Molecular Biology and Evolution, 2025, **42**, 1–11
https://doi.org/10.1093/molbev/msaf148
Advance access publication 9 June 2025
Discoveries

MBE OXFORD

Structural Changes in Gene Ontology Reveal Modular and Complex Representations of Biological Function

Sergi Valverde [1,2,3,*] **Blai Vidiella** [4] **Gemma I. Martínez-Redondo** [5] **Salva Duran-Nebreda** [1] **Rosa Fernández** [5] **Aureliano Bombarely** [6] **Ana M. Rojas** [7] **R. Alexander Bentley** [2,8]

[1]Evolution of Networks Lab, Institute of Evolutionary Biology (CSIC-UPF), Passeig Marítim de la Barceloneta 37-49, Barcelona 08003, Spain
[2]Center for the Dynamics of Social Complexity (DySoC), University of Tennesee, Knoxville, TN 37996, USA
[3]Science of Complexity, European Centre for Living Technology (ECLT), Ca' Bottacin, Dorsoduro 3911, Venice 30123, Italy
[4]Centre for Biodiversity Theory and Modelling, Theoretical and Experimental Ecology Station, CNRS, Moulis, France
[5]Metazoa Phylogenomics and Genome Evolution Lab, Institute of Evolutionary Biology (CSIC-UPF), Passeig Marítim de la Barceloneta 37-49, Barcelona 08003, Spain
[6]Bioinformatics and Evolutionary Genomics, Institute of Molecular and Cellular Biology of Plants (CSIC-UPV), Valencia, Spain
[7]Computational Biology and Bioinformatics, Andalusian Center for Developmental Biology (CABD-CSIC), Sevilla 41013, Spain
[8]Department of Anthropology, University of Tennessee, Knoxville TN 37996, USA

***Corresponding author**: E-mail: s.valverde@csic.es.
Associate editor: Haipeng Li

Abstract

The Gene Ontology is a central resource for representing biological knowledge, yet its internal structure is often treated as static—or as a black box—in computational analyses. Here, we examine 15 years of Gene Ontology evolution using network-based methods, revealing that Gene Ontology changes not only through incremental growth but also through punctuated, curator-driven restructuring. In particular, we document a major reorganization of the Cellular Component branch in 2019, where broad "part" terms were removed and the ontology was modularized into distinct domains for anatomical entities and protein-containing complexes. Semantic modularity aligns Gene Ontology with emerging frameworks such as the Common Anatomy Reference Ontology and Gene Ontology-Causal Activity Modeling, but also disrupts similarity metrics that rely solely on hierarchical proximity. More broadly, the restructuring of the cellular components branch consolidates a shift toward treating Gene Ontology as a multi-layer semantic network—a transformation rooted in a decade-long process of scientific and social consensus across institutions. These findings underscore the need for version-aware, multi-layer models to ensure reproducibility and interpretability—and to better represent biological function across compositional, spatial, and regulatory dimensions as ontologies continue to evolve.

Keywords: gene ontology, computational biology, functional annotation, scale-free networks, cultural evolution, bioinformatics

Introduction

Accurately defining gene function is essential for genomics, biomedicine, and evolutionary biology, but experimental methods for functional characterization remain costly and time-intensive. As a result, computational approaches have become increasingly important for predicting gene function, prioritizing candidate genes for study, and uncovering novel protein interactions. A common strategy is to infer gene function from sequence similarity, leveraging the Gene Ontology (GO)—a structured, expert-curated database that categorizes gene functions into molecular functions (MF), biological processes (BP), and cellular components (CC) (Ashburner et al. 2000; Alterovitz et al. 2010; Consortium et al. 2023). In previous decades, molecular events were characterized from journal abstract databases by text-mining for the textual trigger, event type and participating entities (Kim et al. 2009; Sarafraz et al. 2009).

GO serves as the foundation for many bioinformatics tools that annotate gene function based on sequence homology and evolutionary conservation. For example, *eggNOG-mapper* assigns GO terms to previously uncharacterized genes using orthology inference (Cantalapiedra et al. 2021), while machine learning models such as *FANTASIA* leverage large-scale biological datasets to refine functional predictions (Barrios-Núñez et al. 2024; Martínez-Redondo et al. 2024). These tools accelerate biological discovery but rely on an implicit assumption: that GO provides a stable and accurate framework for functional annotation. However, as biological knowledge expands, GO itself must evolve, raising fundamental questions about how its structural changes influence gene function predictions over time. As both a biological classification system and a curated knowledge infrastructure, the GO encodes not just biological facts but also the evolving agreements and priorities of the research community.

GO is not a static reference; it is a semantic network (see Fig. 1) that grows and reorganizes in response to new research and technological findings. In this network, nodes represent biological concepts (e.g. "ATP binding" or "cell division"), and edges encode hierarchical ("is-a") or compositional ("part-of") relationships. In software development, engineers

Received: February 13, 2025. **Revised:** May 16, 2025. **Accepted:** May 28, 2025

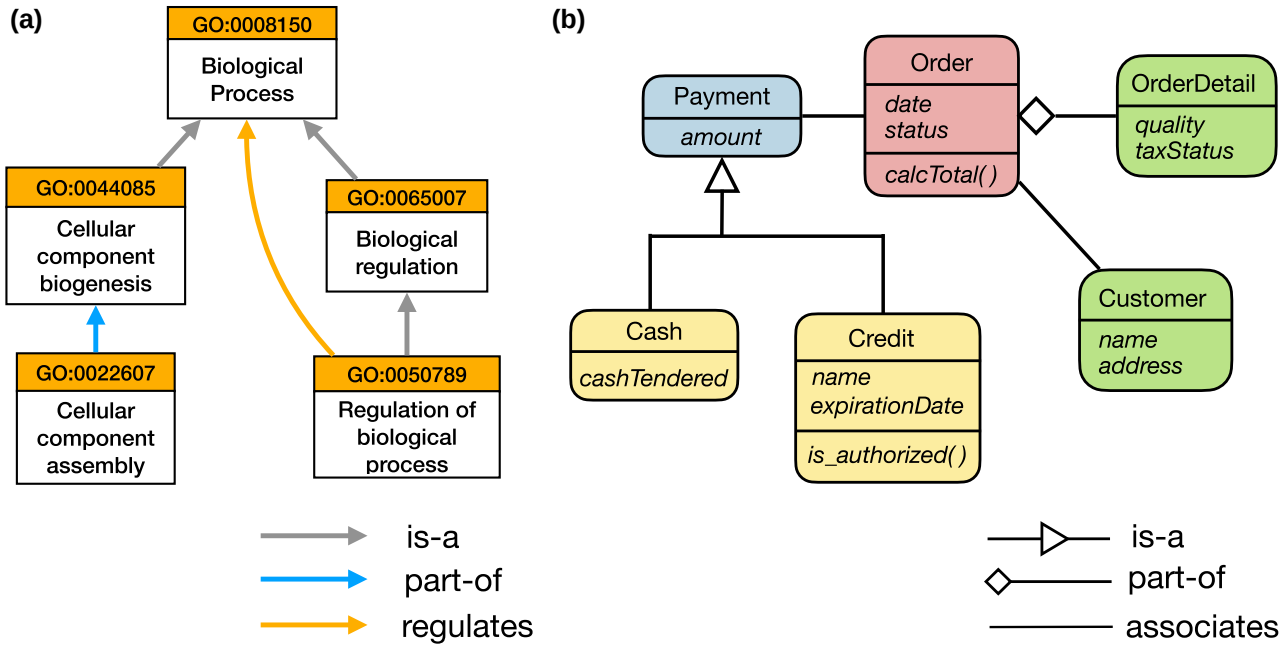

Fig. 1. Two types of semantic networks. a) The GO organizes biological concepts into three main categories (or branches)—biological processes, cellular components, and molecular activities—using a structured vocabulary (e.g. biological regulation, cell projection assembly). Terms within each category are hierarchically linked through relationships such as "is-a", "part-of", and "regulates", forming a directed acyclic graph that represents biological function. b) Class diagrams in software engineering represent the structure and behavior of software systems, where classes (colored nodes) define software components, and edges represent associations, aggregations ("part-of"), compositions, and inheritance hierarchies ("is-a"). Although GO and software ontologies differ in purpose and domain, they share key topological properties, including hierarchical organization, preferential attachment, and scale-free connectivity. These similarities suggest that both systems evolve through a combination of incremental reuse and large-scale restructuring, mechanisms commonly observed in cultural and technological evolution (see text).

frequently modify and refine existing classes, while occasionally performing global reorganizations, i.e. refactorings (Burek et al. 2017; Fowler 2018), to improve system clarity and reduce redundancy. Similarly, GO evolves through the continuous addition of new terms, the merging or removal of outdated categories, and systematic ontology revisions led by expert biocurators. Studies on semantic and technological networks reveal that hierarchical organization (Steyvers and Tenenbaum 2005), preferential attachment (Valverde and Solé 2005b), and rich-club behavior (Valverde and Solé 2007a) are hallmarks of culturally evolving systems (Valverde and Solé 2005a; Solé and Valverde 2020). These same patterns appear in GO's structural evolution, suggesting that its organization is not purely a biological necessity but also a product of scientific and social constraints.

Understanding the evolutionary principles underlying GO's structure is critical for improving functional annotation accuracy, gene similarity metrics, and predictive models in computational biology. As GO evolves, these structural changes directly affect gene function studies, biomedical research, and evolutionary analyses. By applying network analysis to 15 years of GO versions (2009 to 2024), we reveal that its structure exhibits scale-free properties, hierarchical refinement, and rich-club behavior. Our findings suggest that GO's evolution is shaped by both bottom-up reuse and top-down restructuring, similar to other culturally evolving knowledge systems. By characterizing these network patterns, we offer new insights into how biological ontologies develop and propose strategies for improving the stability and interpretability of functional annotations.

Results

To assess how GO's network structure evolves over time, we reconstructed Gene Ontology Networks (GONs) for 40 versions spanning 15 years (2009 to 2024) (see Methods). We analyzed key topological features, focusing on degree distributions, hierarchical refinement, and large-scale restructuring events that shape functional annotation.

Scale-Free Structure Suggests That GO Evolves Through Duplication and Rewiring

The GO continuously expands and reorganizes in response to new biological discoveries, yet its structural evolution follows non-random patterns. One of the most striking features of GO's architecture is that its connectivity follows a scale-free distribution, a hallmark of self-organizing knowledge systems. At a local network level, the in-degree (k_i^{in}) and out-degree (k_i^{out}) of a node represent the number of incoming and outgoing connections, respectively (see inset, Fig. 2d). The in-degree distribution $P(k)$, which describes the probability that a randomly chosen node has k incoming connections, provides insight into the heterogeneity of GO's structure. In many biological and technological networks, this distribution follows a power law,

$$P(k) \sim k^{-\alpha}, \quad (1)$$

indicating that a few highly connected terms act as central hubs, while most terms have fewer connections. To obtain more accurate estimates of the exponent α, we fit the cumulative in-degree distribution:

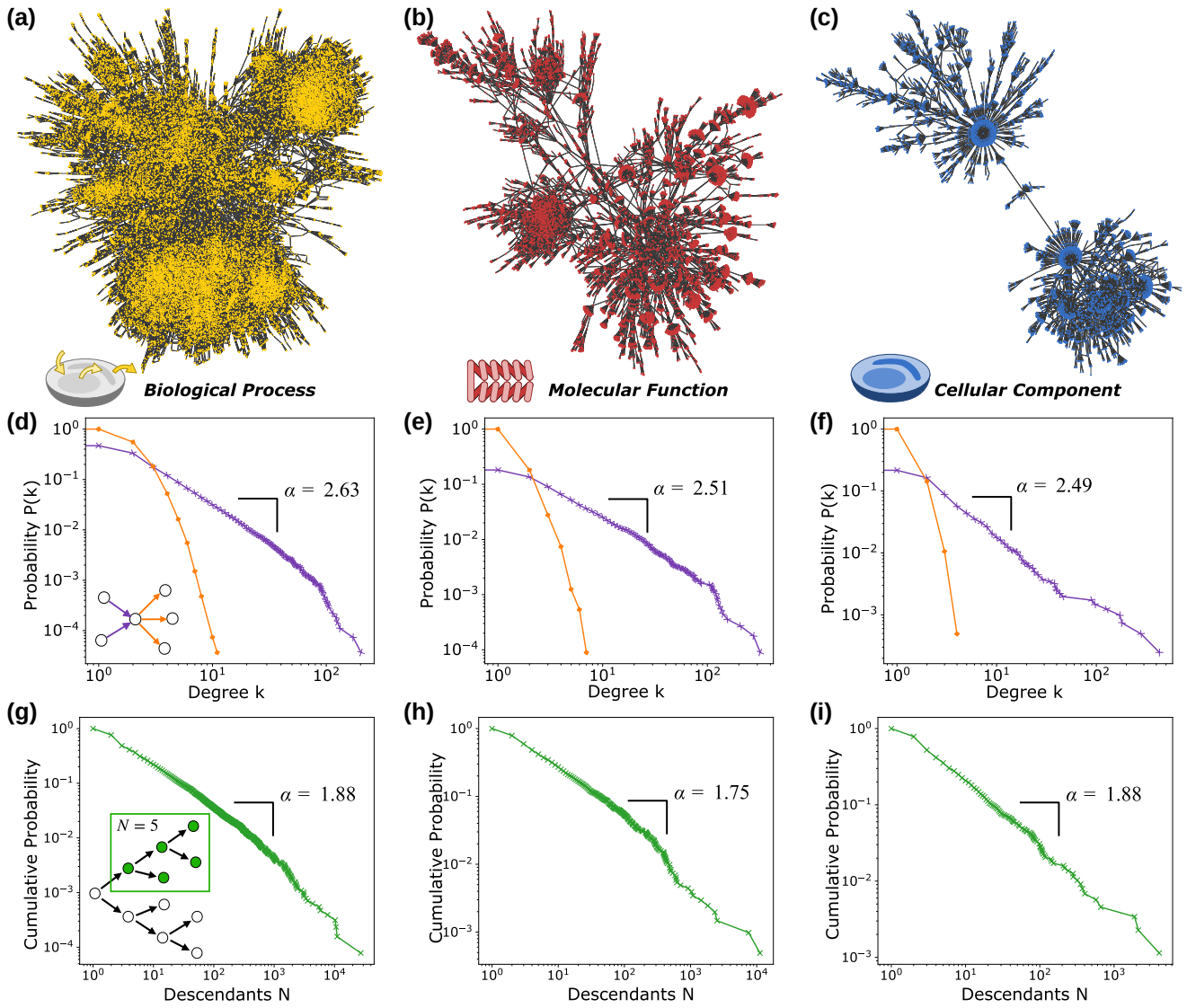

Fig. 2. Structural properties of ontology networks. The GO consists of three directed networks that reflect the functional domains: a) biological processes (G_{bp}), b) molecular functions (G_{mf}) and c) cellular components (G_{cc}). From left to right, the average degree for each network is $k_{bp} = 1.8$, $k_{mf} = 1.21$, and $k_{cc} = 1.15$ (see text). Each node in the ontology can be characterized by the number of incoming (in-degree) and outgoing (out-degree) links (see inset in panel d). The empirical in-degree k distribution follows a power-law $P(k) \sim k^{-\alpha}$ (violet curve) with exponent d) $\alpha = 2.63$ for the BP network, e) $\alpha = 2.51$ for the MF network, and f) $\alpha = 2.49$ for the CC network. The out-degree (orange curve) distribution is exponential for all networks. The distribution of number of node descendants D is a power law, $P(D) \sim D^{-\gamma}$ (green curve) with exponent g) $\gamma = 1.88$ for the BP network, h) $\gamma = 1.75$ for the MF network and i) $\gamma = 1.88$ for the CC network. These ontology networks relate to the GO version from early 2024.

$$P_{>k} = \int_0^\infty P(k)dk \sim k^{-\alpha+1}. \quad (2)$$

Our analysis (Fig. 2) reveals that all three GO networks (MF, BP, and CC) exhibit a power-law in-degree distribution with an exponent close to $\alpha = 2.5$, and thus suggesting a high degree of heterogeneity despite different average degrees. This confirms that GO is a highly heterogeneous, scale-free network, where some terms occupy disproportionately central positions. However, this centrality does not necessarily reflect biological importance–it may instead arise from curatorial decisions and preferential reuse of existing terms (Petersen et al. 2012; Newberry et al. 2017; Bentley et al. 2021; Petersen et al. 2025).

The emergence of scale-free networks in GO is unlikely to be solely a consequence of biological function. Instead, it mirrors universal principles of cultural evolution—the same mechanisms that drive the development of technological and semantic systems, including software networks. A comparison between the degree distributions of GO (Fig. 2) and semantic software networks (see Fig. 2 in Valverde and Solé 2005a) suggests that both systems evolve through reuse and incremental modifications, reinforcing the idea that cultural factors shape GO's structure. Preferential attachment, a well-known mechanism in network evolution, may explain the emergence of scale-free topology (Barabási and Albert 1999). Even in the absence of selection, a generative model based on duplication and rewiring accurately predicts not only the observed in-degree distributions in software networks but also the frequency of common network motifs (Valverde and Solé 2005b). Given the statistical similarity between GO and software networks, we hypothesize that GO's evolution is largely driven by similar generative mechanisms, rather than being purely constrained by biological function.

Many biological networks, including gene regulatory networks, protein-protein interaction networks, and metabolic

pathways, exhibit scale-free properties across diverse taxa, from yeast to mammals (Albert 2005; Sonawane et al. 2019). Given this, one might expect that GO's network structure simply reflects the underlying organization of biological systems. However, our findings suggest an alternative explanation. First, GO aggregates knowledge from multiple biological domains, each with its own connectivity patterns. If GO merely reflected biological networks, we would expect multiple independent sub-networks rather than a unified scale-free structure. Second, GO exhibits statistical properties nearly identical to other human-designed semantic networks, including software ontologies and knowledge graphs (Valverde and Solé 2005a; Solé et al. 2011). And finally, unlike biological networks, GO undergoes direct curatorial intervention, with ontology editors making top-down modifications that influence term connectivity. Thus, while GO serves as a map of biological function, its structural evolution follows cultural besides biological constraints. The fact that the in-degree exponent fluctuated until the end of 2019 (Fig. 4) suggests that GO's architecture is still evolving actively, incorporating both incremental changes and large-scale reorganization.

Tracing The Global Influence Of Terms Through The Distribution Of Descendants

Generative mechanisms like reinforcement and duplication can explain the emergence of scale-free properties in GONs, but they do not fully account for their global complexity. In particular, large-scale, coordinated changes in GONs suggest that some terms exert influence beyond their immediate connections. To quantify this effect, we analyze the descendant distribution $P(D)$, which measures the probability that a randomly chosen term has D descendant nodes reachable through a chain of directed links. This distribution provides insight into how information propagates within the ontology and how structural constraints influence term evolution.

In random networks, descendant distributions typically follow an exponential form, reflecting a uniform and constrained branching process. However, in hierarchical systems with preferential attachment, the distribution tends to follow a power-law $P(D) \sim D^{-\gamma}$, with γ typically between 1 and 2, indicating that a few highly connected nodes dominate the spread of information while most nodes have limited influence (Steyvers and Tenenbaum 2005). To quantify this in GONs, we compute the cumulative distribution of descendants, which smooths fluctuations in $P(D)$ and improves statistical robustness:

$$P_{>D} = \int_0^\infty P(D)dD \sim D^{-\gamma+1} \quad (3)$$

Our analysis of GONs (see Fig. 2) reveals that the descendant distributions for BP (see Fig. 2g) and CC (see Fig. 2h) networks exhibit exponents close to $\gamma = 1.88$, whereas the MF (see Fig. 2i) network has a lower exponent of $\gamma = 1.75$. This suggests that the biological process and cellular component ontologies have a more balanced propagation of influence across terms, while the molecular function ontology is more skewed toward a few highly central terms.

These differences likely reflect underlying domain-specific constraints. The BP and CC networks describe broad biological categories that encompass multiple interdependent pathways, necessitating a more even distribution of relationships. In contrast, molecular functions tend to be more modular, with a handful of highly reused functional categories (e.g. enzyme activities) acting as key hubs. This structural disparity may impact the stability of GO-based functional predictions: changes in a few high-degree nodes within MF could disproportionately alter large portions of the ontology, whereas changes in BP or CC may be more distributed across multiple terms (see Fig. 2).

From a cultural evolution perspective, the observed heterogeneity in descendant distributions aligns with mechanisms of reuse and conceptual refinement. Broad biological processes and cellular components evolve incrementally, with new terms emerging as specializations of existing categories. Molecular functions, by contrast, may be subject to more abrupt shifts driven by technological advances in functional annotation or experimental discoveries (Consortium et al. 2023). These structural properties should be considered when using GO for predictive modeling, as variations in term influence can affect functional annotation consistency over time.

Rich-club Behavior Reveals Non-local Semantic Reorganization

To accurately represent current biological knowledge, the GO is continuously revised and curated (Aleksander et al. 2023). Structural modifications in GO generally fall into three categories: (i) the creation of new terms to capture previously unrepresented concepts; (ii) the replacement of obsolete or inconsistent terms; and (iii) the merging of semantically overlapping terms to streamline the ontology. While many of these changes reflect advances in biological understanding, they are also shaped by systematic curation processes, including internal reviews, quality control checks, and expert feedback.

These revisions manifest in the temporal dynamics of the number of terms $N(t)$ and the number of links $L(t)$ (Fig. 3), as well as the exponents of the topological distributions (Fig. 4), which vary across the three GONs. Notably, the cellular component (CC) network underwent a significant transformation between 2019 and 2020. During this period, the number of links dropped sharply (Fig. 3f), despite 96% of the terms being retained (Fig. 3e). This loss corresponds to the removal of a large high-level subgraph built around compositional categories such as (CC:cell part), (CC:membrane part), and (CC:organelle part)—structures that had historically organized the upper tiers of the ontology (Fig. 3g). Their removal marked a deliberate restructuring event that simplified top-level complexity and introduced a new model of semantic organization.

Fluctuations in Fig. 4 suggest that large-scale reorganizations are transitory. The exponent of the in-degree and descendant distributions fluctuates over time, suggesting that editing efforts produce short-term perturbations rather than persistent structural trends. Still, some reorganizations—particularly those involving upper-level terms—can produce long-term effects detectable by global network metrics, e.g. through the evolution of the rich-club coefficient in the CC network. As shown in Fig. 5a, a sudden increase in the rich-club coefficient appears by the end of 2019, coinciding with the ontology's restructuring. The emergence of a rich-club structure—a densely interconnected core of high-degree nodes—is further reflected in the in-degree distribution, which displays a noticeable hump during this period (Fig. 5d), consistent with a shift toward greater network centralization (Valverde and Solé 2007b).

This rich-club structure is not merely a topological consequence but reflects deeper semantic design decisions. By

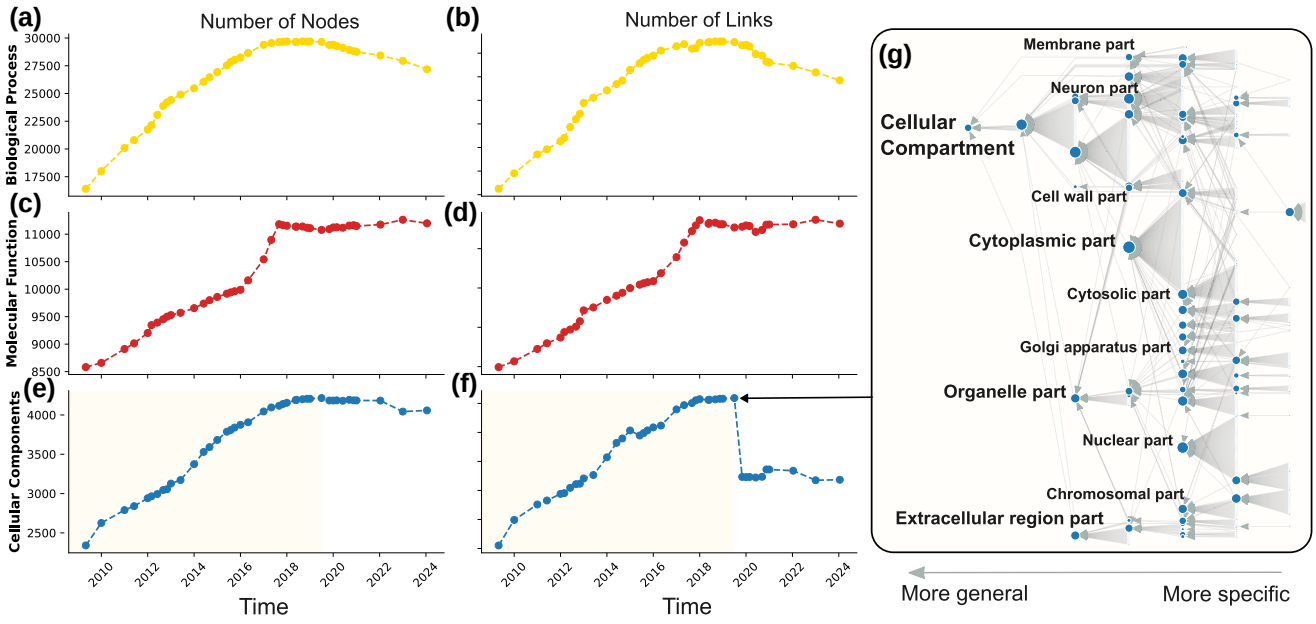

Fig. 3. Structural evolution of the GO branches. Panels a)–f) show the temporal evolution of the number of nodes $N(t)$ and links $L(t)$ in the three GO branches: biological process a)–b), molecular function c)–d), and cellular component e)–f). While biological process and cellular component networks exhibit a clear peak in size around 2018–2019, molecular function stabilizes after 2016. The cellular component ontology (CC) undergoes a particularly notable restructuring in 2019, in which e) 96% of nodes are retained, but f) approximately 50% of links are removed. Panel g) displays the hierarchical subgraph removed by the end of 2019, composed primarily of high-level "part"-style terms such as "cell part", "membrane part", and "organelle part". This subgraph structured CC inheritance until 2019 and its removal marked a major semantic transition of the ontology (see text).

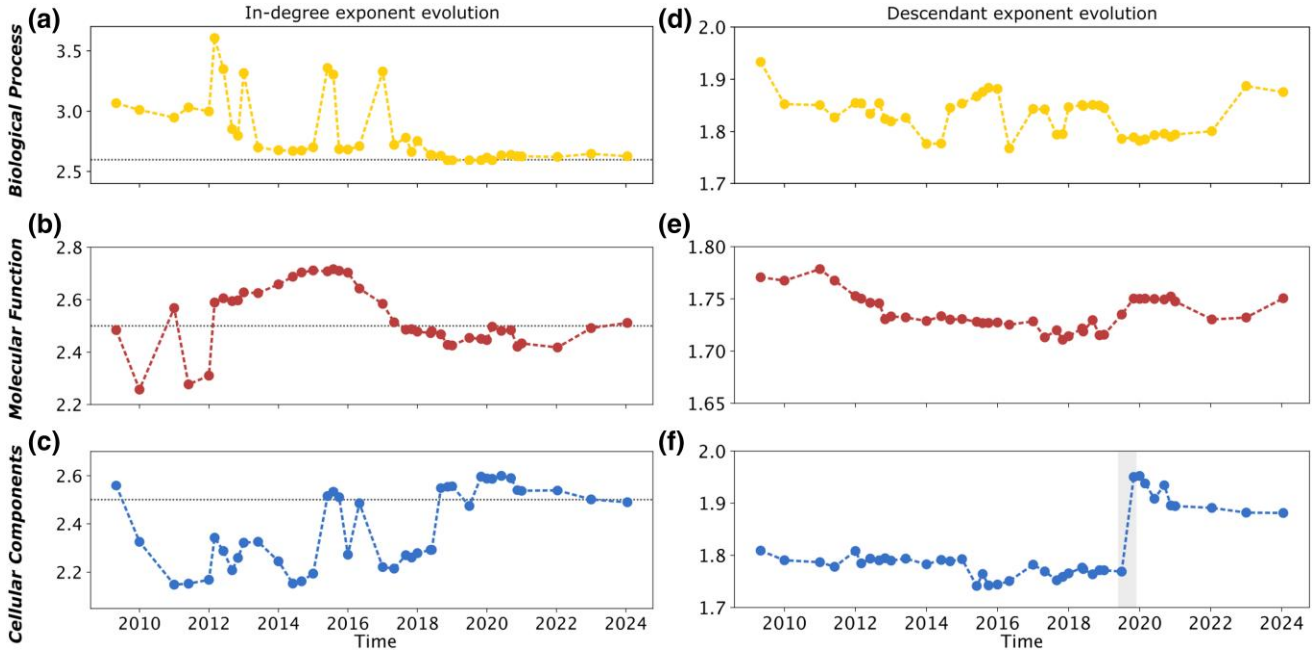

Fig. 4. Temporal changes in descendant and in-degree distributions for terms across Gene Ontology branches. The cumulative distribution of in-degrees $P_{>k} \sim k^{-\alpha(t)+1}$ evolves over time for a) biological processes, b) molecular activities, and c) cellular compartments. Despite different transient dynamics, all systems converge to a comparable exponent value $\alpha(2024) \approx 2.5$ (dashed lines). The cumulative distribution of descendants $P_{>D} \sim D^{-\gamma(t)+1}$ for d) biological processes and e) cellular compartments experienced an abrupt increase (shaded region) towards the exponent $\gamma \approx 1.9$ by the end of 2019, indicating a more random distribution than the f) molecular function network, which displays a steady lower exponent $\gamma \approx 1.75 < 2$.

removing intermediary "part" terms and anchoring components directly to high-level anatomical or complex structures, the ontology now channels semantic inheritance through a smaller set of hubs. Figure 5b shows the resulting rich-club core as of January 2020, organized around two top-level hubs: the newly introduced term "cellular anatomical entity" and "protein-containing complex". These two branches define the current modular architecture of the CC network (see also Fig. 2c). Importantly, this reorganization primarily affected the top layers of the ontology; regional hubs deeper in

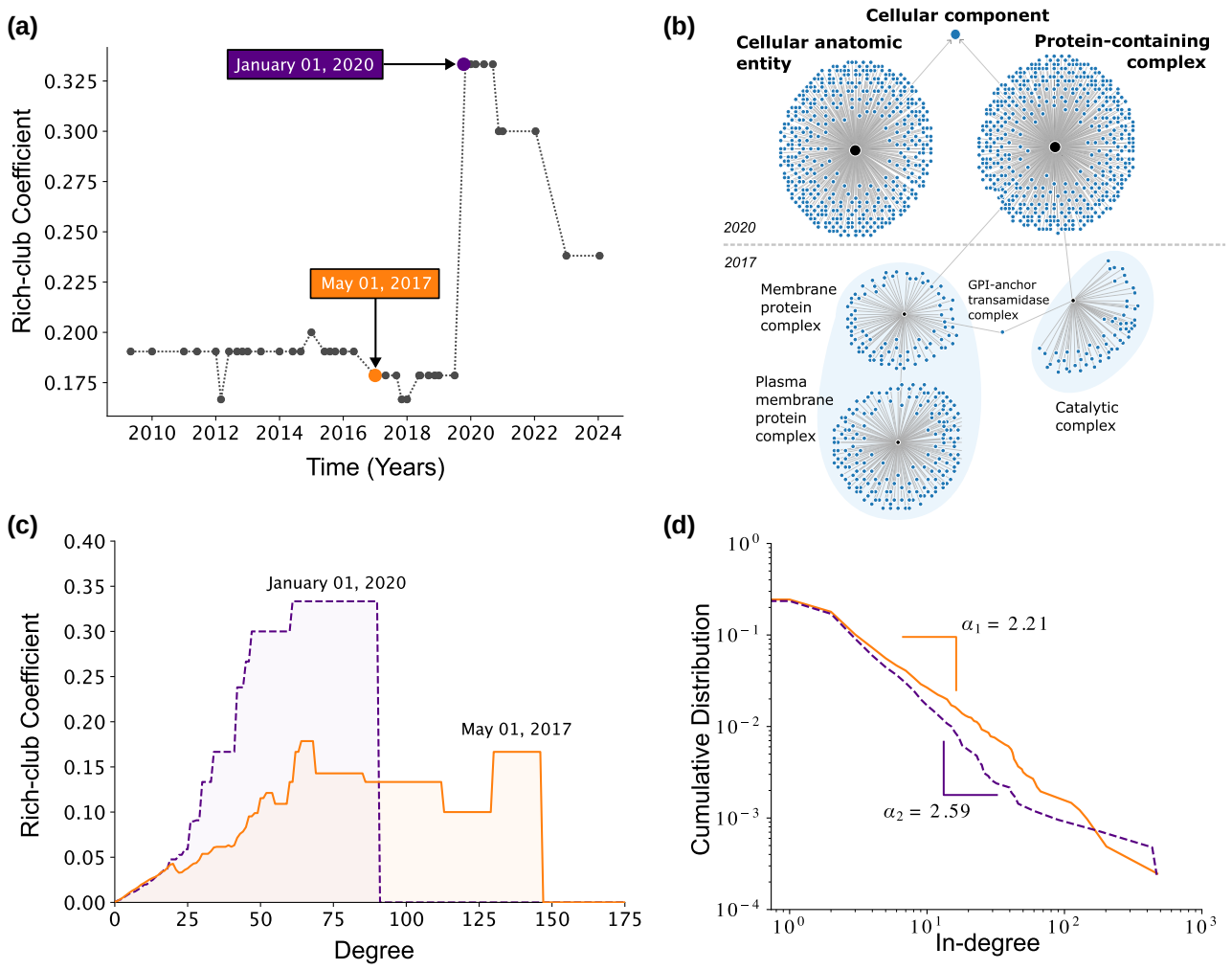

Fig. 5. Rich-club coefficients highlight non-local refactoring in the ontology network. a) The maximum rich-club coefficient of the cellular component (CC) network shows a marked increase at the end of 2019, indicating a sharp rise in topological centralization. Arrows highlight two CC network snapshots whose comparison reveals structural shifts in the semantic architecture. b) The rich-club subgraph of the CC network as of January 1, 2020, includes all nodes with degree $k > k_c \approx 50$ (black points). This subgraph defines the backbone of the reorganized CC hierarchy, rooted in the "Cellular component" term and structured around two major hubs: "cellular anatomical entity" (a new term introduced in 2019) and "protein-containing complex". In contrast, applying the same criteria to the May 1, 2017 version yields a smaller, fragmented subgraph (blue regions), consisting primarily of specific protein-containing complexes. c) The rich-club coefficient scales with degree up to $k_c \approx 50$, after which it saturates. Although neither network version reaches the theoretical maximum, the 2020 version shows higher overall values, reflecting increased centralization. d) The cumulative in-degree distribution $P(k) \sim k^{-\alpha+1}$ of the CC network displays a distinct hump in the January 2020 snapshot (dashed curve) that was absent in earlier versions (solid curve), consistent with an increased exponent ($\alpha_2 > \alpha_1$) and a more centralized semantic core.

the hierarchy were largely preserved (blue regions in Fig. 5b). The increased prominence of these top-level hubs is also reflected in the rise and saturation of the rich-club coefficient for nodes with degree $k > 50$, a pattern not observed in earlier versions of the network (Fig. 5c).

The logic behind the 2019 refactoring aligns with GO's broader goal of improving ontological consistency and supporting compositional reasoning in GO-CAM models (Thomas et al. 2019). To this end, GO was revised to better conform to upper-level frameworks such as the Common Anatomy Reference Ontology (CARO), which explicitly distinguishes anatomical structures from molecular assemblies (Haendel et al. 2008). Previously, spatial containment was partly encoded via inheritance links (e.g. CC:membrane part)—a semantically imprecise but functional workaround. After 2019, these general "part" terms were removed and replaced with explicit "part-of" relationships, decoupling spatial semantics from the taxonomic hierarchy. While this shift increased modular clarity, it also introduced a semantic bifurcation between anatomical entities and protein-containing complexes. As suggested below, annotations that once propagated through shared "part" terms may now appear disconnected unless multi-layer network representations are explicitly considered.

Structural Biases in Ontology Term Similarity

Accurately assessing gene function depends on measuring functional similarity between GO terms. These similarity metrics inform gene function prediction, annotation transfer across species, and computational modeling in functional genomics (Lord et al. 2003; Pesquita et al. 2009). However, because the three GO branches evolve as hierarchical and scale-free networks, similarity scores may be shaped not only by biological relationships but also by cultural and structural constraints. Recognizing these influences is essential for

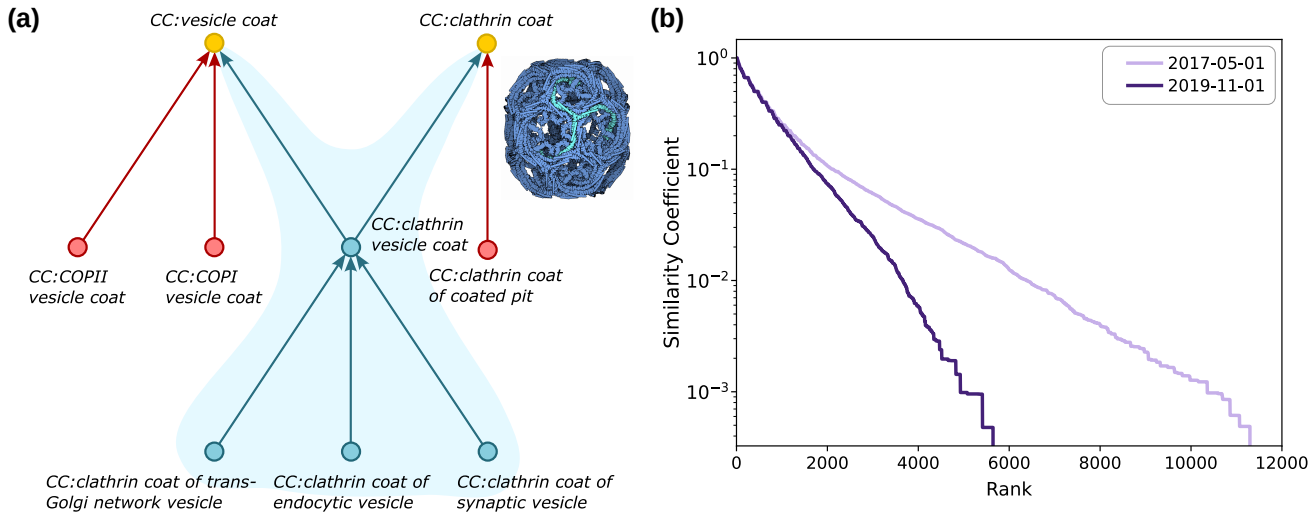

Fig. 6. Impact of refactoring on topological term similarity. a) Topological tree similarity between the terms "CC:vesicle coat" and "CC:clathrin coat" (yellow nodes at the top). This metric (equation (7)) is calculated using the size of shared subgraphs (blue nodes and edges) contrasted with their full subgraphs (including the red nodes and edges which are not shared between terms). In the CC network on May 1, 2017, the semantic similarity between these terms is 0.78. Clathrin coat image adapted from Goodsell (2007). b) The rank-frequency distribution of topological similarity between pairs of terms follows an exponential pattern in 2017, suggesting that similarity scores were widely distributed across the ontology. However, after the sweeping re-centralizing changes of the CC network on January 1, 2020, the distribution becomes significantly steeper, indicating that a few highly connected terms now dominate functional similarity relationships. This shift reflects a decrease in term ambiguity associated with network fragmentation and restructuring, where highly connected terms retained stronger internal coherence while cross-term relationships were pruned or redistributed. These changes imply that post-2020 similarity metrics may provide sharper functional distinctions but may also lose historical connections between functionally related terms.

ensuring that ontology-driven predictions remain both meaningful and unbiased.

Many widely used similarity metrics rely on the topological properties of GO. For example, the Wang method estimates similarity based on the hierarchical distance between terms and their shared ancestors (Schlicker et al. 2006). In this study, we employed a descendant-based similarity measure (see Methods), which captures overlap in inherited annotations. Figure 6 illustrates this computation in the Cellular Component (CC) network for the terms (CC:vesicle coat) and (CC:clathrin coat), which share the descendant (CC:clathrin vesicle coat)—a term corresponding to vesicles that mediate selective cargo sorting (Ashburner et al. 2000). This example shows how highly connected intermediates can act as semantic bridges, boosting similarity between functionally adjacent but mechanistically distinct terms.

Beyond individual comparisons, the overall distribution of similarity scores reveals how network structure influences inference. As shown in Fig. 6b, the rank-frequency distribution of similarity scores in 2017 follows an exponential curve, suggesting moderate ambiguity arising from overlapping hierarchies. However, after the CC branch was restructured in 2019–2020—marked by the removal of generic "part" terms and the emergence of rich-club behavior—the distribution became significantly steeper. Semantic inheritance became more concentrated in a few central nodes (see Fig. 5b), reducing ambiguity but also pruning cross-module relationships.

This transformation reflects more than structural cleanup; it marks a conceptual shift from treating GO as a semantic hierarchy to understanding it as a multi-layer network. In this framework, distinct biological dimensions—such as spatial, compositional, and regulatory context—are encoded as separate relationship types (Thomas et al. 2019). Although GO has long supported multi-type relationships, the 2019 reorganisation event institutionalized their structural role, signaling a transition to more explicit multi-layer representations.

While this shift increases semantic expressiveness (Hammoud and Kramer 2020), it also poses new challenges for similarity metrics. Multi-layer methods often require weighting schemes that are difficult to calibrate (Mazandu and Mulder 2013; Peng et al. 2015), and their added complexity may obscure interpretability. In hierarchical networks, moreover, highly connected hubs—such as (CC:cellular anatomical entity)—can dominate similarity scores regardless of their biological specificity. This effect is amplified by the scale-free nature of GO, where a few high-degree nodes exert disproportionate influence (Huntley et al. 2014). As a result, gene products involved in distinct processes may appear artificially similar due to shared ancestry through semantic hubs (Yon Rhee et al. 2008), reinforcing existing knowledge structures and masking biological complexity.

Discussion

The ability to accurately infer gene function depends on both experimental evidence and computational predictions (Radivojac et al. 2013). While experimental annotations are highly reliable, they are resource-intensive and not scalable. Computational methods, by contrast, can generalize across large datasets—but they depend critically on how biological knowledge is structured in the GO. This structure evolves continuously, as emphasized by the GO Consortium (Aleksander et al. 2023), GO is a dynamic, continuously evolving resource, revised in response to new discoveries, modeling frameworks, and user feedback.

Our findings support this evolutionary view by showing that GO evolves not only through incremental modifications but also through punctuated changes (Vidiella et al. 2022). The ontology evolves through a combination of bottom-up

reuse—anchoring new terms to widely used concepts—and top-down revisions, in which high-level terms are merged, reclassified, or removed to improve global coherence. This multiscale evolution produces scale-free topologies, with a small number of highly reused terms acting as semantic hubs. Such heterogeneous structures are common in cultural and technological systems (Valverde et al. 2002; Valverde and Solé 2007a), where established ideas are frequently reused (Duran-Nebreda et al. 2022) while innovations emerge at the periphery (Borycz et al. 2022).

Our comparison of multiple GO versions reveals not only globally stable hubs but also temporally localized "hotspots" of semantic reorganization, producing rich-club patterns indicative of centralized semantic control (Zhou and Mondragón 2004; Colizza et al. 2006; Valverde and Solé 2007b). We document a significant reorganization that occurred in 2019, when the Cellular Component branch underwent a top-down revision that eliminated broad "part" terms and introduced a clear modular separation between cellular anatomical entities and protein-containing complexes. This restructuring aligned the ontology with emerging principles of ontological clarity promoted by the Common Anatomy Reference Ontology (CARO) and supported compositional reasoning in GO-Causal Activity Modeling (GO-CAM) (Haendel et al. 2008; Thomas et al. 2019).

Semantic modularity, in this context, functions as a consolidation mechanism that resolves competing inheritance paths and clarifies the conceptual boundaries between major subdomains. Rather than flattening biological complexity, modularity preserves it by reorganizing semantic content into more coherent, functionally meaningful layers. This reduces ambiguity in topological inheritance while enabling compositional reasoning across distinct biological dimensions (Thomas et al. 2019). As a result, modularity stabilizes the ontology's evolving architecture and anchors further growth in a more expressive, multi-relational framework.

In this light, the 2019 restructuring can be viewed as the culmination of a longer cultural process of ontological negotiation. The GO reflects not only biological reality, but also the outcomes of evolving scientific priorities, modeling frameworks, and community consensus. Future work could extend this approach to analyze external drivers of semantic change—such as funding priorities, high-impact publications, or automated annotation tools—and their role in shaping GO's evolving architecture. Understanding domain-specific differences—such as those arising in developmental biology versus molecular function—may help explain the uneven impact of restructuring across GO branches and species.

More broadly, carefully documenting (and analyzing) the structural changes to the ontology has practical consequences for computational biology. While semantic modularity improves consistency, our analysis suggests that it can disrupt the performance of similarity metrics based solely on hierarchical proximity (Huntley et al. 2014). Functional similarity scores and annotation pipelines depend on the topology of the ontology. When key terms are restructured or rerouted through new semantic hubs, similarity scores between genes may shift, potentially affecting how genes are clustered, prioritized, or interpreted in downstream applications. In fields like cancer genomics or neurobiology—where GO plays a central role in biomarker discovery and gene enrichment analysis—such shifts can influence how disease-associated genes are ranked or grouped (Yon Rhee et al. 2008). Moreover, as annotations accumulate over time, even minor structural changes can lead to semantic drift, reducing the comparability of past and present results across genome versions (Thomas et al. 2012).

Addressing these challenges will require a new generation of ontology-aware algorithms capable of reasoning across layers and tracking semantic evolution. These models must integrate heterogeneous relationships while also accommodating modularity, structural rewiring, and term migration. Approaches that embed historical context, model semantic drift, and align ontology structure across time can improve the robustness of gene function prediction (Alterovitz et al. 2010). As machine learning and AI-driven annotation become increasingly central to bioinformatics, establishing community standards for semantic quality and structural stability will be crucial to ensure that evolving ontologies remain interpretable, interoperable, and scientifically grounded.

While GO provides guidelines for citing specific ontology releases (Consortium et al. 2023), version control alone is not sufficient. Tools and models must consider not only which ontology was used, but how its internal structure may have changed—and how even local edits can propagate to affect global meaning. This is particularly relevant in large-scale and planetary-scale studies of non-model organisms, where GO-based annotations often remain the only practical approach to assign function at scale. For instance, the LOGAN resource integrates over 27 million metagenome-assembled genomes across global environments, relying heavily on GO-driven pipelines due to the absence of curated annotations for most taxa (Chikhi et al. 2024). Preserving semantic continuity across GO versions is not merely a matter of technical rigor; it is foundational to reproducible science. As biological ontologies grow more expressive and modular, the challenge is not only to keep pace with their evolution, but to ensure that the systems built upon them—algorithms, annotations, predictions—remain coherent, interpretable, and reliable.

Methods

Gene Ontology Networks

To characterize the evolution of the GO, we analyzed 40 distinct ontology versions, sampled biannually between 2009 and 2024, spanning a total of 120 network reconstructions. The GO database undergoes continuous updates, reflecting new biological discoveries and refinements in functional annotations. However, the structural evolution of GO remains poorly understood. To address this, we reconstruct GONs for each version, enabling a quantitative analysis of their changing topology. GO consists of structured, hierarchical relationships that define biological processes, molecular functions, and cellular components (Ashburner et al. 2000; Consortium et al. 2023). Internally, each ontology version comprises three distinct GONs (see Fig. 2), each representing a fundamental aspect of cellular biology:

1. G_{bp}: Biological Processes (BP)—Represents interactions between biological functions, such as signal transduction or cell cycle regulation.
2. G_{mf}: Molecular Functions (MF)—Encodes functional activities at the molecular level, such as ATP binding or enzyme activity.

3. G_{cc}: Cellular Components (CC)—Captures spatial relationships between cellular structures, such as ribosomes or mitochondrial membranes.

Formally, each GON is represented as a directed graph $G = (V, E)$, where nodes (V) represent individual biological terms, and directed edges (E) define semantic relationships between terms. GO allows multiple types of semantic links, including "is-a" (hierarchical classification) and "part-of" (compositional relationships). However, in this study, we focus on the backbone of GO—its hierarchical "is-a" relationships—which define term inheritance and establish ontology structure. Links are stored in an adjacency matrix $A = [A_{ij}]$, where:

$$A_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E \text{ (i.e. } v_j \text{ is a child of } v_i) \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The GO database is available in multiple formats; we specifically use the "basic" edition, which retains only essential hierarchical relationships relevant to our analysis. GO datasets were downloaded from the GO website http://geneontology.org/docs/download-ontology. Earlier versions (pre-2009) were excluded due to format inconsistencies. We processed and analyzed GO networks using the following procedure: (i) imported ontology files into Python's *nxontology* library, (ii) extracted the "is-a" relationships to reconstruct hierarchical networks, (iii) converted networks to *Pajek* format using *networkx* for large-scale graph analysis, and (iv) handled obsolete terms by tracking replacements when available; otherwise, obsolete terms were excluded from the analysis.

As of March 2024, the largest GO network corresponds to biological processes, comprising $N_{bp} = 27,186$ terms, whereas the cellular components network is the smallest, with $N_{cc} = 4,058$ nodes. The molecular functions network is intermediate in size, containing $N_{mf} = 11,198$ terms. A key measure of network connectivity is the average degree (mean number of links per node), denoted as:

$$\langle k \rangle = \frac{L}{N} \quad (5)$$

where $N = |V|$ is the number of nodes (terms), and $L = \sum A_{ij}$ is the total number of links in the network. We find that the two smaller GONs (MF and CC) exhibit a tree-like topology, with an average degree close to $\langle k \rangle \approx 1$, consistent with a predominantly hierarchical structure. In contrast, the biological processes network (G_{bp}) is more densely interconnected, reflecting its higher complexity and functional integration. This difference suggests that GO exhibits heterogeneous scaling properties, requiring more advanced network metrics to fully characterize its evolution. All reconstructed networks can be freely downloaded from our website.

Power-law Fitting

To analyze the structural properties of GONs, we computed degree distributions, descendant distributions, and rich-club coefficients. Power-law fits were conducted using the *powerlaw* package in *python 3.11* (Clauset et al. 2009). For empirical degree distributions, we fit power-law models to the entire range of values using maximum likelihood estimation (MLE), which provides a robust approach for inferring scaling exponents (Clauset et al. 2009). To validate whether a power-law model is the best fit, we compared it against alternative distributions, including (i) exponential distributions, which model rapid decay in connectivity and (ii) log-normal distributions, which account for deviations from strict power laws. We assessed the goodness-of-fit using the Kolmogorov–Smirnov (KS) statistic and calculated P-values to test whether the observed distributions significantly deviated from a power-law model (Clauset et al. 2009). Additionally, for descendant distributions, we followed the same procedure, computing cumulative distributions $P_>(D)$ to reduce statistical noise and fit power-law models to assess the hierarchical structure of GONs.

Rich-club Coefficient

Network analysis reveals the interplay between bottom-up decentralized growth (as seen in the power-law degree distributions) and top-down restructuring (reflected in rich-club ordering). The rich-club coefficient quantifies whether high-degree nodes (hubs) exhibit preferential interconnectivity:

$$\Phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)} \quad (6)$$

where $E_{>k}$ represents the number of undirected edges between hub nodes (nodes with degree $k > k_c$), and $N_{>k}$ is the number of such nodes. A monotonic increase in $\Phi(k)$ suggests rich-club formation, where highly connected nodes become disproportionately linked (Zhou and Mondragón 2004; Colizza et al. 2006). In GO, this effect indicates large-scale refactoring, where a subset of terms centralizes connections due to curation efforts. Our findings reveal that the rich-club coefficient increases up to a threshold k_c, beyond which it saturates, indicating a transition in network organization. Detecting rich-club behavior in GONs provides evidence for non-local, top-down modifications, distinguishing routine term additions from global restructuring events (Valverde and Solé 2007b). This insight is critical for understanding how ontology revisions impact term connectivity and functional predictions over time.

Similarity Metrics

Semantic similarity measures quantify the functional relatedness of GO terms, influencing gene similarity scores, annotation transfer across species, and computational models in functional genomics (Pesquita et al. 2009). Given that GO is a hierarchical network, similarity metrics are often based on shared ancestry and network topology. We used a topology-based measure defined as the Jaccard similarity of descendant sets, which quantifies the fraction of shared descendant nodes between two GO terms:

$$s(i, j) = \frac{|\Gamma_i \cap \Gamma_j|}{|\Gamma_i \cup \Gamma_j|} \leq \frac{2|\Gamma_i \cap \Gamma_j|}{|\Gamma_i| + |\Gamma_j|} \quad (7)$$

where Γ_i denotes the set of descendant nodes of term i. This measure captures the overlap in influence between two terms within the ontology hierarchy, allowing us to assess how structural constraints shape functional similarity scores (Lord et al. 2003; Schlicker et al. 2006).

To assess how GO's structural evolution influences functional predictions, we examined the rank-frequency distribution of similarity scores. A rank-frequency plot displays the sorted frequency of observed similarity values, helping reveal whether similarity scores follow a specific statistical pattern (e.g. exponential or power-law decay). To compute the

rank-frequency distribution in Fig. 6b, we computed all pairwise similarity scores using equation (7), and sorted the unique similarity values in descending order (highest similarity ranked first). Finally, we plotted the frequency of each similarity value as a function of its rank in logarithmic scale. This method allows us to visualize whether similarity scores are broadly distributed or concentrated around a few dominant values.

Acknowledgments

S.V. thanks Rosa Fernandez's invitation to talk about the challenges of prediction gene function at the 2024 LifeHub conference, which inspired this study. The authors thank the participants of the conference for their feedback, particularly Juan Poyatos and Fernando Casares for their in-depth review of an earlier version of our study.

Author Contributions

Conceptualization: S.V. and R.F.; methodology: S.V., S.D-N., and B.V.; data collection and processing: S.V.; investigation: S.V., B.V., S.D-N., and G. I. M.-R; visualization: S.V., B.V., and S.D-N.; writing (original draft): S.V; writing (review and editing): S.V., R.F., S.D-N., A.M.R., A.B., R.A.B., and B.V.; supervision: S.V.

Funding

S.V. is supported by the Spanish Ministry of Science and Innovation through the State Research Agency (AEI), grant PID2020-117822GB-I00 /AEI/10.13039/501100011033. S.D-N. is supported by the Beatriu de Pinós postdoctoral program, from the Office of the General Secretary of Research and Universities and the Ministry of Research and Univertisites (2019 BP 00206) and the support of the Marie Sklodowska-Curie COFUND (BP3 contract no. 801370) of the H2020 program. A. M. R. is supported by the Spanish Ministry of Science and Innovation through the State Research Agency (AEI), grant PID2021-127503OB-I00 / /AEI/10.13039/501100011033. B.V. and S.V. are supported through the 2020-2021 Biodiversa and Water JPI joint call under the BiodivRestore ERA-NET Cofund (GA N°101003777) project MPA4Sustainability with funding organizations: Innovation Fund Denmark (IFD), Agence Nationale de la Recherche (ANR), Fundação para a Ciência e a Tecnologia (FCT), Swedish Environmental Protection Agency (SEPA), and grant PCI2022-132936 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. We acknowledge the support of the network PIE-202120E047-Conexiones-Life, organizer of the 2024 LifeHub conference.

Conflict of Interest

The authors declare that they have no competing interests.

Data Availability

All data and code used in this study are openly available at the following GitHub repository: https://github.com/svalver/go-evolution-networks. The repository includes all Gene Ontology network snapshots and analysis scripts necessary to reproduce the results presented in this manuscript.

References

Albert R. Scale-free networks in cell biology. *J Cell Sci*. 2005:118(21):4947–4957. https://doi.org/10.1242/jcs.02714.

Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, *et al.* The gene ontology knowledgebase in 2023. *Genetics*. 2023:224(1):iyad031. https://doi.org/10.1093/genetics/iyad031.

Alterovitz G, Xiang M, Hill DP, Lomax J, Liu J, Cherkassky M, Dreyfuss J, Mungall C, Harris MA, Dolan ME, *et al.* Ontology engineering. *Nat Biotechnol*. 2010:28(2):128–130. https://doi.org/10.1038/nbt0210-128.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* Gene ontology: tool for the unification of biology. *Nat Genet*. 2000:25(1):25–29. https://doi.org/10.1038/75556.

Barabási AL, Albert R. Emergence of scaling in random networks. *Science*. 1999:286(5439):509–512. https://doi.org/10.1126/science.286.5439.509.

Barrios-Núñez I, Martínez-Redondo GI, Medina-Burgos P, Cases I, Fernández R, Rojas AM. Decoding functional proteome information in model organisms using protein language models. *NAR Genom Bioinform*. 2024:6(3):lqae078. https://doi.org/10.1093/nargab/lqae078.

Bentley RA, Carrignon S, Ruck DJ, Valverde S, O'Brien MJ. Neutral models are a tool, not a syndrome. *Nat Hum Behav*. 2021:5(7):807–808. https://doi.org/10.1038/s41562-021-01149-x.

Borycz JD, Horne BD, Bentley RA. Forking paths of cultural complexity: a case study of research paradigms. *Adv Complex Syst*. 2022:25:2240004. https://doi.org/10.1142/S0219525922400045.

Burek P, Loebe F, Herre H. Towards refactoring the molecular function ontology with a UML profile for function modeling. *J Biomed Semantics*. 2017:8(1):1–11. https://doi.org/10.1186/s13326-017-0152-y.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggnog-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021:38(12):5825–5829. https://doi.org/10.1093/molbev/msab293.

Chikhi R, Raffestin B, Korobeynikov A, Edgar R, Babaian A. Logan: planetary-scale genome assembly surveys life's diversity. bioRxiv. https://doi.org/10.1101/2024.07.30.605881, 2024, preprint: not peer reviewed.

Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM Rev*. 2009:51(4):661–703. https://doi.org/10.1137/070710111.

Colizza V, Flammini A, Serrano MA, Vespignani A. Detecting rich-club ordering in complex networks. *Nat Phys*. 2006:2(2):110–115. https://doi.org/10.1038/nphys209.

Consortium T, Aleksander SA, Balhoff J, Carbon S, Cherry J, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL. The gene ontology knowledgebase in 2023. *Genetics*. 2023:224(1):iyad031. https://doi.org/10.1093/genetics/iyad031.

Duran-Nebreda S, O'Brien MJ, Bentley RA, Valverde S. Dilution of expertise in the rise and fall of collective innovation. *Humanit Soc Sci Commun*. 2022:9(1):1–10. https://doi.org/10.1057/s41599-022-01380-5.

Fowler M. *Refactoring: improving the design of existing code*. Boston: Addison-Wesley Professional; 2018.

Goodsell D. Molecule of the month: clathrin. RCSB Protein Data Bank, 2007.

Haendel MA, Neuhaus F, Osumi-Sutherland D, Mabee PM, Mejino Jr JL, Mungall CJ, Smith B. Caro–the common anatomy reference ontology. In: Burger A, Davidson D, Baldock R, editors. Anatomy ontologies for bioinformatics: principles and practice. New York City: Springer; 2008. p. 327–349. https://link.springer.com/book/10.1007/978-1-84628-885-2.

Hammoud Z, Kramer F. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Anal*. 2020:5(1):2. https://doi.org/10.1186/s41044-020-00046-0.

Huntley RP, Sawford T, Martin MJ, O'Donovan C. Understanding how and why the gene ontology and its annotations evolve: the go within uniprot. *Gigascience*. 2014:3(1):4. https://doi.org/10.1186/2047-217X-3-4.

Kim JD, Ohta T, Pyysalo S, Kano Y, Tsujii J. Overview of BioNLP'09 shared task on event extraction. In: Tsujii J, editor. Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task. Boulder, Colorado: Association for Computational Linguistics; 2009. p. 1–9.

Lord PW, Stevens RD, Brass A, Goble CA. Investigating semantic similarity measures across the gene ontology: the relationship between sequence and annotation. *Bioinformatics*. 2003:19(10):1275–1283. https://doi.org/10.1093/bioinformatics/btg153.

Martínez-Redondo GI, Barrios-Núñez I, Vázquez-Valls M, Rojas AM, Fernández R. Illuminating the functional landscape of the dark proteome across the animal tree of life through natural language processing models. bioRxiv. https://doi.org/10.1101/2024.02.28.582465, 2024, preprint: not peer reviewed.

Mazandu GK, Mulder NJ. Information content-based gene ontology semantic similarity approaches: toward a unified framework theory. *Biomed Res Int*. 2013:2013(1):292063. https://doi.org/10.1155/2013/292063.

Newberry MG, Ahern CA, Clark R, Plotkin JB. Detecting evolutionary forces in language change. *Nature*. 2017:551(7679):223–226. https://doi.org/10.1038/nature24455.

Peng J, Uygun S, Kim T, Wang Y, Rhee SY, Chen J. Measuring semantic similarities by combining gene ontology annotations and gene co-function networks. *BMC Bioinformatics*. 2015:16(1):1–14. https://doi.org/10.1186/s12859-015-0474-7.

Pesquita C, Faria D, Falcao AO, Lord P, Couto FM. Semantic similarity in biomedical ontologies. *PLoS Comput Biol*. 2009:5(7):e1000443. https://doi.org/10.1371/journal.pcbi.1000443.

Petersen AM, Arroyave FJ, Pammolli F. The disruption index suffers from citation inflation: re-analysis of temporal cd trend and relationship with team size reveal discrepancies. *J Informetr*. 2025:19(1):101605. https://doi.org/10.1016/j.joi.2024.101605.

Petersen AM, Tenenbaum JN, Havlin S, Stanley HE, Perc M. Languages cool as they expand: allometric scaling and the decreasing need for new words. *Sci Rep*. 2012:2(1):943. https://doi.org/10.1038/srep00943.

Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor K, Ben-Hur A, *et al*. A large-scale evaluation of computational protein function prediction. *Nat Methods*. 2013:10(3):221–227. https://doi.org/10.1038/nmeth.2340.

Sarafraz F, Eales J, Mohammadi R, Dickerson J, Robertson D, Nenadic G. Biomedical event detection using rules, conditional random fields and parse tree distances. In: Tsujii J, editor. Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task. Boulder, Colorado: Association for Computational Linguistics; 2009. p. 115–118.

Schlicker A, Domingues FS, Rahnenführer J, Lengauer T. A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinformatics*. 2006:7(1):1–16. https://doi.org/10.1186/1471-2105-7-302.

Solé R, Valverde S. Evolving complexity: how tinkering shapes cells, software and ecological networks. *Philos Trans R Soc Lond B Biol Sci*. 2020:375(1796):20190325. https://doi.org/10.1098/rstb.2019.0325.

Solé RV, Valverde S, Rodriguez-Caso C. Convergent evolutionary paths in biological and technological networks. *Evol Educ Outreach*. 2011:4:415–426. https://link.springer.com/article/10.1007/s12052-011-0346-1.

Sonawane AR, Weiss ST, Glass K, Sharma A. Network medicine in the age of biomedical big data. *Front Genet*. 2019:10:445334. https://doi.org/10.3389/fgene.2019.00294.

Steyvers M, Tenenbaum JB. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn Sci*. 2005:29(1):41–78. https://doi.org/10.1207/s15516709cog2901_3.

Thomas PD, Hill DP, Mi H, Osumi-Sutherland D, Van Auken K, Carbon S, Balhoff JP, Albou LP, Good B, Gaudet P, *et al*. Gene ontology causal activity modeling (go-cam) moves beyond go annotations to structured descriptions of biological functions and systems. *Nat Genet*. 2019:51(10):1429–1433. https://doi.org/10.1038/s41588-019-0500-1.

Thomas PD, Wood V, Mungall CJ, Lewis SE, Blake JA, Consortium GO. On the use of gene ontology annotations to assess functional similarity among orthologs and paralogs: a short report. *PLoS Comput Biol*. 2012:8(2):e1002386. https://doi.org/10.1371/journal.pcbi.1002386.

Valverde S, Ferrer-i Cancho R, Solé RV. Scale-free networks from optimal design. *Europhys Lett*. 2002:60(4):512–517. https://doi.org/10.1209/epl/i2002-00248-2.

Valverde S, Solé RV. Logarithmic growth dynamics in software networks. *Europhys Lett*. 2005a:72(5):858–864. https://doi.org/10.1209/epl/i2005-10314-9.

Valverde S, Solé RV. Network motifs in computational graphs: a case study in software architecture. *Phys Rev E*. 2005b:72(2):026107. https://doi.org/10.1103/PhysRevE.72.026107.

Valverde S, Solé RV. Hierarchical small-worlds in software architecture. *Dyn Contin Discrete Impuls Syst Series B Appl Algo*. 2007a:14:1–11. (Special Issue on Software Engineering and Complex Networks, eds K. L. Teo and X. Liu). https://doi.org/10.48550/arXiv.condmat/0307278.

Valverde S, Solé RV. Self-organization versus hierarchy in open-source social networks. *Phys Rev E*. 2007b:76(4):046118. https://doi.org/10.1103/PhysRevE.76.046118.

Vidiella B, Carrignon S, Bentley RA, O'Brien MJ, Valverde S. A cultural evolutionary theory that explains both gradual and punctuated change. *J R Soc Interface*. 2022:19(196):20220570. https://doi.org/10.1098/rsif.2022.0570.

Yon Rhee S, Wood V, Dolinski K, Draghici S. Use and misuse of the gene ontology annotations. *Nat Rev Genet*. 2008:9(7):509–515. https://doi.org/10.1038/nrg2363.

Zhou S, Mondragón RJ. The rich-club phenomenon in the internet topology. *IEEE Commun Lett*. 2004:8(3):180–182. https://doi.org/10.1109/LCOMM.2004.823426.