Robustness, Generative Entrenchment, and the Network Construction of Mathematical Knowledge

Anonymous & Anonymous

In a classic essay [Wim07], philosopher William Wimsatt presents a provocative sketch of evolving scientific activity that represents the relations between scientific facts (broadly defined to include objects, observations, theories, and laws) in the form of a directed network (Figure 1). Very loosely speaking, the nodes in this network represent facts and the arrows represent entailment (e.g., deductive or inductive) relations: an arrow from A to B means that fact A entails fact B.

Motivated by this metaphor, Wimsatt advances two claims about the nature of scientific activity. First, certain scientific facts will be *robust*, in that they are readily obtained or arrived independently from multiple directions. More robust facts, Wimsatt argues, are likely to be discovered earlier than their less robust counterparts, simply because there are more paths that lead to their discovery. Second, some scientific facts will become *generatively entrenched*. This means that such facts will come to form the bedrock upon which numerous downstream discoveries rest. Importantly, should these facts be called into question, large swaths of our scientific understanding would be at risk. As a result, scientists hew away from altering generatively entrenched elements—they are "too big to fail". Wimsatt poses the following conjecture: because robust nodes are likely to be discovered early on, they are prone to becoming generatively entrenched. The converse is not necessarily true. Together the notions of robustness and generative entrenchment help to make sense several otherwise confusing episodes in the history of science and of how certain scientific theories resist modification despite significant empirical limitations.

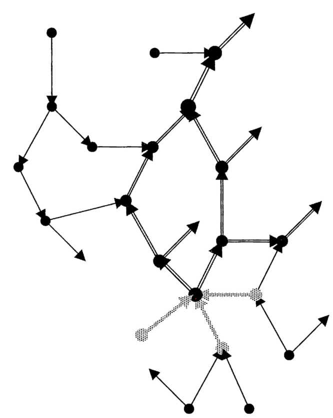


Figure 1: Wimsatt's famous network illustration highlights robust nodes, those pointed to by numerous upstream nodes, and generatively entrenched nodes, those that point to numerous downstream nodes. The original caption to this figure reads "Network with robust node (indicated by multiple dotted arrows) and entrenched by multiple downstream consequences (open arrows) that are thus more reliable than [generatively entrenched] network intersecting with it from the left." (Reproduced from [Wim07])

However, Wimsett is not explicit about how the constituents of his network diagram map to scientific truths or the artifacts of research activity. For instance, he does not provide clean heuristics for distinguishing the nodes of his diagram—scientific facts—from the directed links—entailments. Nor does he explicitly address the complex reality of inference in the natural sciences, where entailments may be deductive or inductive in nature.

Nonetheless, the concepts of robustness and generative entrenchment strike us as fundamental to understanding the construction of scientific knowledge, and deserving of explicit empirical treatment. To bring this about, we explore Wimsatt's ideas in the context of a knowledge system that is more amenable to more formal treatment than that of the natural sciences: formal mathematics. It also allows us to explore Wimsatt's ideas of robustness and entrenchment in a concrete setting that respects and reflects the actual process of scientific activity, consensus, and certainty.

While the boundaries of facts and entailments can be blurry in the natural sciences, their analogues in pure mathematics are more straightforward. In Wimsatt's framework, theorems and lemmas of mathematics correspond to Wimsatt's 'facts' while proofs thereof are descriptions of entailments that derive new theorems from preexisting ones. To assess the structure of mathematical networks of theorems and proofs, we analyze a massive ensemble of proofs and theorems as formalized the automated proof assistant and verifier framework `Lean` [DMKA^{+}15]. As of August 2025, `Lean`'s `mathlib4` project contains 229778 theorems and 112095 definitions.[1]

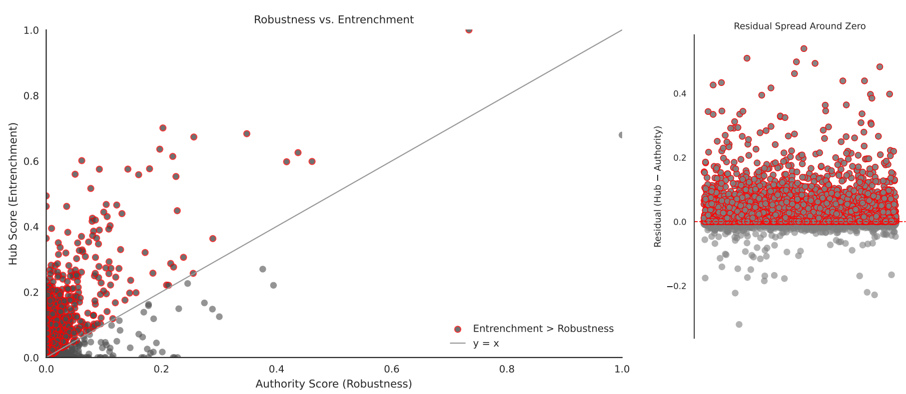
The second step is to instrumentalize Wimsatt's notions of robustness and generative entrenchment. One might follow Wimsatt in associating robustness with high in-degree and generation entrenchment with high outdegree, but doing so ignores the relative importance of the aossicated nodes. A more appropriate rendering draws upon the concepts

[1]https://leanprover-community.github.io/mathlib_stats.html

of *hubs* and *authorities* from Kleinberg's HITS algorithm [Kle99]. The HITS algorithm uses a recursive algorithm to pick out two specific kinds of nodes: hubs are those nodes which point to important authorities, and authorities are those which are pointed to by important nodes. In Wimsatt's terms, authorities are robust because many important hubs point to them and thus they are readily reached through movement on the network. Hubs are generatively entrenched because they point to important nodes that depend on them for their importance and, thus, lie upstream of numerous significant network components.

Methods and results: We analyzed a large, open source, dataset of Lean declaration (theorems, lemmas, definitions, etc.) relations representing hundreds of thousands of hours of scientific activity and used this to test Wimsatt's hypothesis—that the network robustness of a scientific fact leads to the generative entrenchment of that fact, but not visa versa. We induce a dependency graph of Lean declarations in which, for example, if theorem B depends on theorem A, then a directed edge goes from A to B. We then calculate HITS on this graph, quantifying robustness as authority score and generative entrenchment as hub score. We find strong support for Wimsatt's conjecture that robust nodes—declarations such as theorems, in our case—will be generatively entrenched because they are established early and thus disproportionately likely to be used in deriving downstream results. However (and as Wimsatt predicted) the converse is not the case: generatively entrenched results need not be robust in the sense of being readily obtained from multiple independent directions. Our results (Figure 2) provide the first empirical evidence supporting this influential frame in the philosophy of science and provide meaningful, *prima facie* evidence for the claim that robustness and generative entrenchment are characteristic of evolved systems.

Our findings empirically anchor a long-standing philosophical account of how natural and conceptual systems evolve. By showing that robustness and generative entrenchment can be operationalized and observed in a large, living body of mathematics, we provide evidence that these concepts capture deep, structural features of knowledge systems. This also opens the door to comparative studies across domains, offering a framework for understanding why some ideas persist, others fade, and how knowledge systems become either resilient or fragile. If we want to understand how science builds, stabilizes, and sometimes stumbles, we need to take robustness and entrenchment seriously.



(a) Panel A: Authority vs. Hub.

(b) Panel B: Residuals.

Figure 2: (a) shows authority (x-axis) vs. hub scores (y-axis) for Lean declarations. Points above the line are more entrenched than robust. (b) shows the distribution of residuals. Taken together, these offer empirical support for Wimsatt's theory.

References

[DMKA+15] Leonardo De Moura, Soonho Kong, Jeremy Avigad, Floris Van Doorn, and Jakob von Raumer. The lean theorem prover (system description). In *International Conference on Automated Deduction*, pages 378–388. Springer, 2015.

[Kle99] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[Wim07] William C Wimsatt. Robustness and entrenchment: How the contingent becomes necessary. *Re-engineering philosophy for limited beings: Piecewise approximations to reality*, pages 133–145, 2007.