Measuring "Consensus" Across Scientific Fields

Short answer: yes, there is an active (if scattered) research community doing exactly this, and several concrete measures exist, though the question is harder and less settled than it first appears. The work sits at the intersection of a few communities, and the field-by-field comparison you sketch is more tractable than it used to be, but it is still an open research programme rather than a solved problem.

The relevant research communities

- Science of Science / Scientometrics. Journals like *Science of Science*, *Scientometrics*, and *Journal of Informetrics*, plus groups around the Santa Fe Institute, the Allen Institute for AI, and Semantic Scholar, study the large-scale structure of the literature. This is where most quantitative work lives.
- Philosophy of science. Kuhn's distinction between "normal science" (a shared paradigm, high consensus) and "revolutionary science" (competing paradigms, low consensus) is the classic conceptual frame. Lakatos's "research programmes" and Feyerabend's scepticism about incommensurability give qualitative vocabularies. The community rarely proposes a number, but it defines what "consensus" *means* well enough to be operationalised.
- Complex systems / network science applied to science. People at various places treat citation, co-authorship, and topic networks as complex networks and ask structural questions about them.
- Machine-learning / NLP for bibliometrics. Teams at Stanford (Aronow, Johnson), the Allen Institute, and Semantic Scholar use embeddings and topic models to map "where a field is" in semantic space.

Concrete measures that have been proposed or used

None of these is a single agreed-upon "consensus index," but several capture pieces of the idea:

1. Citation-network topology. Build a directed graph of papers in a sub-discipline and ask: how dense is the core? How high is the *modularity* (i.e., does the field split into tightly clustered sub-communities that barely cite each other)? A field with low modularity and a tight core looks more "consensual" than one that fragments into disconnected clusters. Louvain or Leiden community-detection algorithms on citation or co-citation graphs are the usual machinery. Aronow & Johnson (2014, *Machine Learning*) showed that ML "subfields" formed nearly disconnected citation clusters, a vivid picture of low consensus.

2. Semantic fragmentation / topic entropy. Run a topic model (LDA, BERTopic, etc.) over the titles and abstracts of every paper in a field. A field whose literature is spread over

many near-equal-weight topics has high *topic entropy*; one dominated by a few dominant topics looks more unified. Huang, Lu, and Jones (Nature 2020, "The semantic structure of science") mapped ~180 million papers into a semantic space and measured how "spread out" or "clustered" different disciplines are relative to each other.

3 Core-citation concentration. What fraction of all citations in a field points to a small set of "canonical" papers? A tight Pareto-like core suggests a shared body of knowledge; a flatter distribution suggests many parallel research programmes. This is loosely related to Kuhn's idea of a paradigm.

4 Embedding-space dispersion. Represent each paper as a vector (via sentence-BERT, SciBERT, etc.) and compute the *variance* or *effective dimensionality* of the cloud of vectors for a field. A tighter cloud ≈ more agreement about "what this field is about." This is the most literal "are people talking past each other?" test.

5 Open-problem / survey measures. Qualitative but informative: ask 200 researchers in the field to name the ten most important unsolved problems. The *agreement coefficient* across lists (Jaccard overlap, Spearman correlation of rankings) is a rough consensus metric. This is rare because it is expensive, but a few surveys in physics and biology do something like this.

6 Replication / stability of results. The Open Science Collaboration (2015) and follow-ups showed that fields with lower replication rates (psychology, some areas of social science) also tend to have more methodological pluralism. This is an indirect proxy: a field that cannot reproduce its own results arguably lacks consensus on what counts as a result.

7 Disciplinary "distance" metrics. Huang et al.'s 2020 semantic map, and related work by the Semantic Scholar team, produce a distance between any two fields. You can ask not just "is neuroscience fragmented?" but "is the gap between computational neuroscience and molecular neuroscience larger than the gap between condensed-matter physics and materials engineering?"

The neuroscience vs. materials science vs. quantum-computing comparison

This is a genuinely interesting contrast, and the literature gives you enough machinery to attempt it, but you should expect the answer to be messy:

- Neuroscience is enormous and old, with deep sub-fields (molecular, cellular, systems, cognitive, social, clinical, computational, optogenetics, …) that often share very few references. You would expect high modularity, high topic entropy, and a somewhat diffuse core. But there is also a deep "trunk" (the neuron, the synapse, the action potential, the cortical column) that every sub-field cites, so the picture is not pure fragmentation.

- Materials science / condensed matter is also large but arguably more "engineered" into a

grid of sub-fields (metals, polymers, semiconductors, nanomaterials, …). The sub-fields are large but each is internally tight, and they share a common language (band theory, crystallography, spectroscopy). You might see *clustered* consensus: high consensus *within* sub-fields, lower consensus *between* them.

- Quantum computing / quantum information is small, young, and rapidly expanding. The community is small enough that citation networks are dense, but the field is so new that the "settled core" is tiny. You would probably see high density (everyone cites everyone) but low *topological maturity*—the field is not yet settled into the sub-structure that older fields show. In Kuhnian terms it may be in or near a "revolutionary" phase, where consensus is genuinely low but in a different way: the disagreement is about the *foundations* rather than about sub-topics.

A crucial confound: field size, age, and publication volume all bias every measure above. A tiny field will look "tight" for trivial reasons. A huge old field will look "fragmented" partly because it has had time to splinter. Any honest comparison needs to normalise for these, and there is no fully settled way to do that yet. You can stratify by publication counts, log-size, and age, but the residual comparison is still interpretively open.

Honest limitations

- There is no single agreed "consensus index." The measures above capture different facets (structural, semantic, social, epistemic) and can disagree. A field can be citation-dense but semantically heterogeneous.
- Citations are not beliefs. Citing a paper is not the same as agreeing with it; adversarial citations exist. This makes pure bibliometric measures imperfect proxies for "agreement."
- The unit of "a sub-discipline" is itself contested. Where does neuroscience end and computational biology begin? Where does quantum computing end and quantum chemistry begin? The boundaries you draw change the numbers.
- The philosophical literature insists that consensus is partly a *social* and *normative* phenomenon (what counts as a good argument, what counts as a solution), not purely a structural one, and is hard to extract from publication data.

Where to look next

- Huang, Lu, et al., "The semantic structure of science," *Nature* 589 (2020) and the follow-up papers in the same group.
- Aronow & Johnson, "Mapping and predicting research in machine learning," *Machine Learning* (2014) for the fragmentation idea.
- The Science of Science journal and the Scientometrics journal for the broader community.
- Kuhn, *The Structure of Scientific Revolutions* (1962, 1970) and Lakatos, *The Methodological Background of Scientific Research Programmes* for the conceptual frame that "consensus" and "paradigm" sit in.
- The AllenNLP / Semantic Scholar toolkits and the OpenAlex database (free, 250+

million works, rich metadata) for building your own field-level measures today without needing to scrape.

The short version: the tools exist, the questions are being asked by a small but growing community, and a field-by-field comparison of the kind you describe is very doable as an empirical exercise. What is not yet available is a single, validated, discipline-agnostic "degree of consensus" number that everyone would trust. Getting from a handful of complementary measures to something like that is itself the open research problem.