ecological Indicators”



Similar in Life & Earth Science...but also new patterns



Similar in Life & Earth Science...but also new patterns

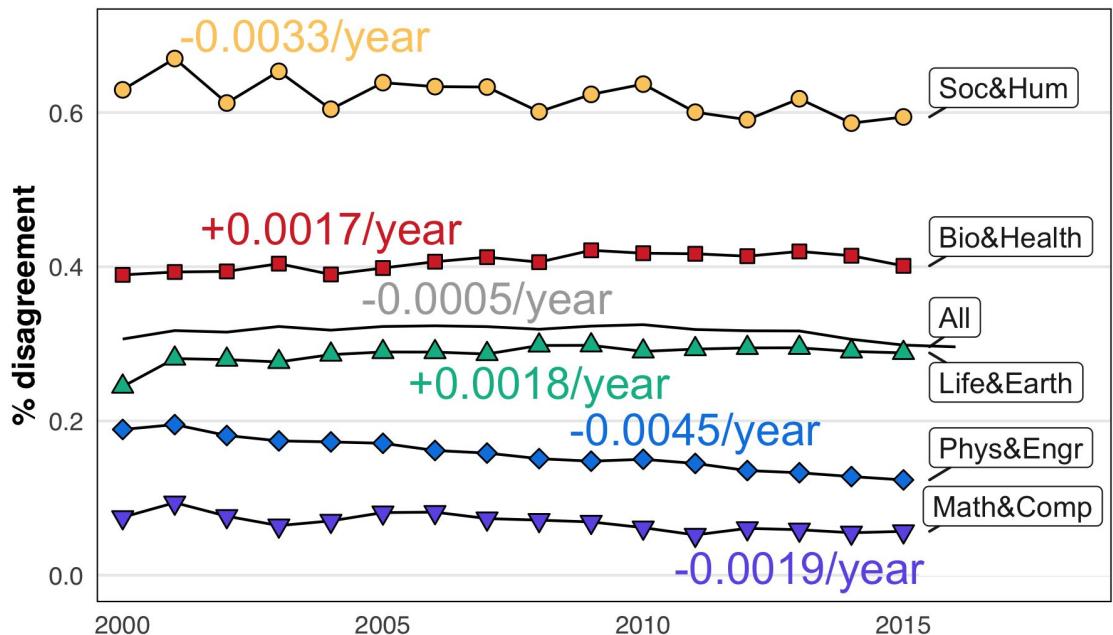


Takeaways

- Re-confirm the hierarchy of sciences using disagreement
 - Additional evidence that our approach is robust
 - Supports past bibliometric studies of the hierarchy (Fanelli & Glänzel, 2013)
- Identify heterogeneity
 - Based on epistemic characteristics of the fields

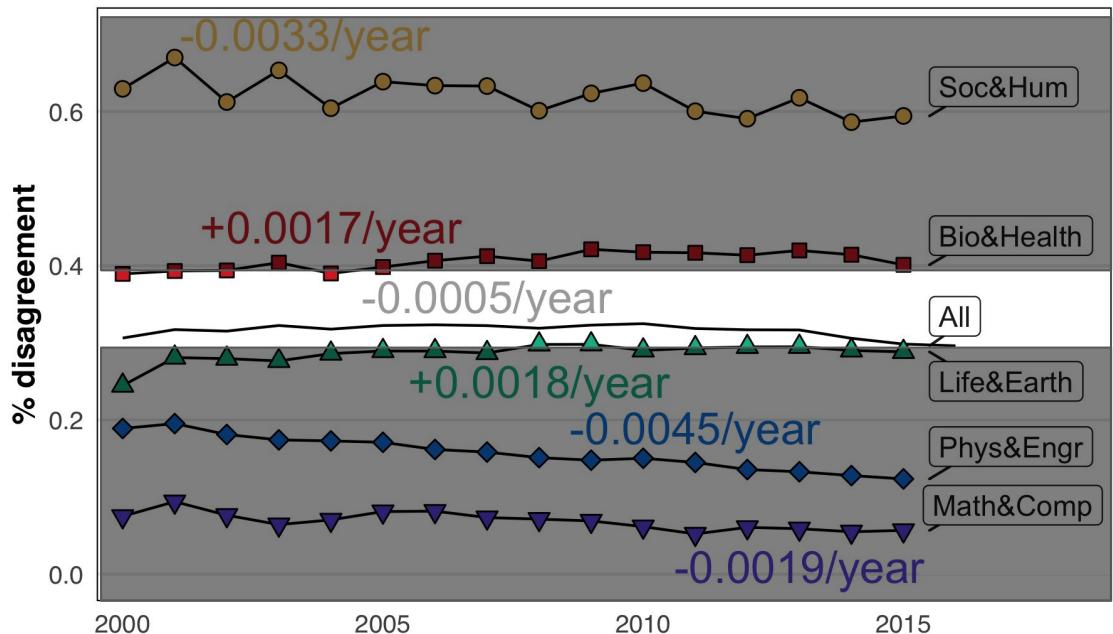
Measuring
disagreement
over time

Disagreement over time



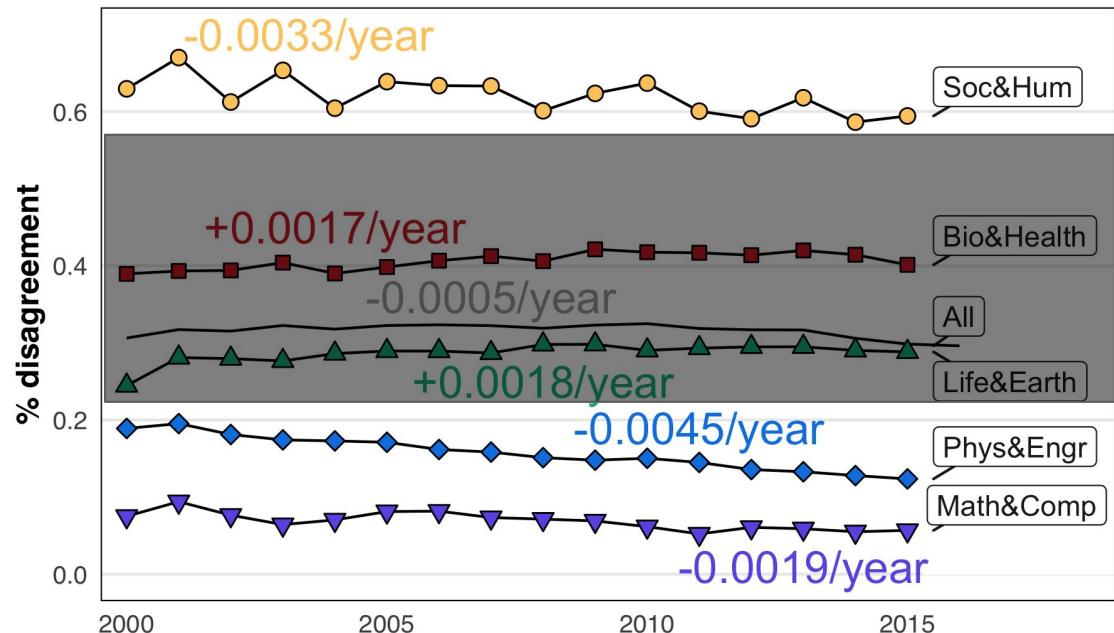
Disagreement over time

- Mostly stable, slight decrease



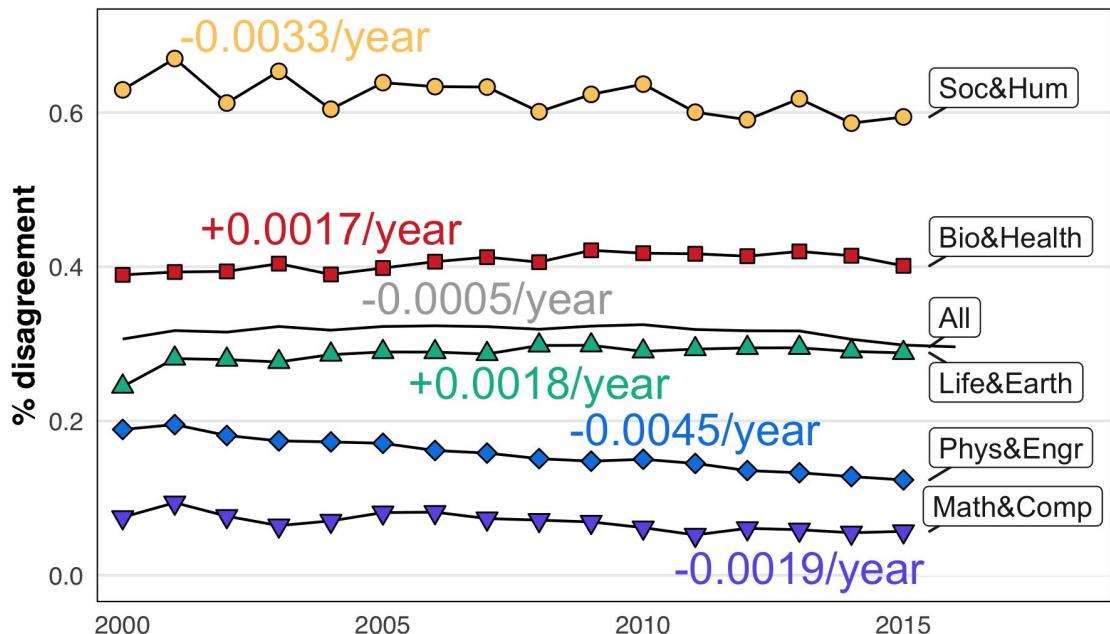
Disagreement over time

- Mostly stable, slight decrease
- Mostly driven by a few fields



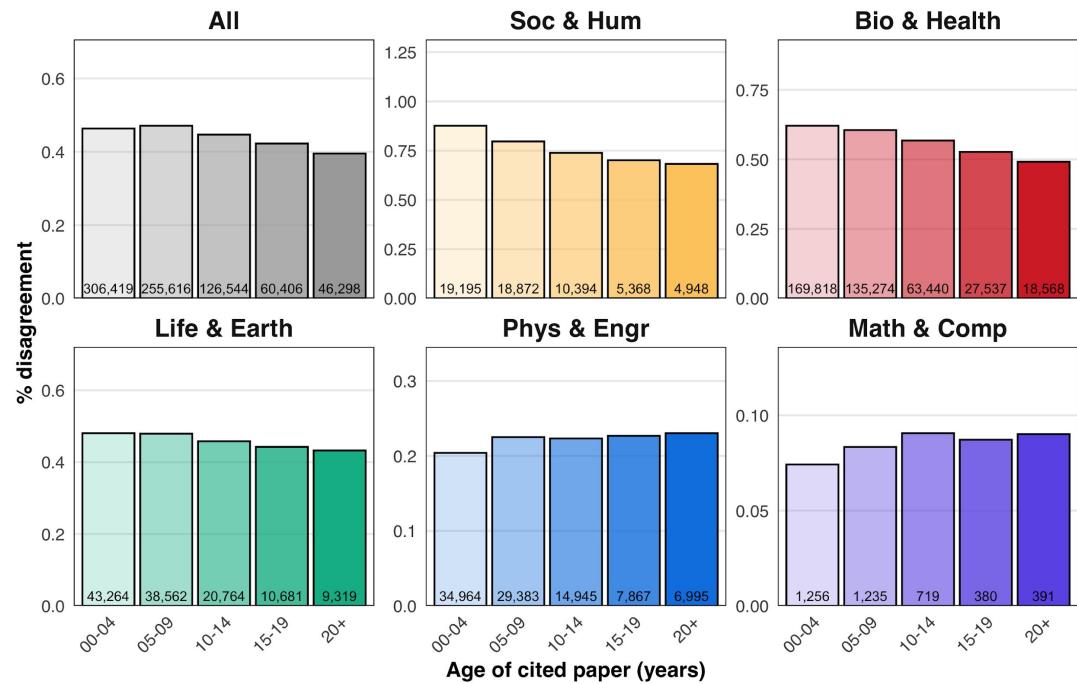
Disagreement over time

- Mostly stable, slight decrease
- Mostly driven by a few fields
- Takeaways
 - Is this real?
 - Are these changes significant?



Measuring disagreement *by age of cited paper*

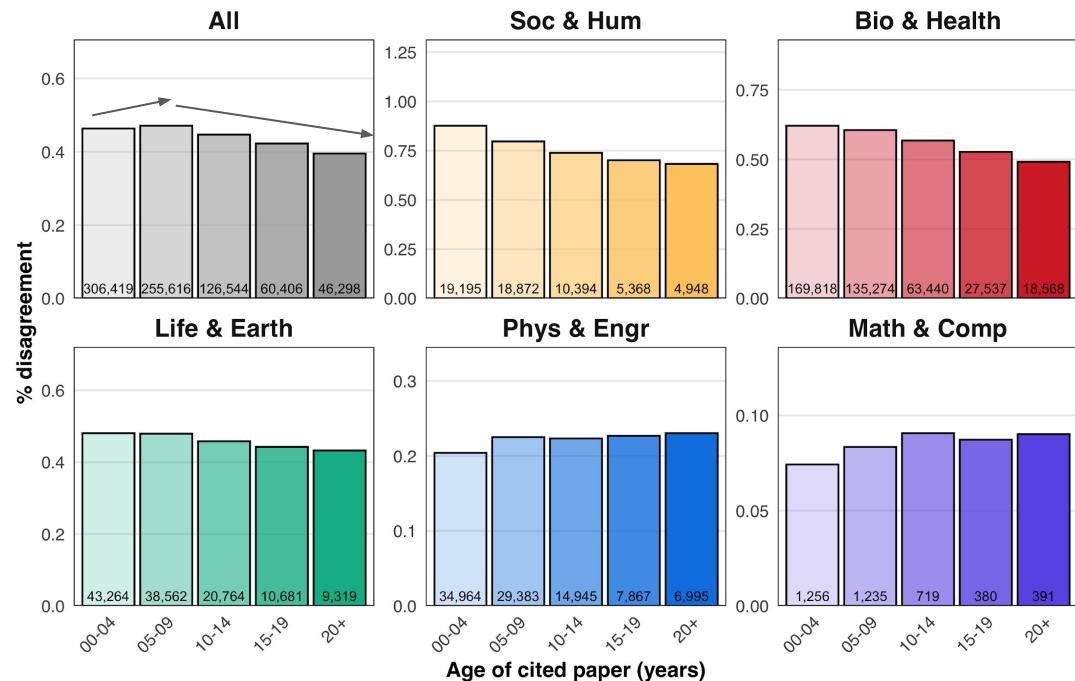
Disagreement and age



Relative age of cited paper (years)

Disagreement and age

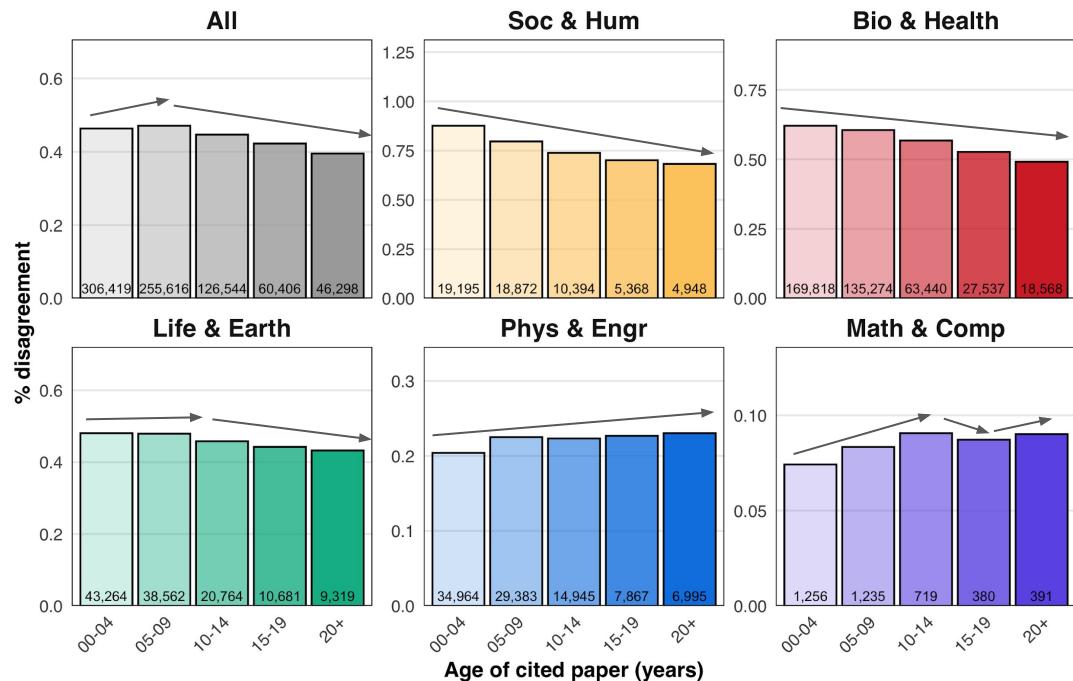
- After a *bump*, older papers are disagreed with less



Relative age of cited paper (years)

Disagreement and age

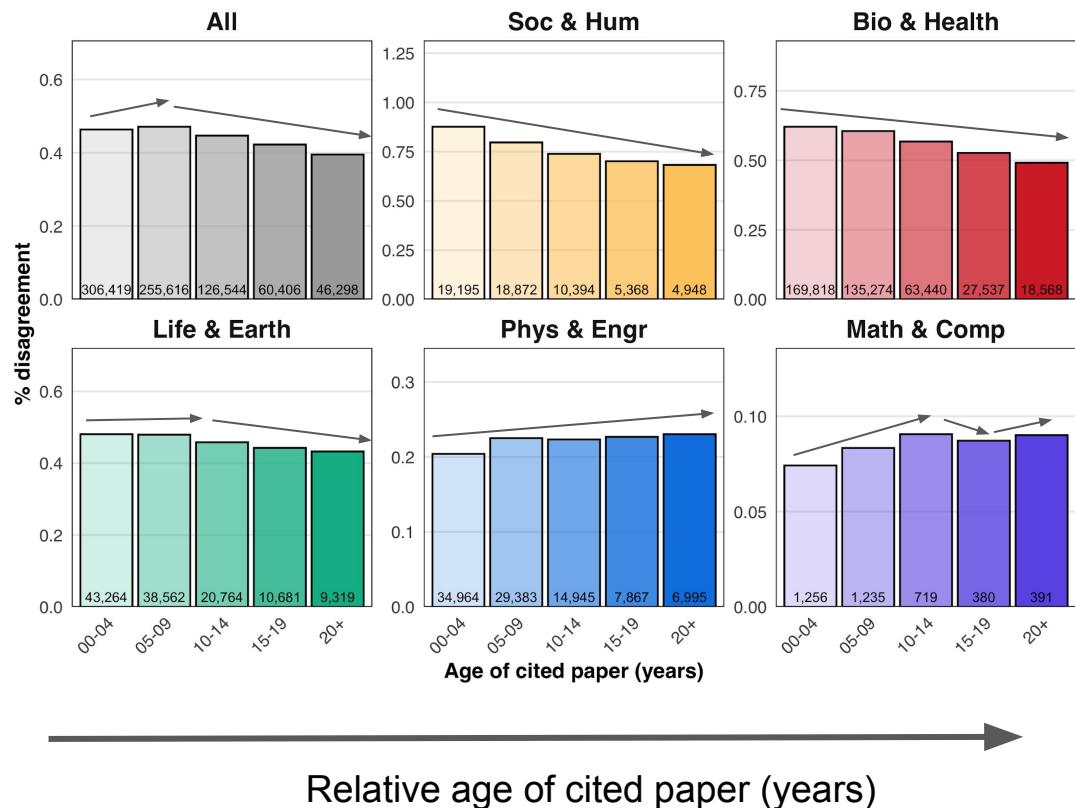
- After a *bump*, older papers are disagreed with less
- Field differences



Relative age of cited paper (years)

Disagreement and age

- After a *bump*, older papers are disagreed with less
- Field differences
- Takeaways
 - Disagreement relates to age
 - Selection effect?



Measuring
disagreement...

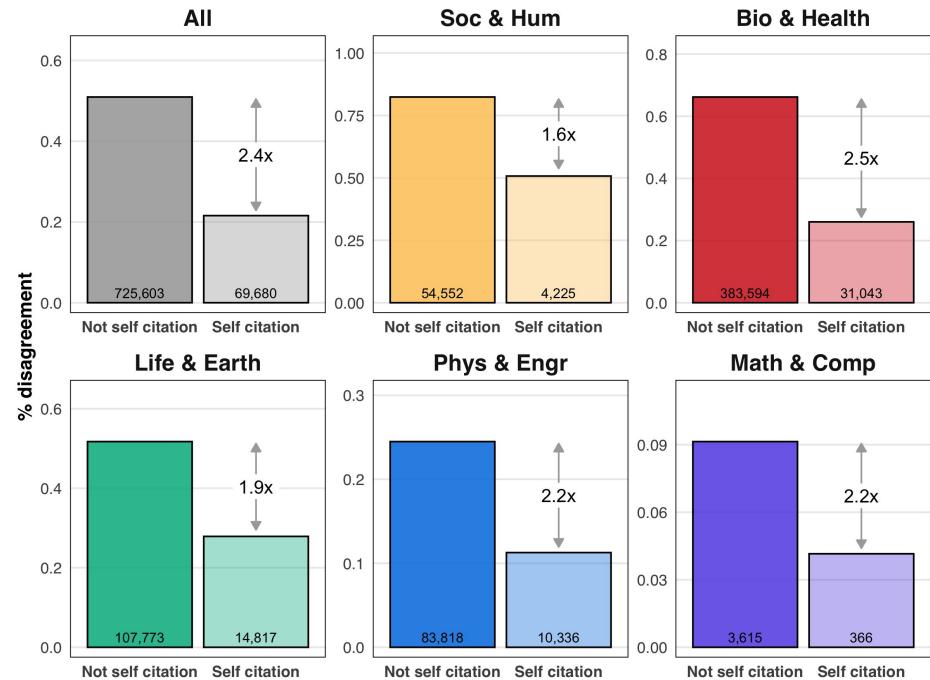
by self citation

Disagreement and Self-citation

- People *should* disagree with themselves less

Disagreement and Self-citation

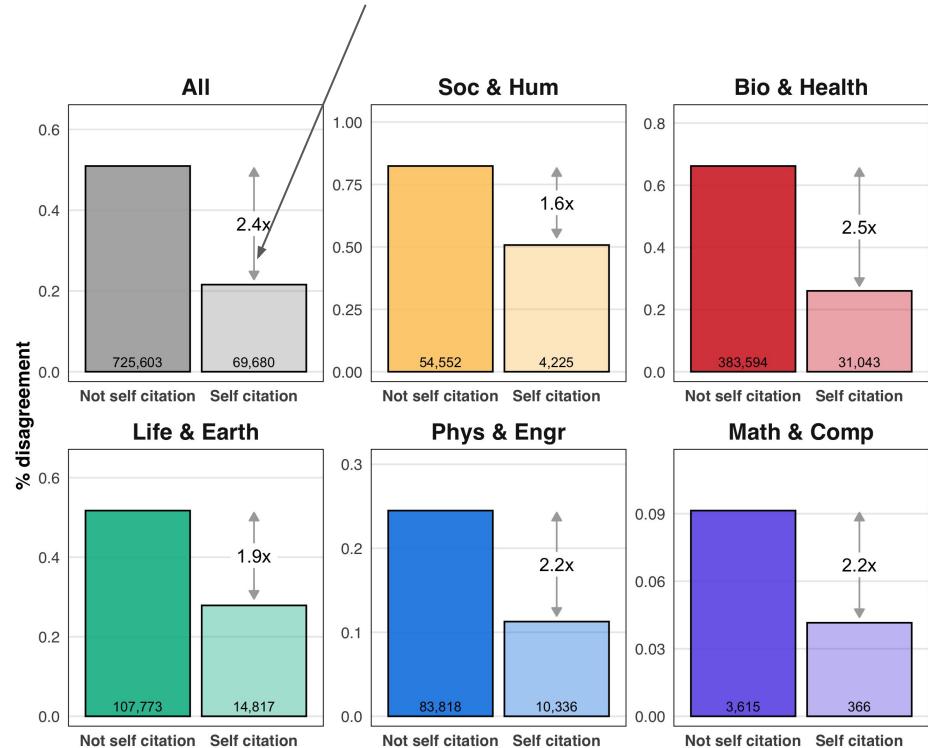
- People *should* disagree with themselves less



Disagreement and Self-citation

- People *should* disagree with themselves less
- Similar across fields

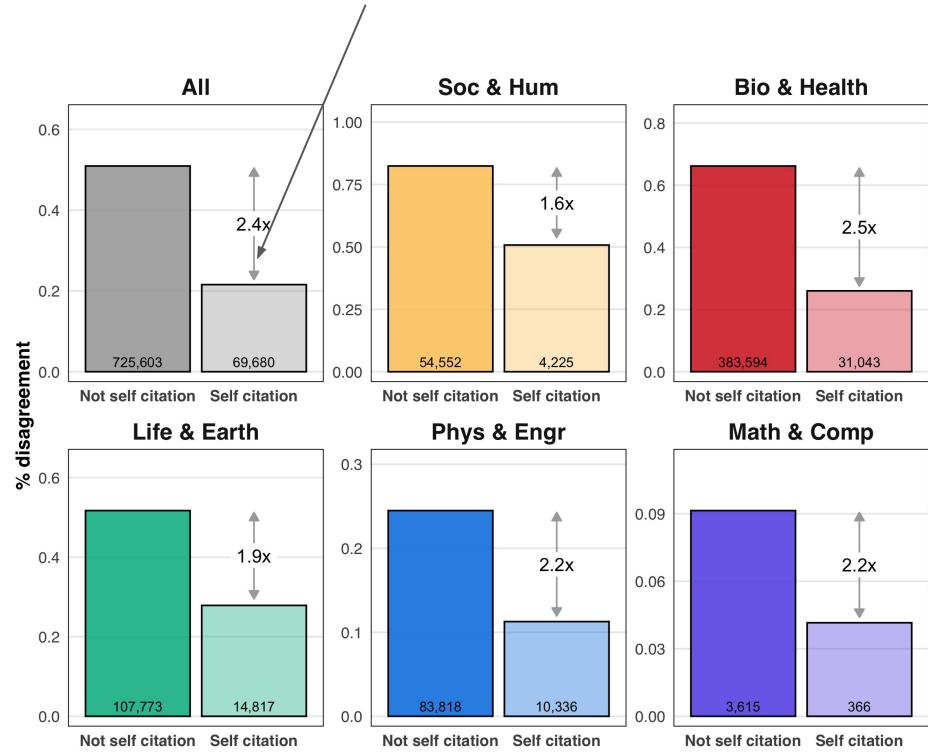
Presence of author overlap



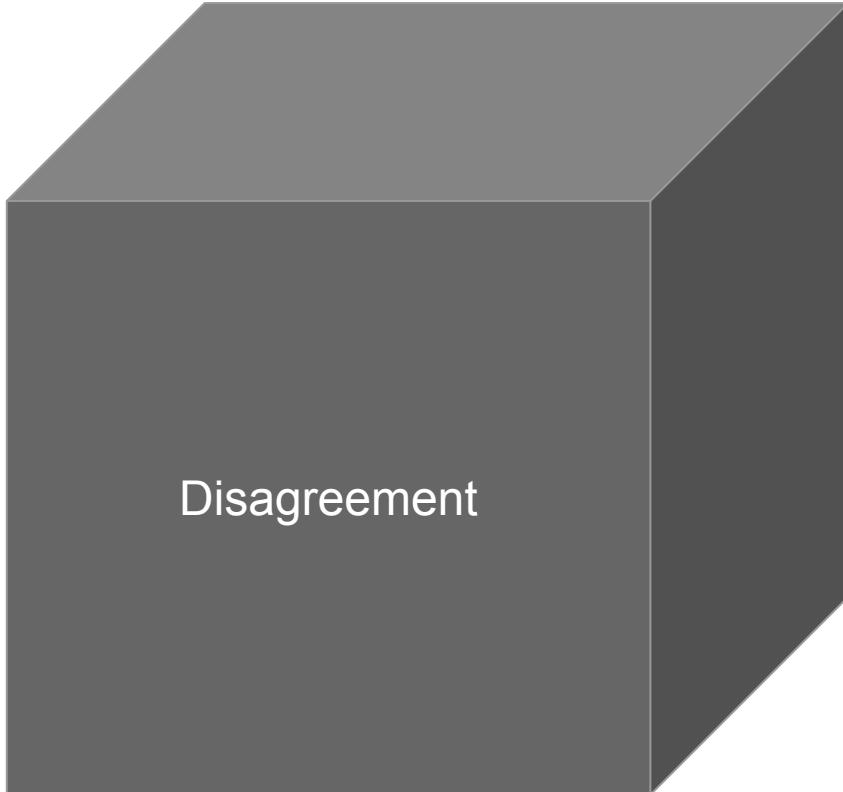
Disagreement and Self-citation

- People *should* disagree with themselves less
- Similar across fields
- Takeaways
 - Confirms expectations
 - Perfunctory citations?

Presence of author overlap

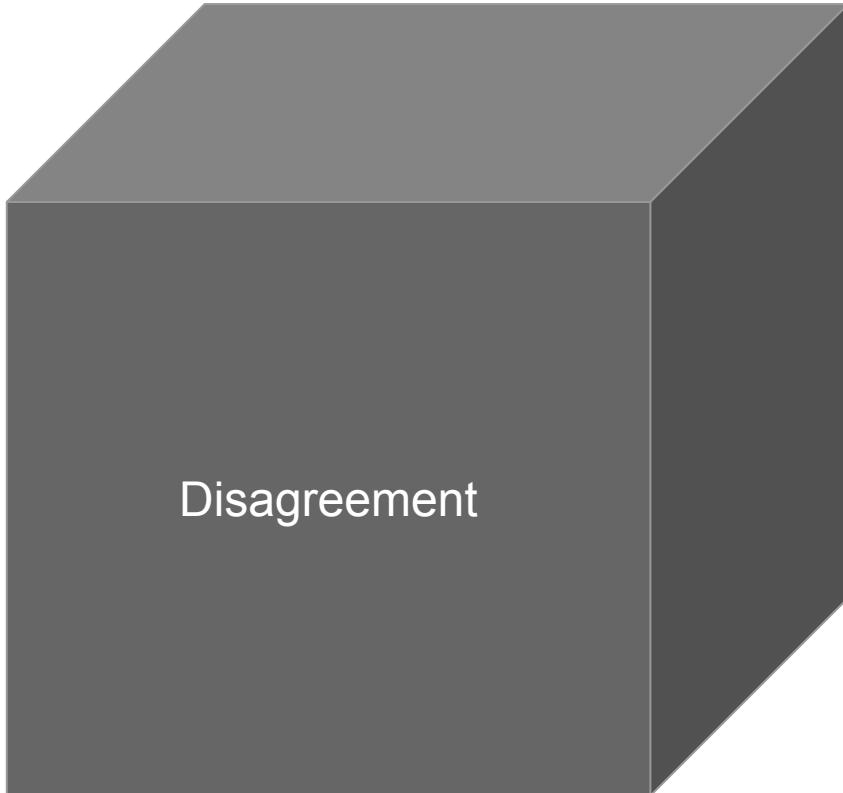


So far, our measure has revealed insights based on...



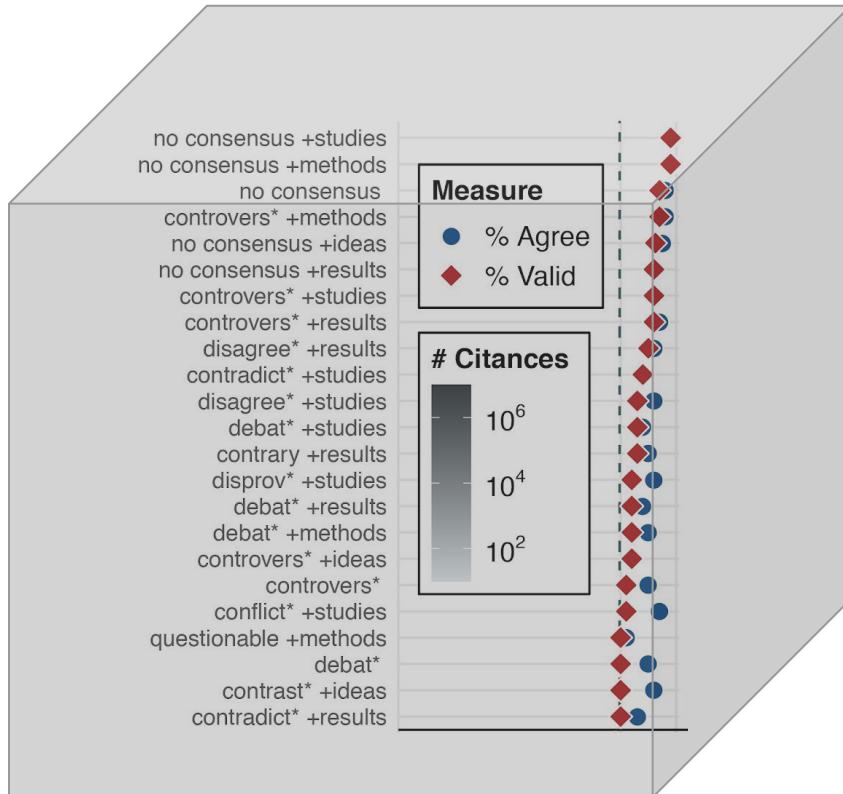
- Discipline
- Time
- Age of cited paper
- Self citation

So far, our measure has revealed insights based on...



- Discipline
 - Time
 - Age of cited paper
 - Self citation
-
- And the process of keyword-based studies itself

Our approach is transparent



- We can *open the black box* and see exactly where issues occur
 - In contrast to common Machine Learning techniques
- Contextualizes our analysis
- Informs future keyword-based studies

When a signal is a proper noun

“...the conflict monitoring theory [6] provides a slightly different way to account for conflict effects.”

When a signal is an object of study

“Caselli and Coleman study conflicts based on ethnicity....”

When a signal is just in reference to the wrong thing

“...consumers are usually self-motivated and enthusiastic in exchanging, sharing and debating ideas... (Anderson & Weitz, 1992).”

When it's somebody's name

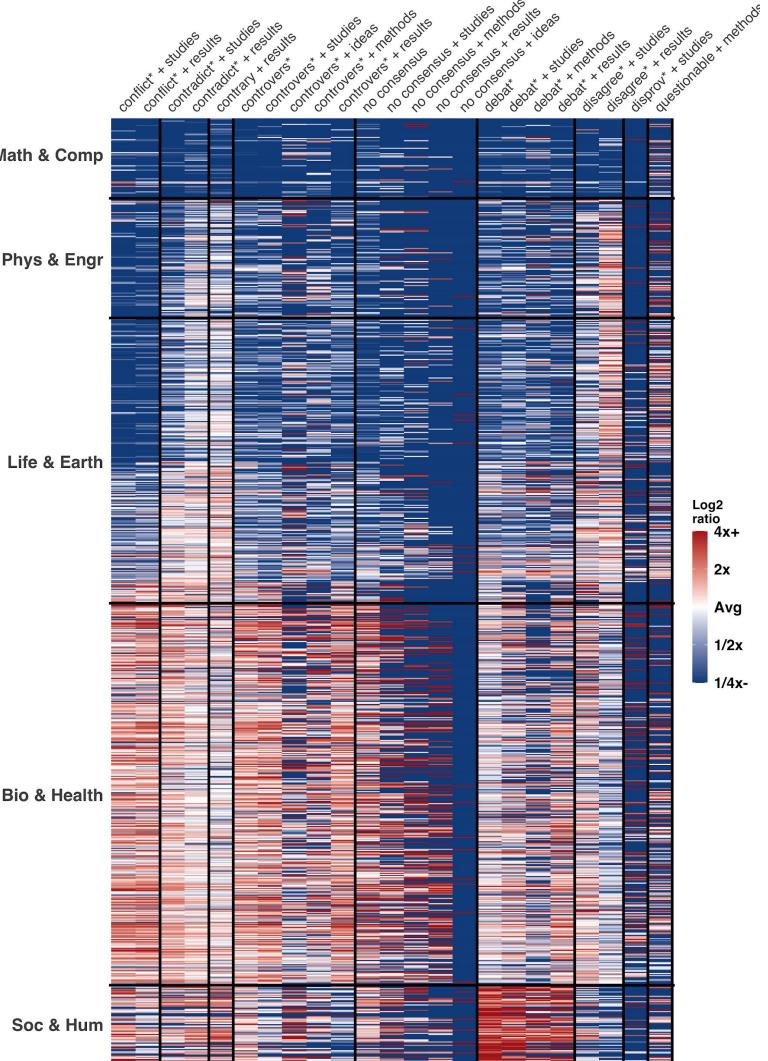
“Measurement items for perceived risks were also developed from **Debatin** et al.’s (2009) study.”

Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field

Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field

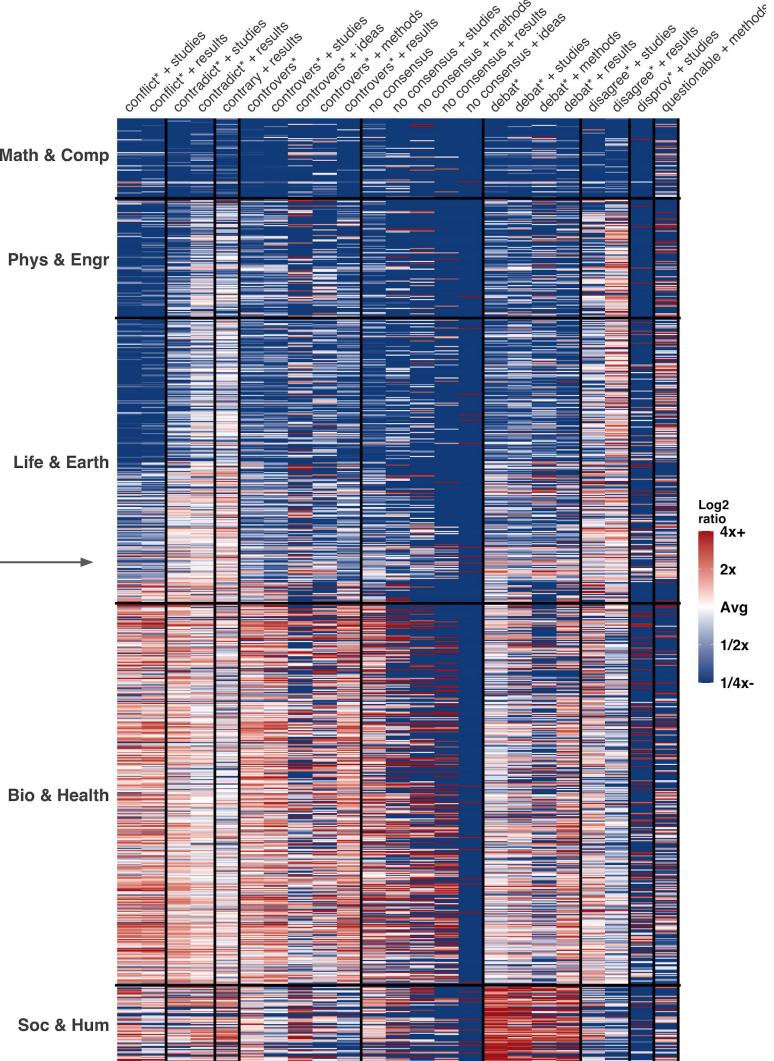


Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field

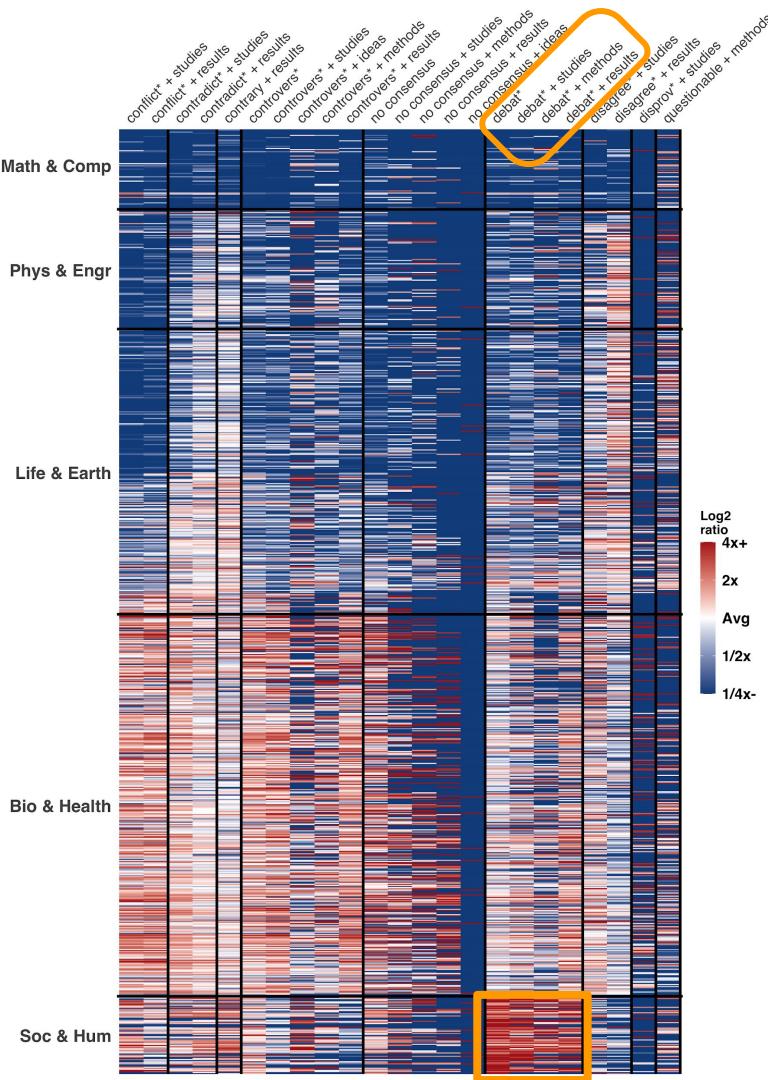
Each line corresponds
to a meso-field

Color matches ratio to
average across all fields



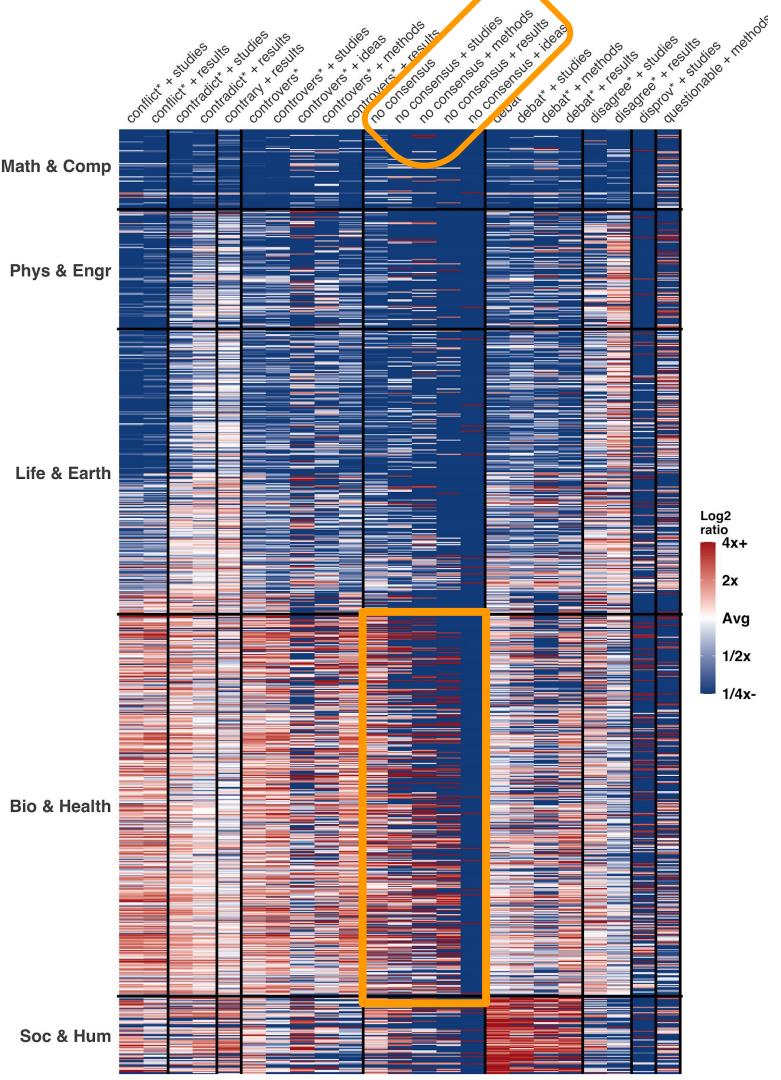
Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field
- Debat* overrepresented in Soc & Hum



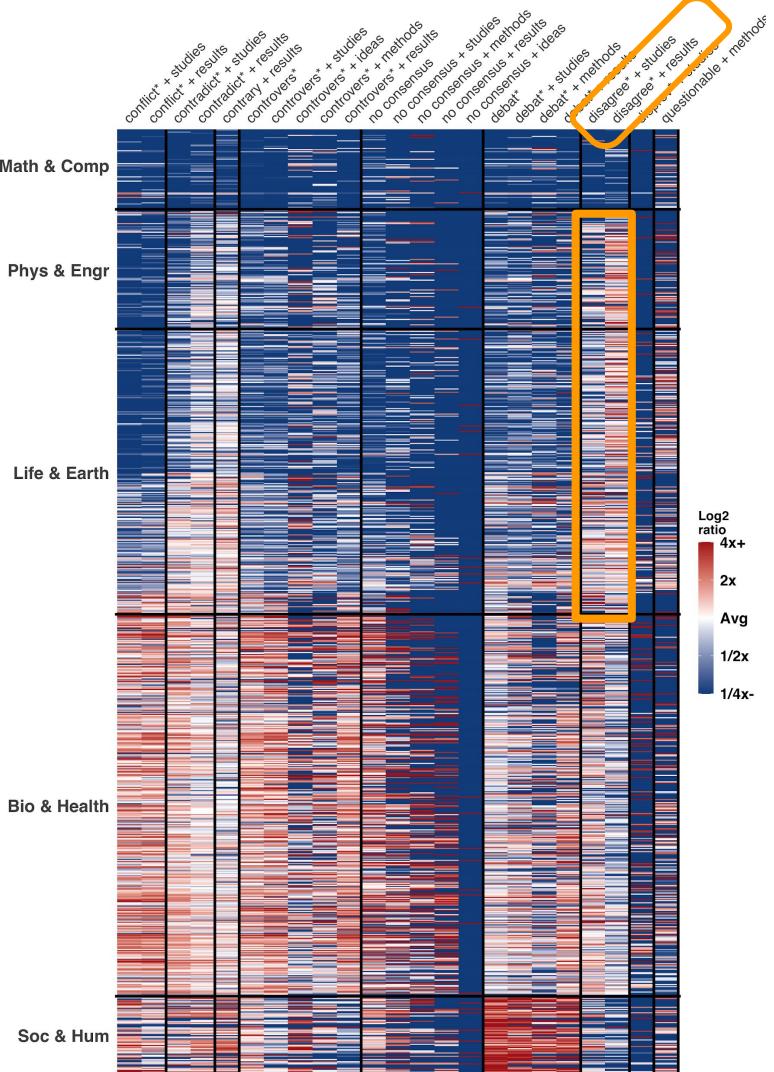
Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field
- Debat* overrepresented in Soc & Hum
- No consensus in Bio & Health



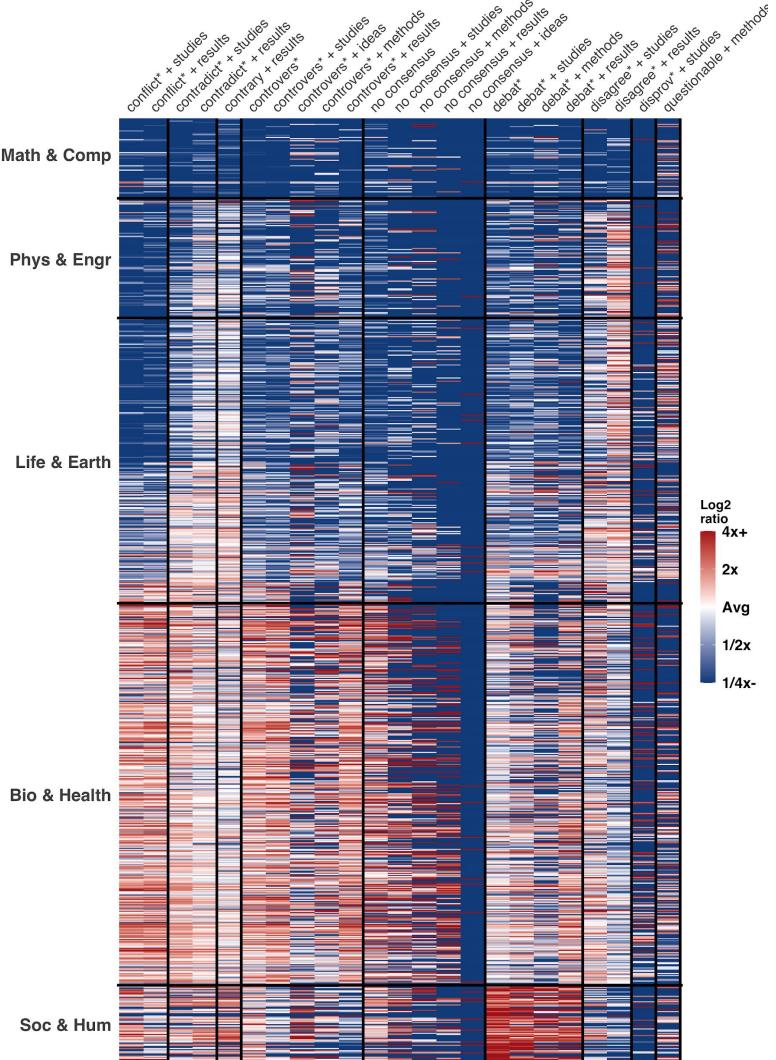
Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field
- Debat* overrepresented in Soc & Hum
- No consensus in Bio & Health
- Disagree* in Phys & Engr, Life & Earth



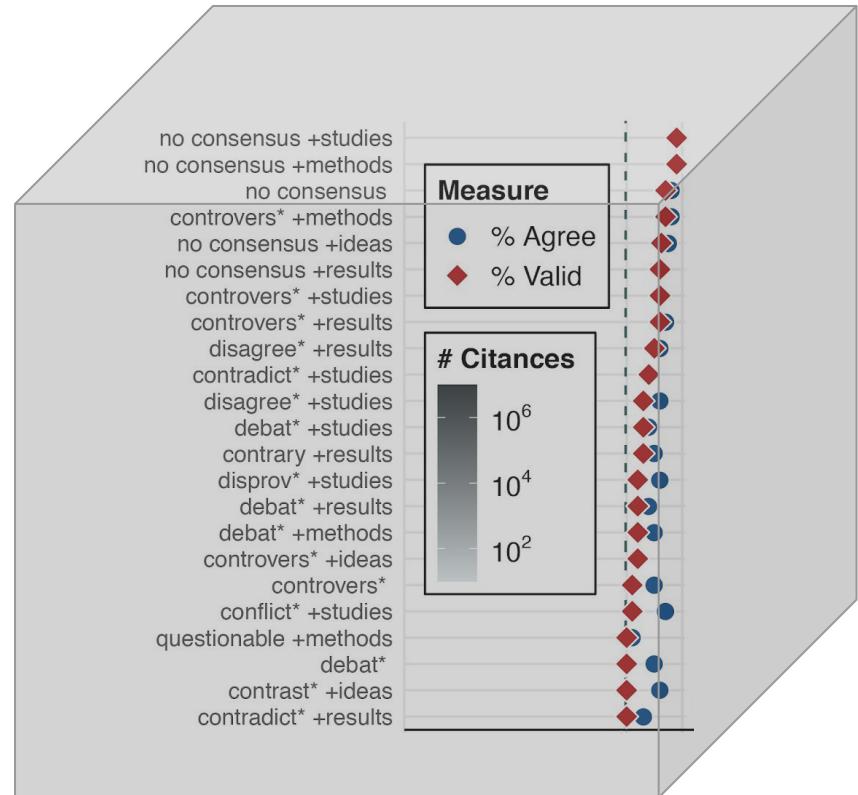
Isolated, or systematic?

- Overall, we capture disagreement,
 - Heterogeneity by field
- Debat* overrepresented in Soc & Hum
- No consensus in Bio & Health
- Disagree* in Phys & Engr, Life & Earth
- Are these driven by artefacts?



Our approach has limitations, but...

- They can be identified
- Can be easily adjusted
 - Different validity threshold
 - Refine existing queries
 - Add new signal and filter terms
- Future analyses of scientific text will need to consider these artefacts



In Summary

In Summary

- Developed a framework for identifying instances of disagreement
 - Robust
 - Transparent
 - Adjustable

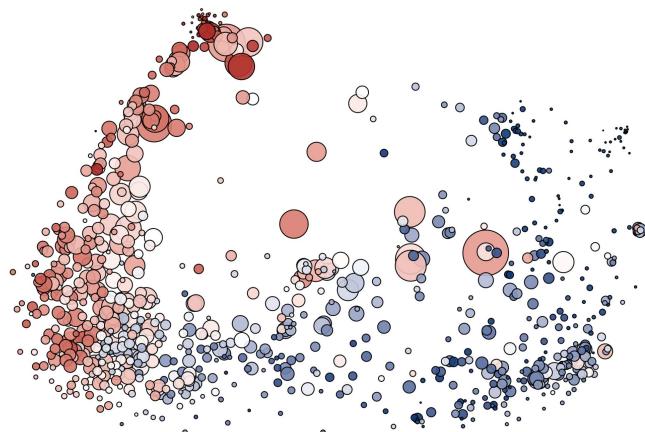
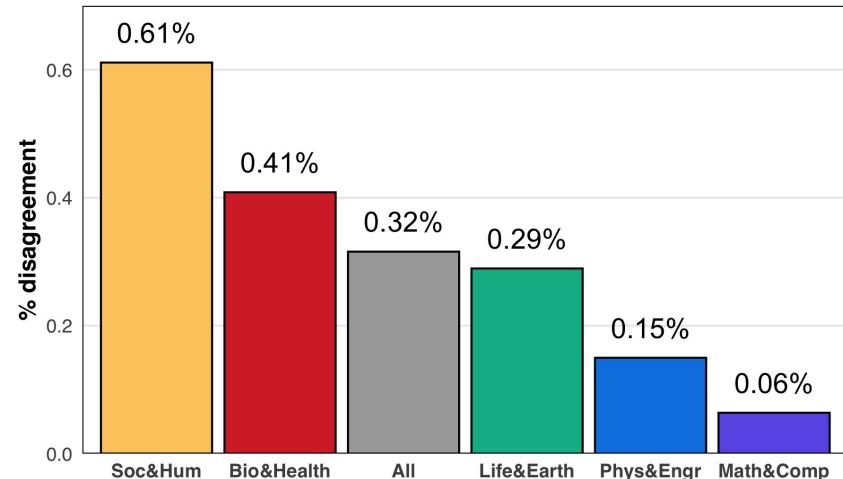
---- Signal terms ----

---- Filter terms ----

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

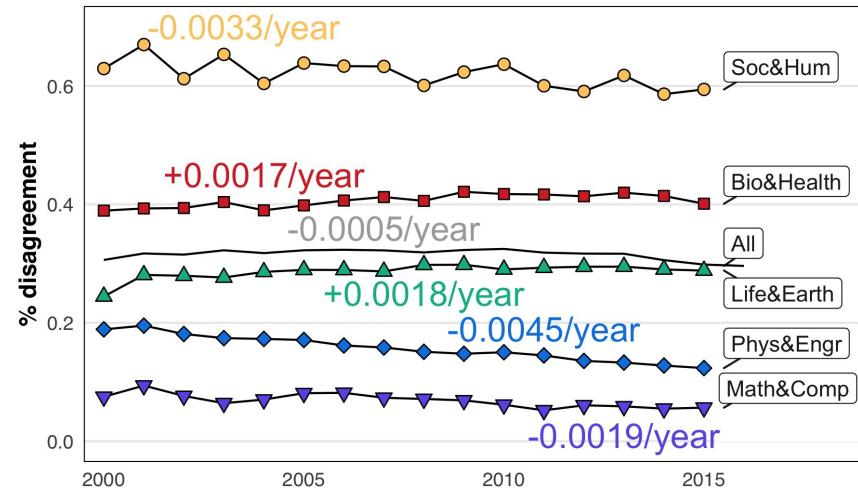
In Summary

- Developed a framework for identifying instances of disagreement
 - Robust
 - Transparent
 - Adjustable
- Confirmed a hierarchy of sciences
 - But with greater heterogeneity at the field level



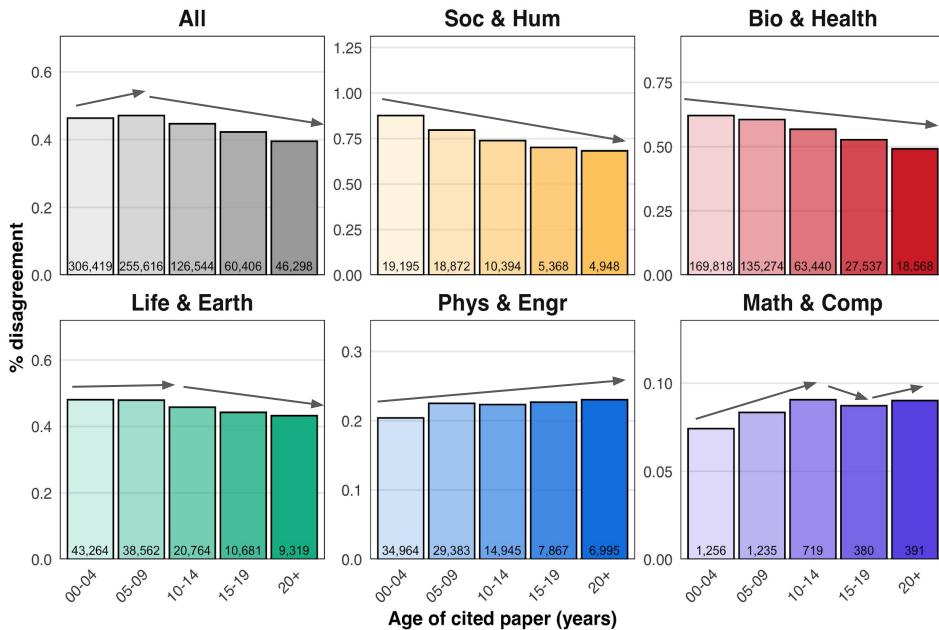
In Summary

- Developed a framework for identifying instances of disagreement
 - Robust
 - Transparent
 - Adjustable
- Confirmed a hierarchy of sciences
 - But with greater heterogeneity at the field level
- Observed minor temporal changes



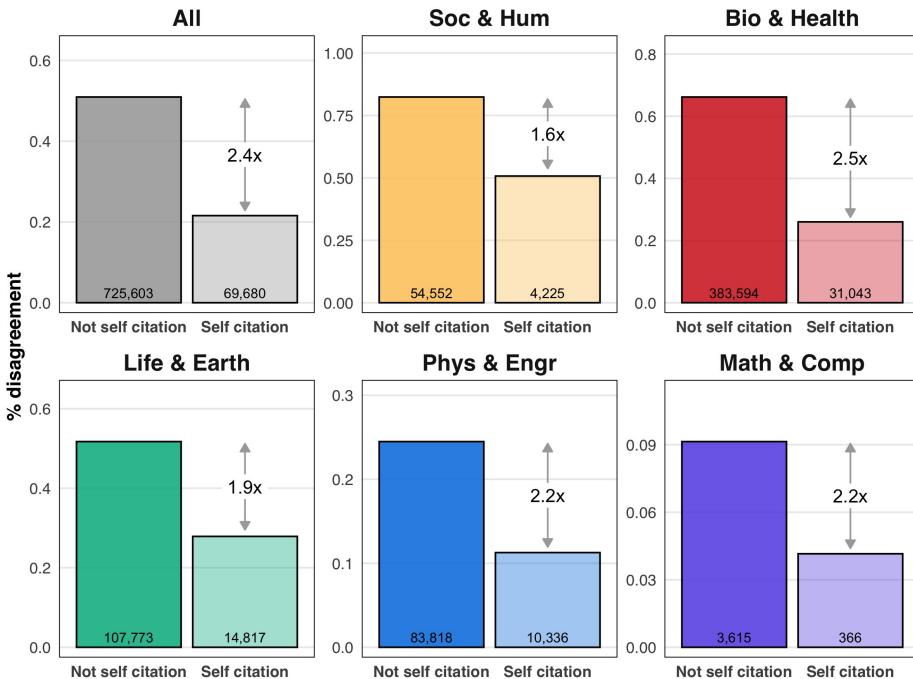
In Summary

- Developed a framework for identifying instances of disagreement
 - Robust
 - Transparent
 - Adjustable
- Confirmed a hierarchy of sciences
 - But with greater heterogeneity at the field level
- Observed minor temporal changes
- Relative age matters



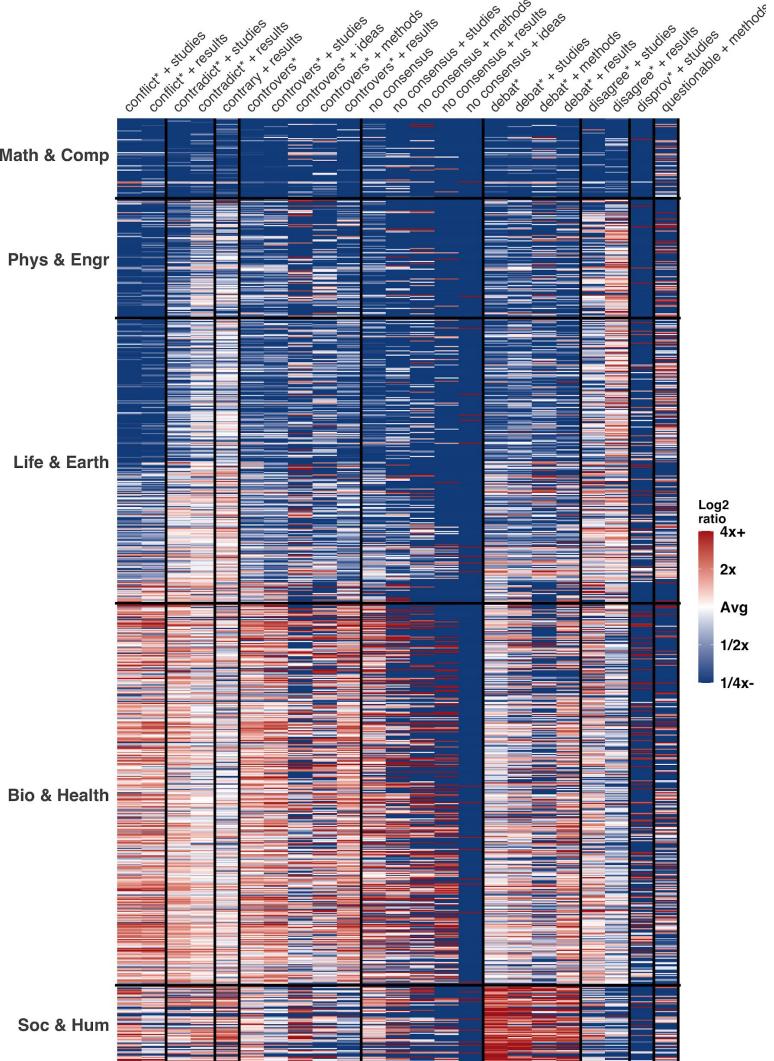
In Summary

- Developed a framework for identifying instances of disagreement
 - Robust
 - Transparent
 - Adjustable
- Confirmed a hierarchy of sciences
 - But with greater heterogeneity at the field level
- Observed minor temporal changes
- Relative age matters
- Differences by self-citation



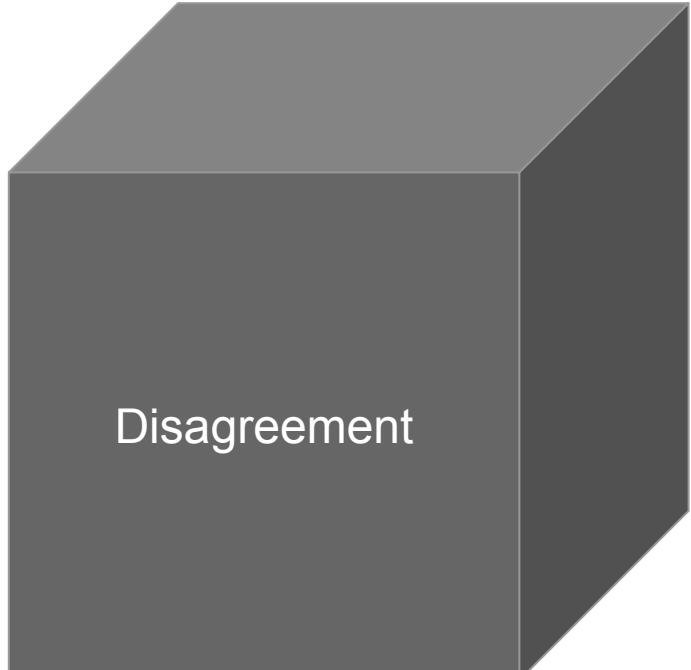
In Summary

- Developed a framework for identifying instances of disagreement
 - Robust
 - Transparent
 - Adjustable
- Confirmed a hierarchy of sciences
 - But with greater heterogeneity at the field level
- Observed minor temporal changes
- Relative age matters
- Differences by self-citation
- Disciplinary heterogeneity in word use requires consideration in scientific text analysis



Now that we have an indicator of disagreement

- Answer questions about how disagreement differs based on author characteristics
 - Gender
 - Race
 - National affiliation
 - Career stage
- Incorporate into traditional scientometric analysis
 - How does disagreement relate to citation impact?
 - Are novel or interdisciplinary papers more likely to be disagreed with?
 - Does team size relate to disagreement?



Disagreement

More soon!

Thank you!



Dakota Murray
Indiana University Bloomington
dakmurr@iu.edu



Wout Lamers
CWTS Leiden University
w.s.lamers@cwts.leidenuniv.nl



Kevin Boyack
SciTech Strategies
kboyack@mapofscience.com



Vincent Larivière
Université de Montréal
vincent.lariviere@umontreal.ca

@dakotasmurray

dakotamurray.me

Slides at:



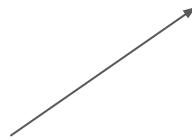
Cassidy R. Sugimoto
Indiana University Bloomington
sugimoto@indiana.edu



Nees Jan van Eck
CWTS Leiden University
ecknjpvan@cwts.leidenuniv.nl



Ludo Waltman
CWTS Leiden University
waltmanlr@cwts.leidenuniv.nl



Supporting slides

Signal terms

Signal term	Variants	Specific negations	Results
challenge*			405,613
contradict*			115,375
contrast*			1,257,866
contrary			171,711
conflict*			212,246
disagree*	“not agree*”, “no agreement”	“range”, “scale”, “kappa”, “likert”, “agree*” and/or “disagree” within a ten-word range of each other.	52,615
differ*		“different*”	2,003,677
controvers*			154,608
disprov*		“prove*” and “disprove*” within a ten-word range	2,938
refut*		“refutab*”	10,322
debat*		“parliament* debat*”, “congress* debat*”, “senate* debat*”, “polic* debat*”, “politic* debat*”, “public* debat*”, “societ* debat*”	150,617
no consensus	“lack of consensus”	“consensus sequence”, “consensus site”	16,632
questionable			24,244

Filter terms

studies	Studies; study; previous work; earlier work; literature; analysis; analyses; report; reports
ideas	idea*; theory; theories; assumption*; hypothesis; hypotheses
methods	model*, method*, approach*; technique*
results	result*; finding*; outcome*; evidence; data; conclusion*; observation*

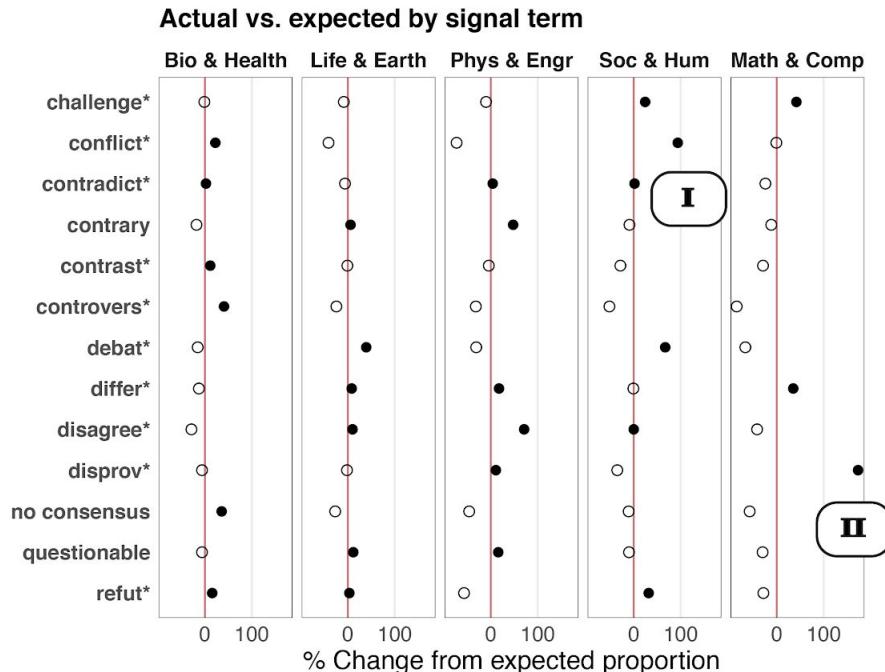
Representation

The expected (red line) vs. actual (dot) number of citations

*conflict** more common in Social Sciences

*disprov** more common in Maths

*disagree** more common in Physics & Engineering

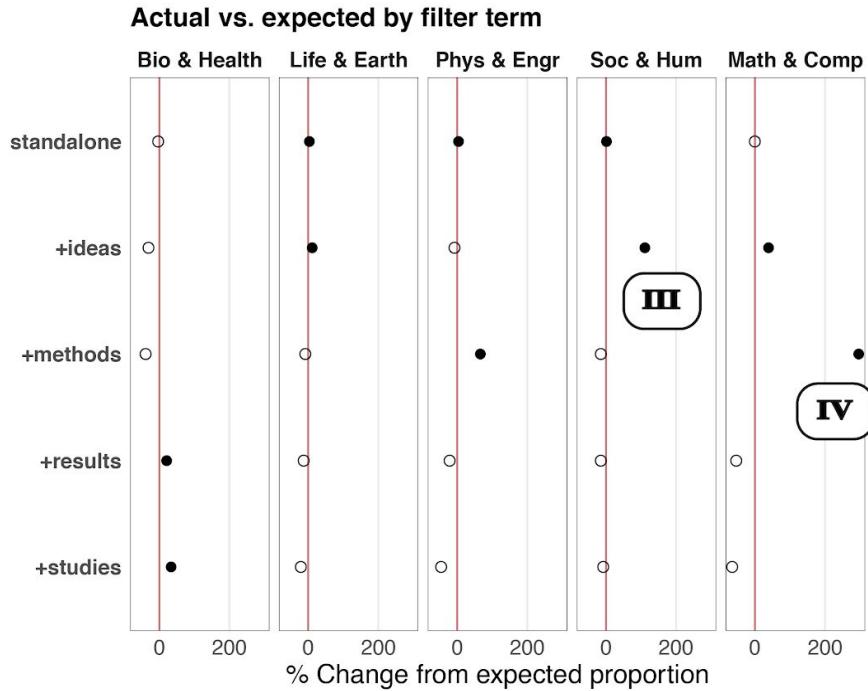


Representation

Same for filter terms

+ideas over-represented in Social Sciences and Humanities

+methods over-represented in Math & Comp Sci, also Physics and Engineering

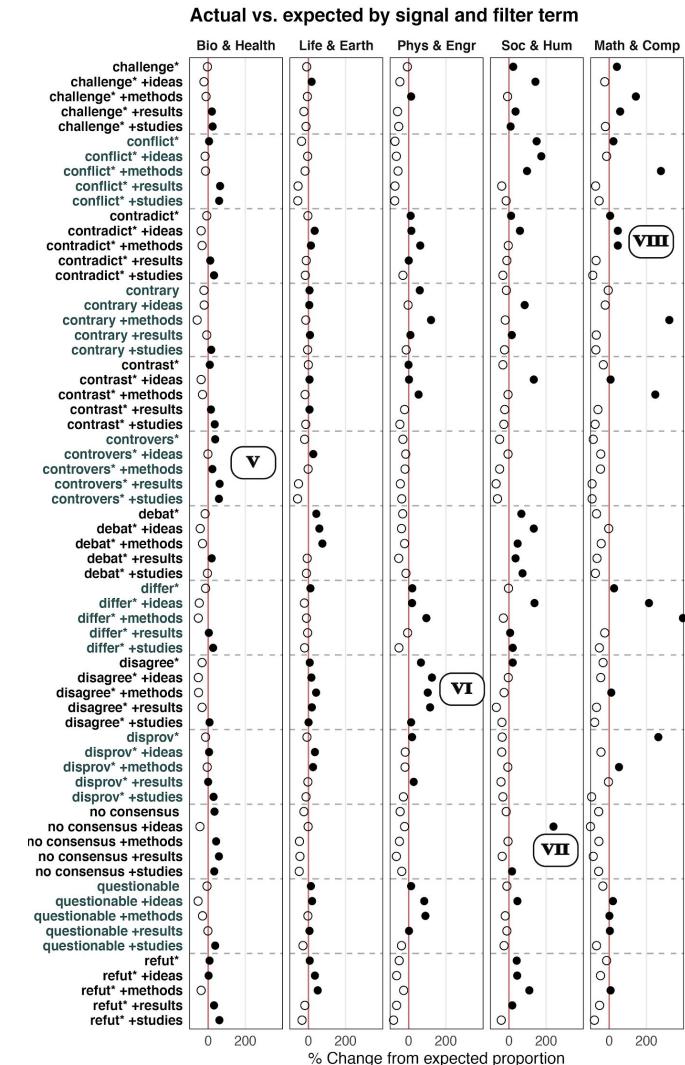


Representation

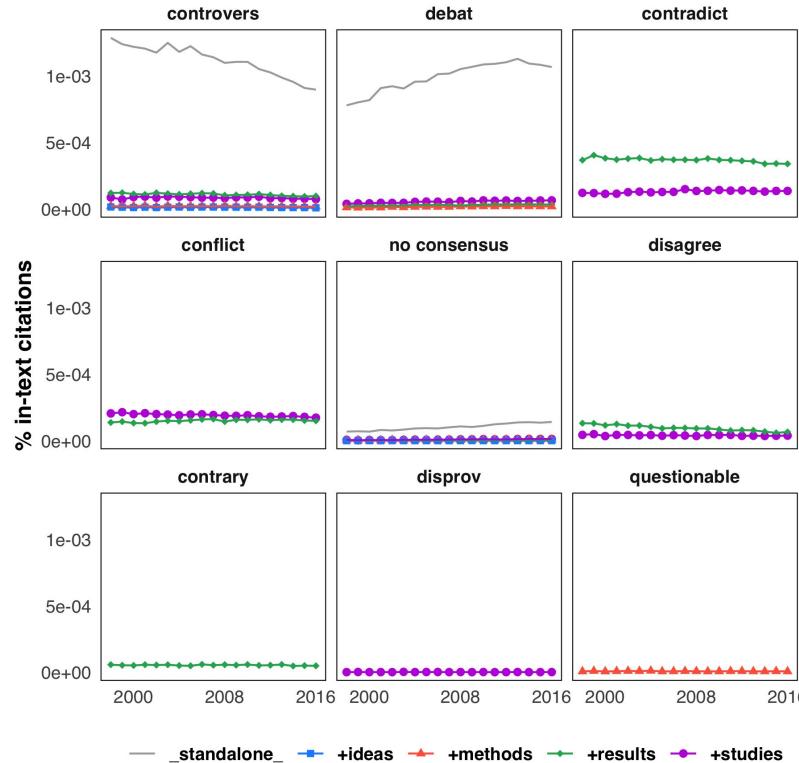
But there is huge heterogeneity

*controversy** over-represented in Bio & Health sciences, except with *+ideas*

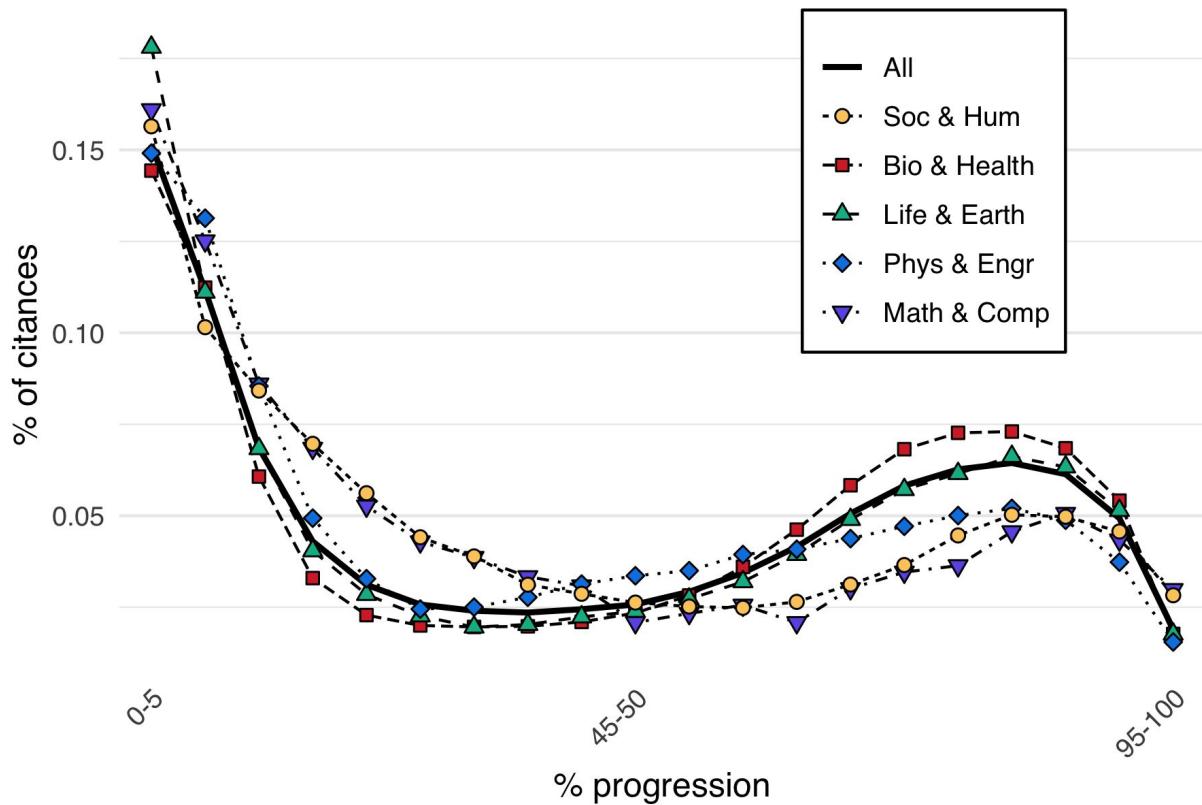
*contradict** over-represented in Math & Comp Sci, except with *+results* and *+studies*



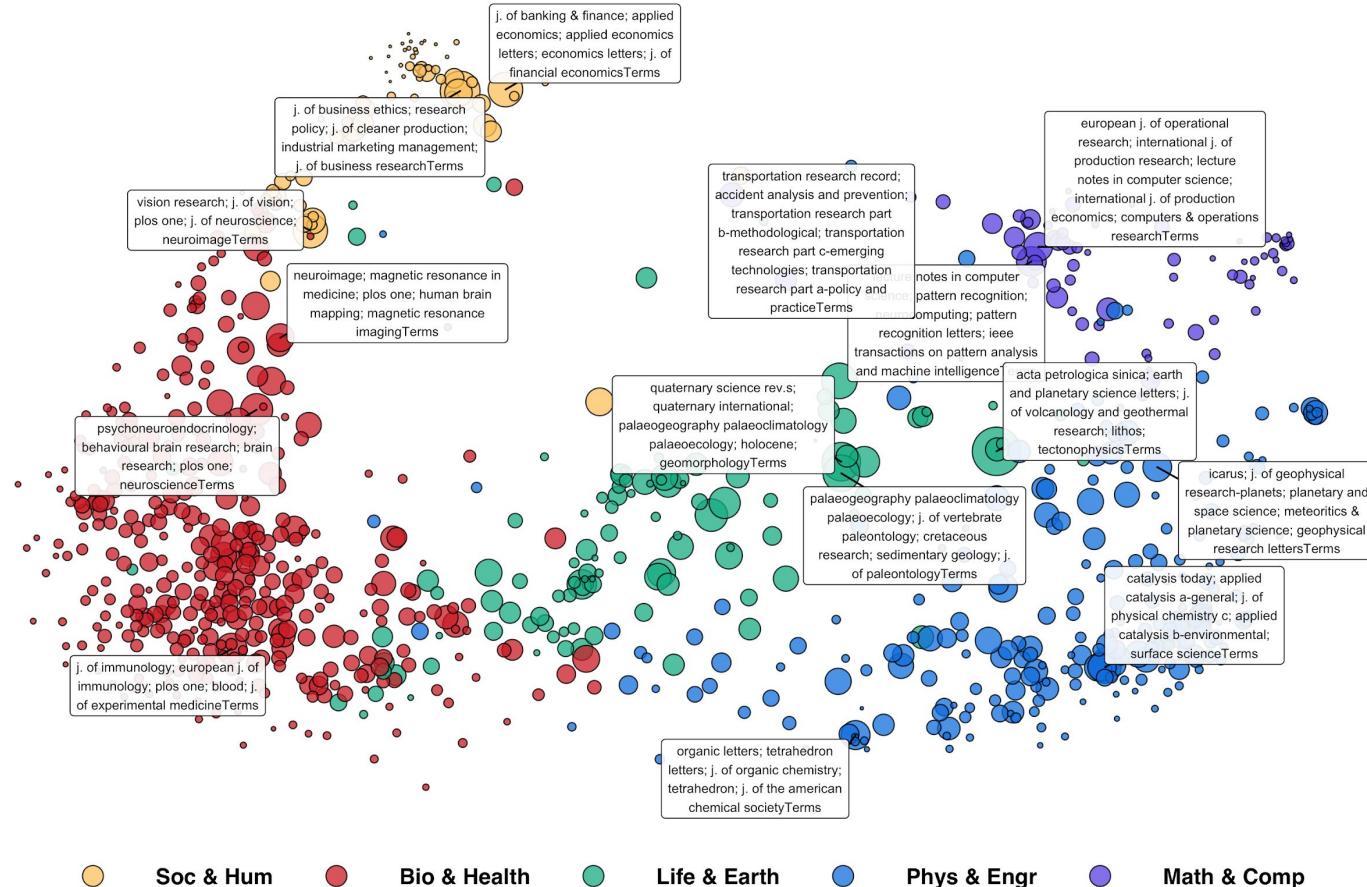
Different temporal dynamics by query



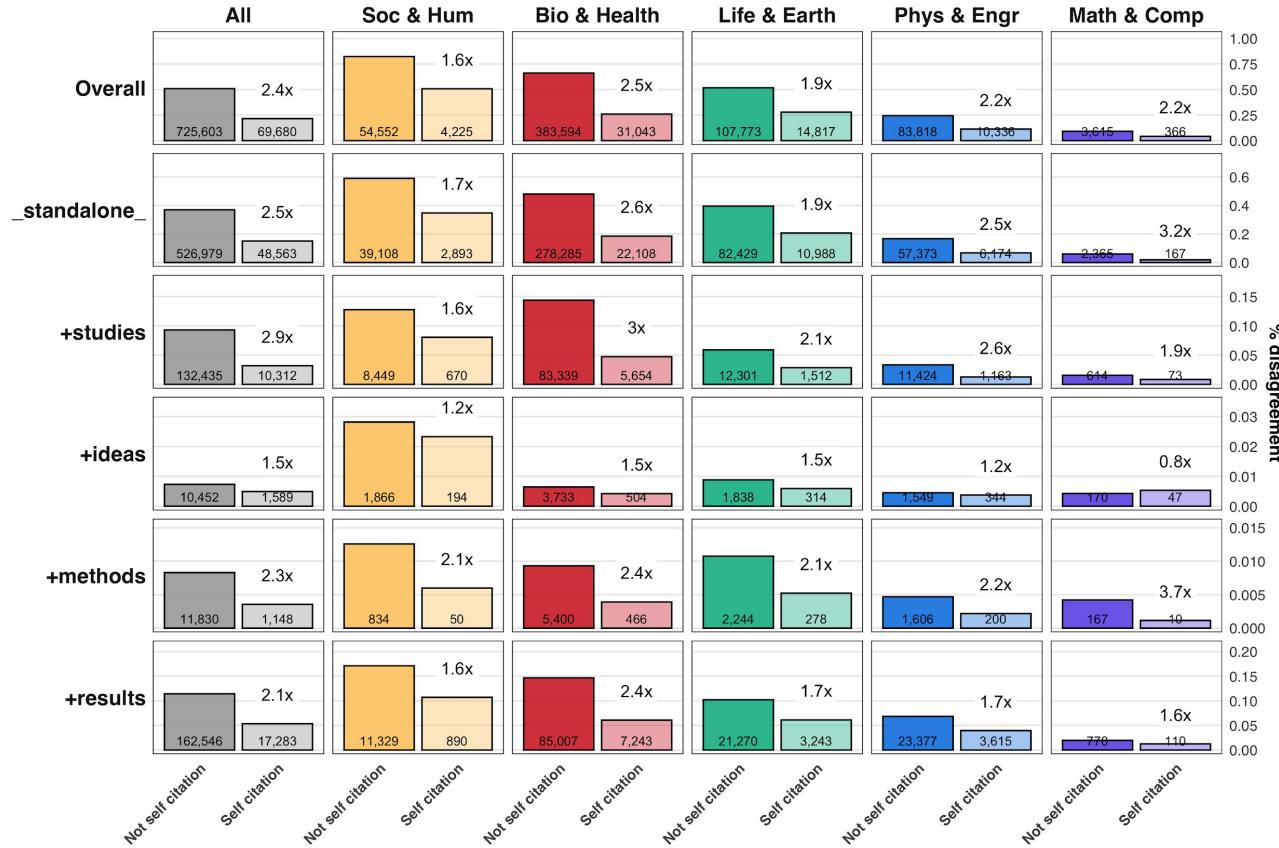
Text progression



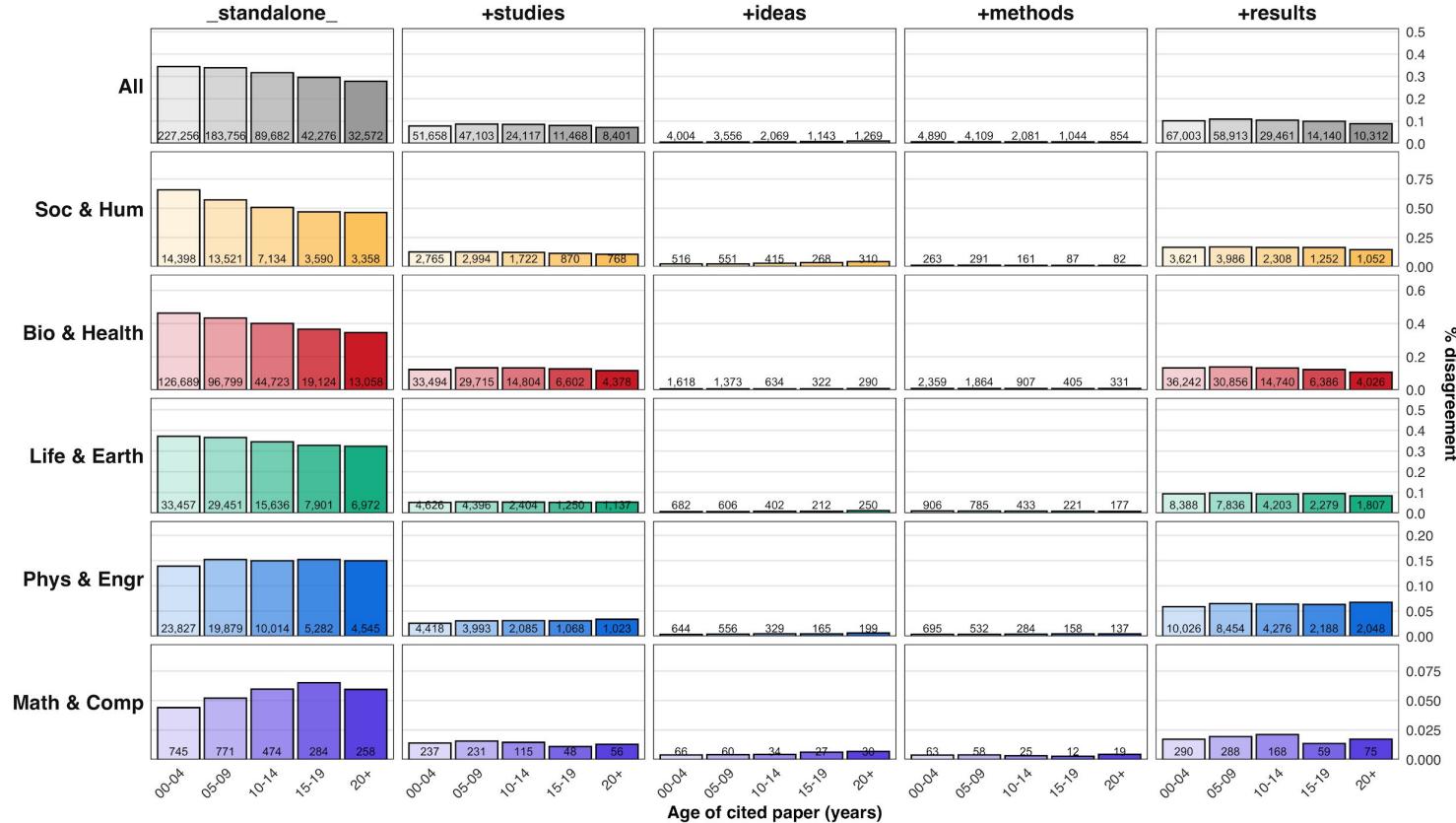
Reference meso-field map



Self-citation and filter term



Relative age and filter term



Disagreement and gender

Based on gender of the author of the citing paper

Not really any convincing evidence of an effect

