Agentic Analysis for Agentic Infrastructure: An LLM-Powered Pipeline for Comparative Governance of DAO and Corporate AI Protocols

Yutian Wang[†]
Duke Kunshan University
Kunshan, China

Luyao Zhang[*†]
Duke Kunshan University
Kunshan, China

Abstract

As AI agent protocols proliferate, the governance structures shaping their interoperability standards remain empirically underexamined. We introduce an LLM-powered comparative pipeline for large-scale governance discourse analysis, integrating automated annotation, neural topic modeling, and multi-layer network analysis to study socio-technical power structures at scale. We validate it on two contrasting standards for agent interoperability: ERC-8004 (a permissionless on-chain standard advanced by rough consensus) and Google A2A (a corporate-led protocol governed by a Technical Steering Committee). Analyzing the governance participation records, we combine LLM-assisted coding, topic modeling, and multi-layer network analysis to examine how institutional design shapes thematic priorities and community structure. We find that while governance form influences substantive focus, both regimes exhibit comparable levels of participation inequality and community fragmentation. Discourse alignment is denser in the permissionless setting, suggesting that open governance may foster greater thematic convergence despite decentralized participation. These findings illustrate how LLM-assisted methods can advance the empirical study of technology governance, with implications for designing more equitable agentic AI standards. All data and code are openly available.

Keywords

DAO governance, consortium governance, AI agents, LLM research pipelines, blockchain

1 Introduction

Interoperability is an inherited characteristic of decentralized applications [21]. As AI agent protocols proliferate, the governance of their interoperability standards remains empirically underexamined. We compare governance in a decentralized autonomous organization (DAO; a blockchain-native organization governed by smart contracts and community consensus rather than corporate hierarchy) [5, 22] with corporate-led governance. Who controls the rules by which autonomous agents discover, negotiate, and coordinate across organizational boundaries? This question sits at the intersection of artificial intelligence and institutional design [16, 53], yet the public discourse through which such standards are negotiated has been difficult to study at scale. Prior work relies predominantly on manual coding with fixed categories, limiting both thematic discovery and structural analysis across large text corpora.

We address this gap with an LLM-powered comparative pipeline for large-scale governance discourse analysis, integrating automated annotation, neural topic modeling, and multi-layer network analysis. We validate the pipeline on two standards: ERC-8004 and Google A2A. The two protocols address the same technical problem of AI agentic communication across systems, and their governance processes are both public on GitHub or Forums. Therefore, comparing them isolates the effect of governance form on who participates and what they discuss.

ERC-8004, which belongs to Ethereum Improvement Proposals (EIP), defines a smart-contract interface standard for on-chain AI agent identity. It "proposes to use blockchains to discover, choose, and interact with agents across organizational boundaries without pre-existing trust" [14]. Its lifecycle follows the stages defined in EIP-1, and is advanced by rough consensus—a decision rule requiring the absence of sustained substantive objection rather than a formal majority vote [13, 49]—among forum participants and EIP editors, with no formal votes and permissionless implementations. Google A2A introduces "An open protocol enabling communication and interoperability between opaque agentic applications" [1]. It was initiated by Google and donated to the Linux Foundation in June 2025, now governed by an eight-seat Technical Steering Committee (TSC) composed entirely of corporate representatives [2, 19].

Comparing the two cases, we ask:

RQ: Compared to corporate hierarchy, does the governance structure of a permissionless DAO really achieve a higher degree of decentralization?

We decompose this into three sub-questions:

- **RQ1 (Decision Architecture):** How do the formal decision procedures, entry rights, and authority structures of the two regimes differ?
- **RQ2 (Discourse Composition):** How does governance form shape the topical and argumentative composition of participation discourse?
- **RQ3 (Relational Networks):** How does governance form shape co-participation ties, discourse-level consensus and conflict structures, and actor–topic divisions of collaborative labor?

To answer these questions, we first reconstruct the two governance mechanisms from their official documents and visualize them as decision-flow diagrams. We then collect 4,323 governance participation records from their public repositories and apply an LLM-assisted annotation pipeline to label each record's argumentative function, stance, and stakeholder affiliation [8]. On top of this annotation, we run two topic-discovery methods: unsupervised BERTopic [20], and LLM-inductive Thematic-LM [43]. Moreover,

[*]The corresponding author: Luyao Zhang (email: lz183@duke.edu, address: Duke Kunshan University, No.8 Duke Ave. Kunshan, Jiangsu 215316, China).
[†]**Acknowledgments**: Yutian Wang gratefully acknowledges the Summer Research Scholar Program supervised by Professor Zhang at Duke Kunshan University.

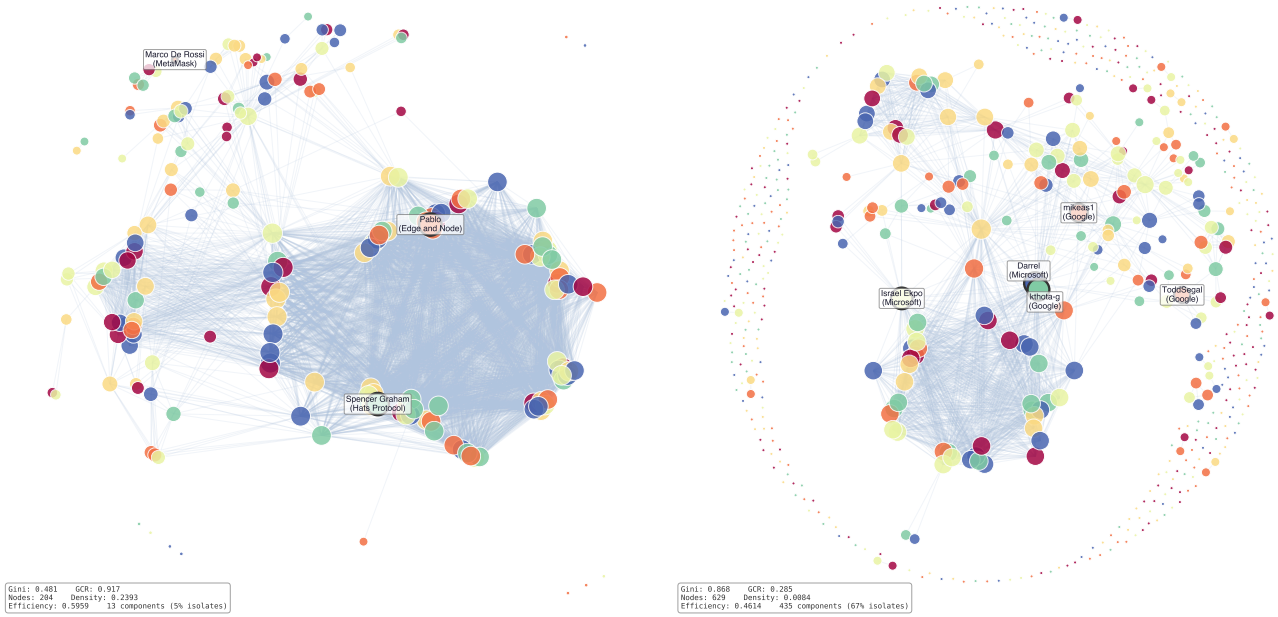

Figure 1: Co-participation network (see section 3.3) visualization of the 34 ERC AI protocol clustering and Google A2A. Left: ERC clustering; right: Google A2A.

we run three network analyses: co-participation social network analysis (SNA; mapping actors as nodes and shared thread participation as undirected edges) [4], discourse network analysis (DNA; projecting actor–theme stance matrices into congruence and conflict networks) [32], and socio-semantic bipartite networks (SSN; linking actors to the themes they discuss as a two-mode network, projecting to actor–actor co-engagement and theme–theme co-occurrence structures) [48]. The methodological triad mirrors the three sub-questions.

Three findings emerge. First, the two cases instantiate structurally opposite decision architectures: ERC-8004 advances by rough consensus with permissionless deployment, while A2A vests binding authority in a corporate technical steering committee. Second, governance form shapes the thematic focus of deliberation: DAO governance concentrates discourse on security mechanisms and protocol principles, whereas corporate governance distributes deliberation across engineering-execution workstreams. Third, discourse congruence is nonetheless denser in ERC-8004, reflecting tighter within-community consensus formation. Both networks exhibit comparable levels of participation inequality and community fragmentation only considering the single protocol of ERC-8004. When we cluster 34 ERC AI protocols, the ERC community become significantly more connected, see Appendix A and figure 1.

This paper makes three contributions. Methodologically, the architecture can generalize to any large-scale governance discourse analysis, offering a computational lens on the black box of public text negotiation, communication, and coordination. Empirically, it provides the first matched-case, multi-method comparison of DAO and corporate governance of agentic standardization with text corpora at scale. Theoretically, it pioneers the fundamental problem of who controls the future AI agentic infrastructure, with the interdisciplinary standing of institutional design and AI.

2 Related Work

Three streams of literature converge on this study. For a more specific comparison, see Appendix H.

Decentralized governance via blockchain. First, Harvey provides a comprehensive overview of the technologies, theories, and practices inherent to decentralized finance (DeFi) and blockchain technology [22]. Beck et al. [5] provided the foundational Information System framework mapping blockchain governance along decision rights, accountability and incentives dimensions. Ziolkowski et al. [55] identified six major governance challenges specific to the blockchain system, and half of them are unique to this technology. Kiayias and Lazos [28] systematised the field with a systematization of knowledge (SoK) review of blockchain governance mechanisms. More recently, Ellinger et al. [11] explored the balanced polycentric governance of digital commons. Reineke et al. [47] developed an integrative theoretical framework, showing the evolution of decentralization concepts. Sunyaev et al. [51] called for purposeful, design-oriented decentralization rather than ending at ideology. Motea and Oba [40] interrogated the democratic legitimacy of blockchain governance structures.

Governance of decentralized versus corporate structure. Murray et al. [41] examined how smart contracts and DAOs alter agency costs in corporate contracting, providing leading opinions. Lumineau et al. [35] highlight tacitness of transactions and social implications of this technology. Rahman et al. [45] and Hunt et al. [24] analyzed the dynamics and power accumulation of platforms. Hui and Tucker [23] address AI along with a decentralized

ecosystem, proposing an innovative governance framework. However, these studies rely on theoretical or interview-based methods; none of them use governance-participation data or computational methods to test structural difference between DAO and corporate governance in the same domain.

Computational studies of cooperative work. Researchers have long examined coordination in open, online communities. Mockus et al. [39] documented participation inequality in Apache and Mozilla. Im et al. [25] analyzed Wikipedia's Requests for Comments (RfC), a deliberation mechanism structurally analogous to EIP rough consensus, and revealed a persistent imbalance between deliberation and resolution. Germonprez et al. [18] catalogued structural realities of contemporary open-source projects, including pervasive corporate engagement; Li et al. [33] examined code-of-conduct conversations on GitHub as a window into informal governance norms in open-source repositories. Kulakowski and Frasincar [30] introduced CryptoBERT, a domain-adapted BERT variant pre-trained on 3.2 million cryptocurrency social-media posts, establishing a specialized embedding backbone for blockchain-native corpora. More recently, Wu et al. [44] applied sentiment and discourse analysis to six DAO forums; Stine and Agarwal [50] proposed comparative discourse analysis via topic models; Qiao et al. [43] introduced Thematic-LM for LLM-assisted inductive thematic analysis of large corpora; Ao et al. [4] used social network analysis on on-chain Aave data to show voting-power concentration; Leifeld [32] proposed discourse network analysis concentrating on participants' stance; Roth and Cointet [48] connects semantic analysis with social topology. Wang et al. [52] found the threat of power concentration via scaled empirical analysis of scale *SnapShot* data. Jungnickel et al. [27] quantified voting power concentration and participation apathy across DAO governance models, especially in token-based organizations. Chen et al. [9] applied a quasi-experimental propensity-score matching with difference-in-differences (PSM-DID) design to 98,000 Steemit users, finding that governance token ownership boosts curation effort but reduces creation novelty.

3 Methodology

We selected MiniMax-M2.5 [38] as the LLM backbone for its reasoning capability and low cost. The model is trained on complex real-world environments and achieves 80.2% on SWE-Bench Verified; frontier models of this capability tier have been shown to complete complex multi-turn analytical tasks with high reliability [36], making them suitable for assigning nuanced governance-process labels such as *Argument Type* and *Consensus Signal*. For design of the comparative case, see Appendix B.

3.1 Data Collection and LLM Annotation

We collect governance participation records from the public repositories. For ERC-8004, data originate from two sources: (1) 113 posts from the Ethereum Magician forum thread [14]; and (2) 36 GitHub records from the nine pull requests that directly modified `ERCS/erc-8004.md` [12]. For Google A2A, data originate from three streams in the `a2aproject/A2A` repository [1]: (1) 3,104 issue and issue-comment records, (2) 1,955 pull requests and review-comment records, and (3) 822 GitHub Discussion records.

The raw records totaled 6,030. We removed the following data: records with fewer than 20 characters of body text, which primarily are CI notifications, merge-conflict markers, and bot-generated status messages; and records attributed to verified bot accounts. After such filtering, 4,323 records are retained (ERC-8004: 142; Google A2A: 4,181). SHA-256 checksums for all raw data fields are provided in the authors' GitHub repository [1]. See also Appendix C.

The retained records were annotated with four categorical fields:

- **Stakeholder Institution**: Google / MetaMask / Ethereum Foundation / Coinbase / Independent / Unknown
- **Argument Type**: Technical / Governance-Principle / Economic / Process / Off-topic
- **Stance**: Support / Oppose / Modify / Neutral / Off-topic
- **Consensus Signal**: Adopted / Rejected / Pending / N/A

The institution affiliations of the top 109 contributors were also manually reviewed. The cascade is detailed in Appendix F. Robustness checks with expanded data and multi-model consensus annotation are reported in Appendix A.

3.2 Discourse Composition Analysis

To characterize the topical and argumentative composition of the two governance discourses, we apply three methods of increasing inductive depth.

3.2.1 Supervised argument typing. Treating the LLM-assigned *argument_type* as the record's communicative function, we test cross-case independence with a chi-square test [10] and within-ERC-8004 temporal change across three consecutive two-month phases spanning submission to mainnet deployment (phase boundaries in Appendix D).

3.2.2 BERTopic comparative discourse analysis. Following Stine and Agarwal [50], we fit BERTopic [20] jointly on the combined corpus. BERTopic uses class-based TF-IDF (c-TF-IDF; a term-weighting scheme that adjusts for topic-specific word distributions) [20] to represent topics by their most discriminative terms. Texts are embedded with `all-MiniLM-L6-v2` [46], reduced via UMAP (Uniform Manifold Approximation and Projection, a non-linear dimensionality reduction technique [37]; $n_{\text{neighbors}} = 15$, cosine, seed= 42) and clustered with HDBSCAN (Hierarchical Density-Based Spatial Clustering, a density-based clustering algorithm; min cluster size 10) into $K = 19$ topics plus a noise class. For each topic t and case c, the within-case share is

$$p_t^{(c)} = n_t^{(c)} / n^{(c)}.$$

Cross-case divergence is measured by Jensen–Shannon divergence (JSD; a symmetric, smoothed measure of distributional distance bounded between 0 and 1) [34]:

$$\mathrm{JSD}(p, q) = \tfrac{1}{2}\mathrm{KL}(p, m) + \tfrac{1}{2}\mathrm{KL}(q, m)$$

with

$$\mathrm{KL}(p, m) = \sum_i p(i) \log \frac{p(i)}{m(i)}, \qquad m = \tfrac{1}{2}(p + q),$$

and JSD = 0 denotes identical distributions, JSD = 1 disjoint.

To assess whether ERC-8004's topic concentration is an artefact of a general-purpose embedding model rather than a genuine

[1] https://github.com/kl41r3/erc8004-a2a-case-study

structural property, we also embed the 142 ERC-8004 records with CryptoBERT, a model further pre-trained on cryptocurrency social-media posts (StockTwits, Reddit, Telegram, Twitter) [30].

3.2.3 Thematic-LM inductive themes. To complement the embedding based view with human-interpretable labels, we apply the method of Thematic-LM [43], a four-stage multi-agent pipeline: (1) *open coding* assigns a short code to each record; (2) *aggregation* groups 300 sampled codes into 14 raw clusters; (3) *codebook review* merges and refines to 19 themes (T01–T19); (4) *theme assignment* labels every record with its best-fitting theme, or *Unclassified* when confidence is insufficient. We again report per-theme shares and JSD between cases.

3.3 Relational Network Analysis

We construct three complementary networks for each case, each layer adding discursive information to the previous. All graphs use actors as nodes; what changes is the semantic content of an edge.

3.3.1 Co-participation network (SNA). Following Ao et al. [4], an undirected edge connects two contributors who both posted to the same discussion thread (forum topic, GitHub issue, pull request, or discussion); edge multiplicity counts co-occurrences. We report six structural measures:

- Density: the realized fraction of possible ties. Computed by
$$\rho = \frac{2E}{N(N-1)}$$
where E denotes the number of edges and N denotes the number of nodes.
- Degree Gini: inequality of interaction counts (d_i is node degree). Computed by
$$G = \frac{\sum_{i,j} |d_i - d_j|}{2N \sum_i d_i}$$
where d_i is the number of edges of node i. $G = 0$ implies perfectly equal participation; $G \to 1$ implies that interactions are concentrated in a small elite.
- Components and giant-component ratio: fragmentation into disconnected threads. Computed by
$$\mathrm{GCR} = N_{\max}/N.$$
- Newman–Girvan modularity: reported for institution partitions and Louvain partitions [6] [42]. Computed by
$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$
where m is the total number of edges, A_{ij} is the adjacency matrix entry, k_i is the degree of node i, and $\delta(c_i, c_j) = 1$ if nodes i and j belong to the same community, 0 otherwise.
- Core–periphery: Borgatti–Everett coreness [7]. Computed by
$$\rho_{\mathrm{BE}} = \mathrm{corr}(A, \Delta), \qquad \Delta_{ij} = \delta_i \cdot \delta_j$$
where $\delta_i \in \{0, 1\}$ is the coreness label of node i. Statistical significance is evaluated by comparing the observed ρ_{BE} against a null distribution of random graphs that preserve the empirical degree sequence (configuration model), following the q-s test of Kojaku and Masuda [29].
- Betweenness centrality and network efficiency: identifies governance brokers—actors who mediate between otherwise disconnected groups:
$$b(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$
where σ_{st} is the total number of shortest s–t paths and $\sigma_{st}(v)$ the number that pass through v.
- Network efficiency: mean normalized harmonic centrality, measured by
$$\bar{h} = \frac{1}{n(n-1)} \sum_{u \neq v} d(u, v)^{-1}$$
where $d(u, v)^{-1} = 0$ for unreachable pairs [31].

3.3.2 Discourse Network Analysis (DNA). Following Leifeld [32], we construct the co-participation layer with stance-aware edges that capture shared discursive positions. Let A be the actor set and $T = \{T_{01}, \ldots, T_{19}\}$ the Thematic-LM codebook. For each record we encode stance into Support= +1, Modify= +0.5, Neutral= 0, Oppose= −1 and obtain an actor–theme stance matrix M similar to the pseudo-matrix below:

	Theme A	Theme B	Theme C
Alice	+1	−1	—
Bob	+1	+0.5	0
Carol	—	−1	+1

Two networks are projected from M: the *congruence network* G^+ connects actors who take same-sign stances on at least one shared theme (Alice and Bob agree on Theme A), and the *conflict network* G^- connects actors who take strictly opposite-sign stances (Alice and Bob conflict on Theme B); edge weights count the number of themes satisfying each criterion. We report density, Louvain modularity and betweenness centrality $b(v)$ of the top-5 discourse brokers.

3.3.3 Socio-semantic bipartite network. Following Roth and Cointet [48], we construct a two-mode (bipartite) network $\mathcal{B} = (A \cup T, E)$, where A is the set of actors, T the set of Thematic-LM themes, and an edge $(a, t) \in E$ exists whenever actor a authored at least one record assigned to theme t. The weight B_{at} equals the number of such records, forming a non-negative integer matrix $B \in \mathbb{Z}_{\geq 0}^{|A| \times |T|}$.

Two one-mode projections are derived from B. The *actor–actor projection* $W^A = BB^\top$ links any two actors by the number of themes they co-discussed (self-loops removed); the *theme–theme projection* $W^T = B^\top B$ links any two themes by the number of actors who engaged in both.

Per-actor topic diversity is measured by Shannon entropy:
$$H(a) = -\sum_{t=1}^{|T|} \hat{p}_{at} \log_2 \hat{p}_{at}, \qquad \hat{p}_{at} = \frac{B_{at}}{\sum_{t'} B_{at'}},$$
where $H(a) = 0$ for a pure specialist (all posts in one theme) and $H(a) = \log_2 |T|$ for a perfect generalist (uniform spread across all $|T|$ themes).

We further characterize the distribution of $\{H(a)\}_{a \in A}$ via its Gini coefficient. The higher the Gini, the more thematic breadth is concentrated in a few actors. We also measure per-theme actor concentration as the Gini of the column vector $\{B_{at}\}_{a \in A}$ for each t.

The thematic overlap coefficient is measured by

$$\Omega = \frac{|T_1 \cap T_2|}{\min(|T_1|, |T_2|)},$$

where T_c denotes the active theme set of case $c \in \{1, 2\}$, quantifies cross-case thematic alignment; $\Omega = 1$ means one community's thematic space is a subset of the other's.

Actor-set sizes across the three filtering stages are detailed in Appendix G.

4 Results

4.1 Decision Architectures

ERC-8004 and Google A2A represent contrasting governance archetypes, as detailed in figure 2. ERC-8004 remains at the stage of "Draft" when the data was fetched (Mar 2026); nevertheless, its three canonical registry contracts were deployed on Ethereum mainnet on January 29, 2026 [15]. Proposals are discussed through the Ethereum Magicians forum and are recorded on GitHub. Ownership of Google A2A transitioned from Google-controlled to Linux Foundation governance; contested decisions result in a GitVote [3]. For more details and pseudocode, see Appendix E.

4.2 Discourse Composition

4.2.1 Supervised argument typing. Figure 3 compares the LLM-assigned argument type distributions. Both communities are predominantly Technical (74.3% in ERC-8004 vs. 62.1% in A2A), but A2A devotes almost twice the share to Process arguments (25.4% vs. 13.9%). The cross-case difference is significant with a small effect size ($\chi^2(3) = 52.88$, $p < .001$, Cramér's $V = .103$): both cases remain technically grounded, but corporate governance carries substantively heavier coordination overhead.

Within ERC-8004, argument composition shifts significantly across three two-month phases ($\chi^2(6) = 25.32$, $p < .001$, $V = .315$). Technical arguments dominate Phases 1–2 (≥80%); in Phase 3 Process discussion surges to 53% as deliberation moves from substantive design to editorial ratification. This two-stage pattern is consistent with rough-consensus norms in open standards bodies. Small-n caution applies to Phases 2–3 ($n = 11$ and $n = 17$).

Figure 4 cross-tabulates stance with argument type. Support and Modify stances are dominated by Technical arguments in both cases (ERC Support 73%; A2A Support 87%): substantive engagement is predominantly technical regardless of governance form. Governance-Principle arguments, particularly under *Oppose*, are more visible in ERC-8004, reflecting the principled deliberation characteristic of the EIP process. The sharpest cross-case contrast appears in the *Neutral* row: 44% of A2A Neutral records are Process in nature, against 26% in ERC-8004—corporate governance generates procedural commentary even among non-committal participants.

4.2.2 BERTopic. BERTopic identifies 19 topics plus a noise class over the combined corpus. The global Jensen–Shannon divergence between the two cases' topic distributions is $\mathrm{JSD}_{\mathrm{BERTopic}} = 0.288$, a moderate but meaningful structural separation (Figure 5). ERC-8004 is strikingly concentrated: 67.6% of its records fall into Topic 0 (*agent, agents, for, of*), the broad agent-discourse cluster, versus 21.2% for A2A. Implementation-heavy topics—Task/Message management (8.2%), JSON/proto spec (7.7%), PR-contribution workflows (7.2%), SDK samples (5.9%)—carry substantial A2A weight but zero ERC-8004 records, confirming that low-level engineering is absent from the EIP forum. The only topic with a higher ERC share than A2A is Topic 9 (*suggestion, sap, linkedin*; 2.8% vs. 0.9%), signalling corporate voices (SAP, LinkedIn) entering the EIP thread.

CryptoBERT on ERC data shows more precise results. It divides the "agent" topic into T0 (Onchain, Reputation) and T2 (Trust, Feedback), and the total percentage is 73.2%, which aligns with the 67.6% of pure BERTopic (figure 6).

4.2.3 Thematic-LM. The Thematic-LM pipeline yields a 19-theme codebook (T01–T19). Figure 7 overlays per-case record share (diverging bars) with per-theme actor participation rate (lines with markers). Table 1 reports the six themes with the largest cross-case divergence. The global Jensen–Shannon divergence is JSD = 0.216.

The most discriminating theme is **T08 Trust & Security Mechanisms**. It covers agent-trust scoring, reputation schemes and on-chain credential verification—the foundational problem of establishing trustworthy agency. By contrast, A2A spreads deliberation across T06 Documentation & Examples, T18 Clarifications & Information Requests, T07 Community Collaboration & Contributions, and T01 Protocol Specification & Versioning. A2A additionally covers three themes: T09 (Transport & Protocol Mechanisms), T14 (Project Governance & Process), and T16 (Streaming & Real-time Communication). T09 and T16 are purely engineering-execution concerns entirely absent from the EIP forum; T14's ERC-8004 records were authored by automated review bots and thus fall outside the actor network.

The two methods agree on the direction and approximate magnitude of divergence, with BERTopic's JSD = 0.288 slightly exceeding Thematic-LM's JSD = 0.216. The gap is expected: Thematic-LM compresses divergence by absorbing related concerns into shared conceptual themes, while BERTopic retains finer-grained lexical distinctions (ERC's "trustless", "reputation" versus A2A's "samples", "proto"). Together the two estimates bracket the structural divergence and confirm its robustness: the two governance processes occupy meaningfully different but overlapping discourse spaces, with ERC-8004 concentrated on the constitutive layer of the design space ("what to build and why") and A2A on execution ones (how to build, document, and ship it).

Table 1: Thematic-LM themes with largest cross-case divergence. The result is aligned with figure 5.

ID	Theme	ERC %	A2A %	Δ
T08	Trust & Security	**34.5**	4.0	+30.5
T01	Protocol Specification	**13.4**	7.9	+5.5
T06	Documentation & Examples	2.1	**10.8**	−8.7
T05	SDK Development	0.7	**4.6**	−3.9
T15	Tooling & Automation	0.7	**4.5**	−3.8
T09	Transport Mechanisms	0.0	**3.2**	−3.2

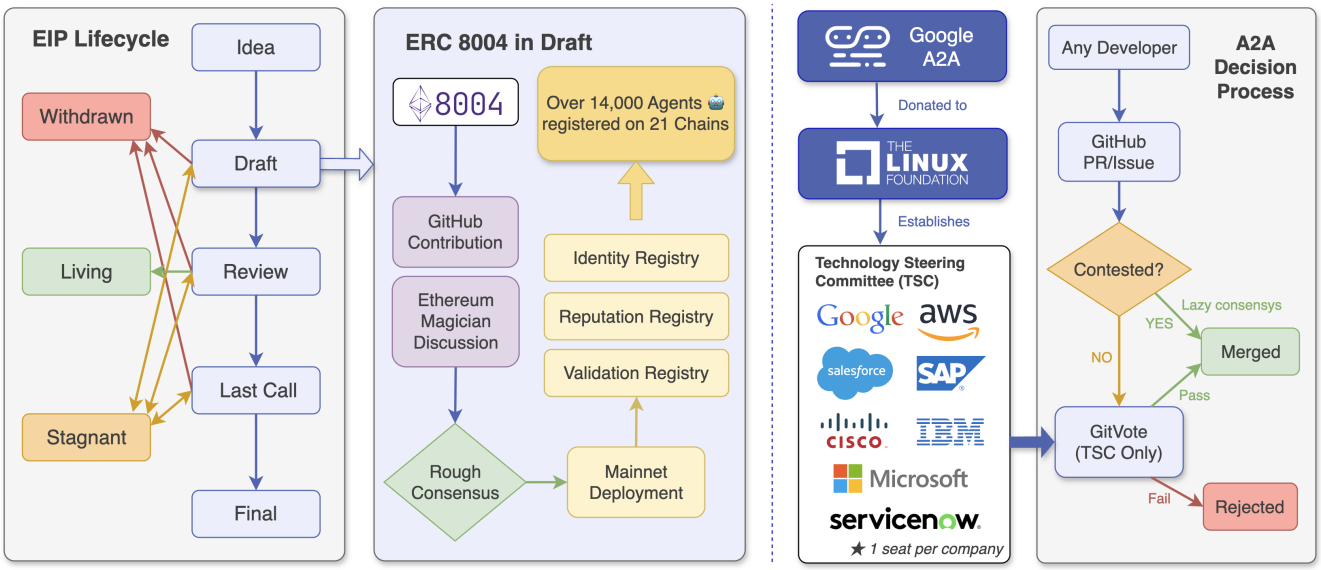

Figure 2: Left: ERC-8004 governance decision flow. Right: A2A governance structure and decision flow.

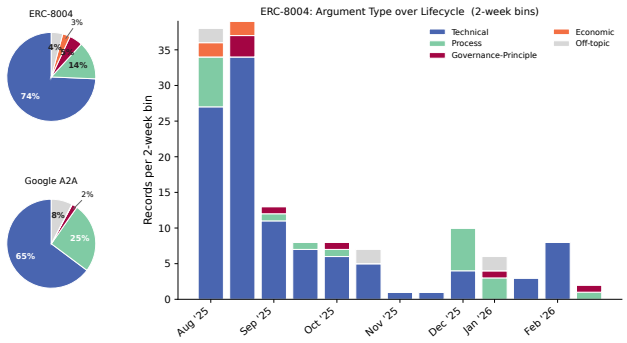

Figure 3: Left: argument-type distribution of both cases. Right: ERC-8004 argument-type distribution across three two-month phases.

4.3 Relational Networks

4.3.1 Co-participation network (SNA): core and periphery. Table 2 reports structural metrics for the co-participation networks (ERC-8004: $N = 67$; Google A2A: $N = 771$). For network visualization, see Figure 1.

Both communities exhibit a high degree of inequality. The Gini coefficient of degree is 0.804 for ERC-8004 and 0.779 for A2A, indicating that in both cases, a small elite drives the majority of interactions. The top-3 contributors account for 32.3% of all ERC-8004 interactions and 14.9% of A2A interactions. In A2A, two of the top three are Google employees, and the third is from Microsoft. In ERC-8004, the three contributors are: MarcoMetaMask (Marco De Rossi), whose username suggests affiliation with *MetaMask*; spengrah (Spencer Graham), cofounder of *Hats Protocol*; and pcarranzav (Pablo Carranza Vélez), who is from *The Graph*, a decentralized protocol for indexing and querying blockchain data. Their formal organizational mandate is indeterminate from public records, but their persistent participation accrues reputational standing.

Both networks are structurally fragmented, and neither shows statistically significant core-periphery structure. Louvain community detection recovers 46 communities for ERC-8004 and 358 for A2A, closely matching the component counts and confirming that participation organizes around parallel threads rather than a coherent deliberative body. Institution labels also do not predict interaction structure in either case. The giant component covers 32.8% of ERC-8004 nodes and 53.4% of A2A nodes; 333 of 771 A2A participants (43.2%) are complete isolates with no co-participation edges. Density indicates that the vast majority of possible co-participation pairs are unrealized. Under the Borgatti-Everett test [7], ERC-8004 returns $p = .139$ and A2A returns $p = 1.000$; neither crosses the $\alpha = .05$ threshold. For A2A, the BE algorithm classifies 769 of 771 nodes as core, a degenerate solution driven by the network's extreme fragmentation (density 0.004, 346 components), confirming the absence of a coherent core–periphery structure.

4.3.2 Discourse network: congruence and conflict. The stance layer (Table 3) reveals structures that are invisible in the co-participation graph.

ERC-8004 achieves denser within-community consensus, while A2A generates far greater absolute conflict volume. Congruence density is 0.148 for ERC-8004 versus 0.082 for A2A: within the tighter EIP community, participants more often share positions. Conflict edges in A2A are 34× larger in absolute count (2,531 vs. 74), which is not only an effect of actor numbers, but also from the broader range of technical positions supported by a multi-vendor engineering project.

The top-3 betweenness share in the congruence network drops to 34.5% for ERC-8004 and 12.2% for A2A, compared with 70.8%

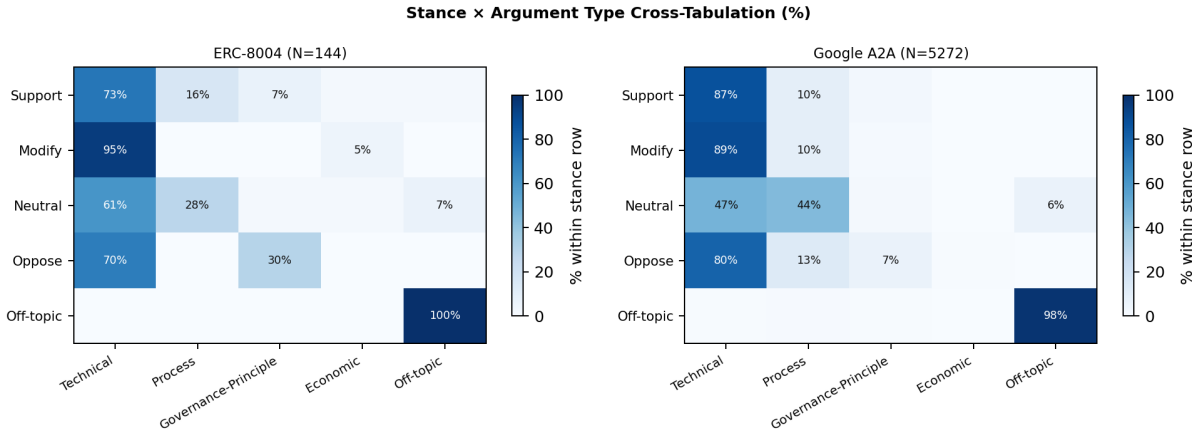

Figure 4: Stance × argument-type cross-tabulation (% within stance row).

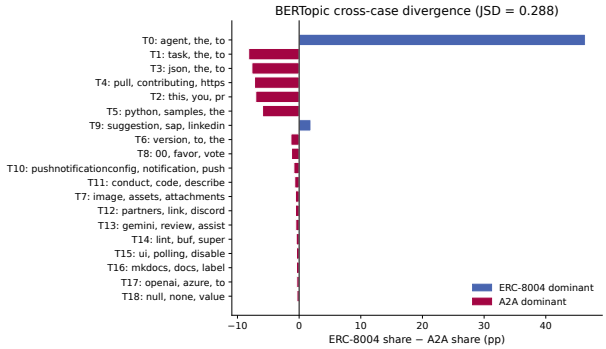

Figure 5: BERTopic cross-case divergence. ERC-8004 concentrates on foundational agent architecture, while A2A distributes deliberation across engineering-execution workstreams.

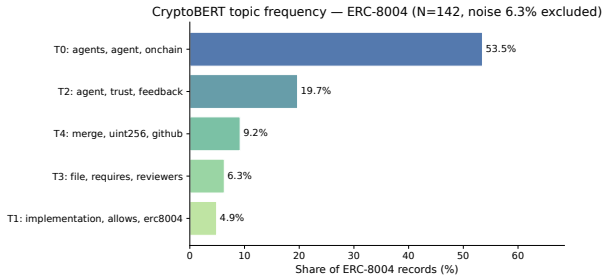

Figure 6: CryptoBERT topic frequency for ERC-8004.

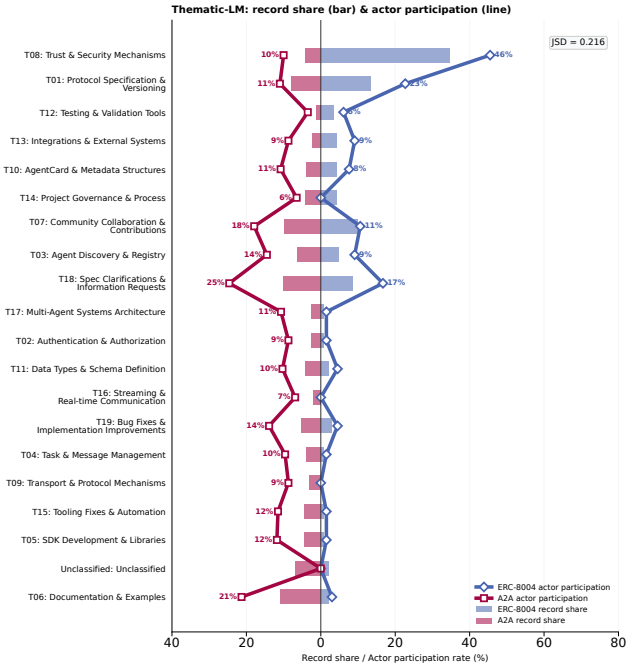

Figure 7: Thematic-LM overlaid chart. Bars show per-case record share; lines show actor participation rate. The DAO concentrates deliberation on security and trust; the corporate project distributes effort across documentation, tooling, and engineering-execution themes.

and 48.5% in the co-participation network (Table 2). Discourse brokerage is more distributed than structural interaction brokerage in both cases; yet ERC-8004's congruence network remains markedly more concentrated, consistent with a compact EIP core mediating contested positions on trust and protocol security.

4.3.3 Socio-semantic bipartite network: thematic concentration of discursive labor. The actor–theme bipartite layer (Table 4, Figure 8) reveals where each governance community directs its deliberative effort. Peripheral participation is the structural baseline in both cases: median actor Shannon entropy $H = 0$ in both ERC-8004 and A2A, meaning the majority of contributors engage a single theme

Table 2: Governance Network Structural Metrics. Despite structurally opposite decision architectures, both networks exhibit comparably high participation inequality: a small elite drives most interactions regardless of governance form.

Metric	ERC-8004	Google A2A
Nodes	67	771
Edges	65	1,230
Density	0.029	0.004
Gini (degree)	0.804	0.779
Top-3 degree share	32.3%	14.9%
Components	43	346
Giant component ratio	0.328	0.534
Modularity—institution	−0.059	−0.034
Modularity—Louvain	0.425	0.473
Louvain communities	46	358
CP p-value (BE test)	0.139	1.000
CP significant ($p < .05$)	No	No
Top-1 betweenness centrality	0.069	0.136
Betweenness Gini	0.931	0.979
Top-3 betweenness share	70.8%	48.5%
Network efficiency $\bar{h}$	0.050	0.110

CP = core-periphery; BE = Borgatti–Everett.

Table 3: Discourse Network Analysis metrics. The DAO achieves denser within-community agreement; the corporate regime generates a much larger volume of conflict edges, reflecting the broader technical surface of a multi-vendor engineering project.

Metric	ERC-8004	Google A2A
Actors	66	710
Active themes	17 / 19	19 / 19
Congruence edges	318	20,638
Congruence density	0.148	0.082
Congruence modularity (Louvain)	0.2886	0.2453
Top-1 betweenness centrality	0.061	0.018
Betweenness Gini	0.859	0.889
Top-3 betweenness share	34.5%	12.2%
Conflict edges	74	2,531

across their entire record [39]. This is not a governance-specific feature but an intrinsic property of large-scale deliberative systems.

What differs across governance forms is *which* themes absorb deliberative effort, and how broadly the most active participants range. Mean actor entropy is 0.348 for ERC-8004 and 0.617 for A2A: A2A's core contributors span roughly twice as many themes on average. Gini(H) remains high in both cases (0.773 vs. 0.707), confirming that thematic breadth is itself concentrated in a small number of generalist actors in both communities.

The cross-case divergence is sharpest at the theme level. In ERC-8004, 34.5% of actors participated in T08 (Trust & Security Mechanisms), and 13.1% participated in T01 (Protocol Specification & Versioning). For A2A, the pattern inverts for engineering-execution themes: Documentation & Examples (T06) engaged 10.8% of A2A actors but only 2.1% of ERC-8004 actors. Mean theme actor-concentration Gini is 0.453 for A2A against 0.085 for ERC-8004,

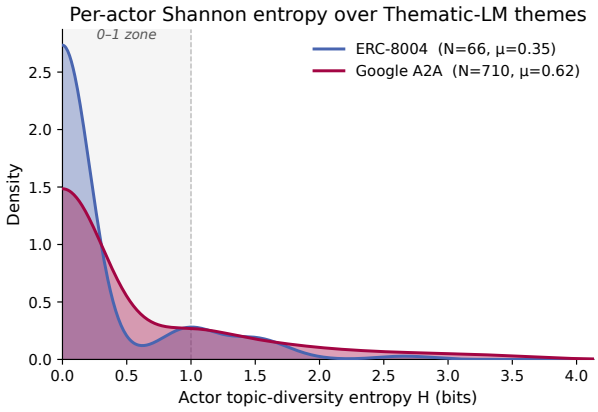

Figure 8: Per-actor Shannon entropy over Thematic-LM themes. $H = 0$ marks pure specialists; the $H = 0$ bin dominates both distributions.

Table 4: Socio-semantic bipartite network metrics. Both regimes share the same thematic space, but governance form drives which themes absorb deliberative effort and how tightly specialized those subgroups become.

Metric	ERC-8004	Google A2A
Actors	66	710
Active themes	17/19	19/19
Actor–actor projection edges	645	59,007
Mean actor entropy H	0.348	0.617
Median actor entropy H	0	0
Gini(H)	0.773	0.707
Max H	2.664	3.834
Mean theme Gini (actor concentration)	0.085	0.453
Most-discussed theme	T08	T06
Thematic overlap Ω	1.000	

indicating that each A2A theme attracts a narrower, more dedicated subgroup. The thematic overlap coefficient $\Omega = 1.0$: all 17 active ERC-8004 themes reappear in A2A.

5 Discussion

5.1 Decentralization More a Design than a Fact

Decentralized governance is appealing for its idealistic design of distributed authority and trustless coordination. In practice, however, the ERC-8004 process exhibits a centralization paradox.

Routine decisions are not subject to voting. The power to decide, for example, which proposals are "ready", which objections are "substantive", which positions count as "rough consensus", accumulates in the hands of a few minorities, mirroring the participation inequality patterns repeatedly documented in voluntary online cooperation [18, 39].

More fundamentally, no formal hierarchy is provided to distribute the power of decision. One possible reason for this is that the small scale of community size does not require such hierarchy; however, one direct drawback is that people tend to engage in groupthink, resulting in higher discourse congruence.

These two features compound: a certain minority holds the voice. The decentralization promised by the EIP architecture exists at the level of entry rights; in practice, routine decision authority concentrates around whoever has the time and reputation to remain in the room.

5.2 Role of Open Source and the Limitations

A natural objection is that the convergence stems not from governance form but from open-source software (OSS) norms shared by both cases. First, the topical divergence between the two cases (JSD = 0.288 on BERTopic, 0.216 on Thematic-LM) cannot be reduced to OSS norms; it is governance-driven, reflecting the constitutive–regulative distinction between proposing a standard (what the protocol *is*) and shipping an implementation (*how* it is built). Second, ERC-8004's discourse-congruence density is nearly twice that of A2A, the opposite of what an OSS-scale-only account would predict. The cautious framing that follows is therefore layered: open-source publication establishes a baseline of high participation inequality and fragmented community structure; governance form then redirects which themes this skewed participation engages and which actors accrue informal authority.

Here are two more possible limitations. First, ERC-8004 contributes 142 records against A2A's 4,181, so themes of low frequency carry wide confidence intervals. Second, A2A's TSC meetings, internal Google design reviews, and partner negotiations occur outside the public repository. The structural concentration we observe in A2A may therefore understate deliberation among its core members.

5.3 Who Controls Future Directions?

The answer is, in both cases, a small group of elites, despite the different constitutions. In ERC-8004, they are those who stay in the room; for A2A, corporate representatives are the gatekeepers. The deeper risk of the DAO model is not that elites emerge—they emerge in any sustained deliberation [18, 39]—but that, absent formal accountability structures, their influence operates through reputational authority that is opaque to outsiders and difficult for newcomers to contest.

Moreover, the governance form determines what problems communities treat as worth deliberating about. ERC-8004's deliberation is dominated by trust and security, while A2A also spreads its deliberation across engineering and execution issues. The concentration differences of DAO and corporate discourse reflect a value distribution that propagates into deployed systems: communities that deliberate about accountability will architect for accountability; those that deliberate about velocity will architect for velocity [53]. As agentic AI scales from research prototypes to critical infrastructure, these upstream deliberative choices become societal choices.

Three implications follow for those positioned to act on these findings. *Standards bodies* should pair open entry rights with procedural accountability mechanisms, because open access alone does not redistribute deliberative authority. *Protocol architects* should treat the deliberative agenda itself as a design artifact: topics absent from the docket become embedded defaults, so a periodic audit of what is not discussed is as essential as auditing what is. *Practitioners and policymakers* adopting either standard should consult its deliberation record alongside its specification, because the values that shaped the rules are rarely fully recoverable from the rules themselves.

References

[1] A2A Authors. 2025. Agent2Agent (A2A) Protocol. GitHub. https://github.com/a2aproject/A2A Accessed: Mar. 28, 2026.

[2] A2A Authors. 2025. GOVERNANCE.md. GitHub. https://github.com/a2aproject/A2A/blob/main/GOVERNANCE.md Accessed: Mar. 28, 2026.

[3] a2aproject. 2025. .gitvote.yml. GitHub. https://github.com/a2aproject/A2A/blob/main/.gitvote.yml

[4] Ziqiao Ao, Lin William Cong, Gergely. Horvath, and Luyao Zhang. 2023. Is Decentralized Finance Actually Decentralized? A Social Network Analysis of the Aave Protocol on the Ethereum Blockchain. arXiv:2206.08401 [econ.GN].

[5] Roman Beck, Christoph Mueller-Bloch, and John King. 2018. Governance in the Blockchain Economy: A Framework and Research Agenda. *Journal of the Association for Information Systems* 19 (10 2018), 1020–1034. doi:10.17705/1jais.00518

[6] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008. doi:10.1088/1742-5468/2008/10/P10008

[7] Stephen P Borgatti and Martin G Everett. 2000. Models of core/periphery structures. *Social Networks* 21, 4 (2000), 375–395. doi:10.1016/S0378-8733(99)00019-2

[8] Natalie A. Carlson and Vanessa Burbano. 2026. The use of LLMs to annotate data in management research: Foundational guidelines and warnings. *Strategic Management Journal* 47, 3 (2026), 699–725. arXiv:https://sms.onlinelibrary.wiley.com/doi/pdf/10.1002/smj.70023 doi:10.1002/smj.70023

[9] Kun Chen, Yifan Fan, Yulin Fang, and Xin (Robert) Luo. 2025. Beyond Money: Incentive Effects of Tokenized Ownership on User Contribution in DAOs. *Journal of Operations Management* 71, 7 (2025), 988–1016. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/joom.1351 doi:10.1002/joom.1351

[10] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Hillsdale, NJ.

[11] Eleunthia Wong Ellinger, Robert Wayne Gregory, Tobias Mini, Thomas Widjaja, and Ola Henfridsson. 2024. Skin in the game: The transformational potential of decentralized autonomous organizations. *MIS quarterly* 48, 1 (2024), 245–272.

[12] ERC-8004 Authors. 2026. erc-8004.md. GitHub. https://github.com/ethereum/ERCs/blob/c8cd50348b111d13f62e1d71d2b4c8101bd3514f/ERCS/erc-8004.md Accessed: Mar. 10, 2026.

[13] Ethereum Foundation. 2015. EIP-1: EIP Purpose and Guidelines. Ethereum Improvement Proposals. https://eips.ethereum.org/EIPS/eip-1 Accessed: Mar. 28, 2026.

[14] Ethereum Magicians. 2025. ERC-8004: Trustless Agents. Ethereum Magicians Forum, topic 25098. https://ethereum-magicians.org/t/erc-8004-trustless-agents/25098 Accessed: Mar. 10, 2026.

[15] Etherscan. 2026. ERC-8004 Identity Registry Contract (0x8004A169…). Ethereum Mainnet. https://etherscan.io/address/0x8004A169FB4a3325136EB29fA0ceB6D2e539a432 Deployed Jan. 29, 2026.

[16] James Evans, Benjamin Bratton, and Blaise Agüera y Arcas. 2026. Agentic AI and the next intelligence explosion. *Science* 391, 6791 (2026). doi:10.1126/science.aeg1895

[17] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (1971), 378–382.

[18] Matt Germonprez, Georg J.P. Link, Kevin Lumbard, and Sean Goggins. 2018. Eight Observations and 24 Research Questions About Open Source Projects: Illuminating New Realities. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 57 (Nov. 2018), 22 pages. doi:10.1145/3274326

[19] Google Cloud. 2025. Google Cloud Donates A2A to Linux Foundation. Google Developers Blog. https://developers.googleblog.com/en/google-cloud-donates-a2a-to-linux-foundation/ Accessed: Mar. 28, 2026.

[20] Maarten Grootendorst. 2022. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. arXiv:2203.05794.

[21] Campbell R. Harvey and Daniel Rabetti. 2024. International business and decentralized finance. *Journal of International Business Studies* 55, 4 (2024), 840–863. doi:10.1057/s41267-024-00705-7

[22] Campbell R. Harvey, Ashwin Ramachandran, and Joey Santoro. 2021. *DeFi and the Future of Finance.* Wiley, Hoboken, NJ.

[23] Xiang Hui and Catherine Tucker. 2025. Decentralization, Blockchain, Artificial Intelligence (AI): Challenges and Opportunities. *Journal of Product Innovation Management* 42, 5 (2025), 947–957. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jpim.12800 doi:10.1111/jpim.12800

[24] Richard A. Hunt, David M. Townsend, Robert Nugent, Joseph J. Simpson, Maximilian Stallkamp, and Esin Bozdag. 2024. Digital Battlegrounds: The Power Dynamics and Governance of Contemporary Platforms. *Academy of Management Annals* 19, 1 (2024). doi:10.5465/annals.2022.0188

[25] Jane Im, Amy X. Zhang, Christopher J. Schilling, and David Karger. 2018. Deliberation and Resolution on Wikipedia: A Case Study of Requests for Comments. 2, CSCW, Article 74 (Nov. 2018), 24 pages. doi:10.1145/3274343

[26] Jiarui Ji, Ran Lei, Xueqi Pan, Zhenhua Wei, Hao Sun, Yankai Lin, Xu Chen, Yujiu Yang, Yong Li, Bolin Ding, and Ji-Rong Wen. 2026. Leveraging LLM-based agents for social science research: insights from citation network simulations. *Humanities and Social Sciences Communications* 13 (2026), 127. doi:10.1057/s41599-025-06193-w

[27] Markus Jungnickel, Ferda Özdemir Sönmez, Cathy Mulligan, and William J. Knottenbelt. 2025. DAO Governance: Voting Power, Participation, and Controversy - A Review and an Empirical Analysis. *Distrib. Ledger Technol.* (Nov. 2025). doi:10.1145/3777416 Just Accepted.

[28] Aggelos Kiayias and Philip Lazos. 2023. SoK: Blockchain Governance. In *Proceedings of the 4th ACM Conference on Advances in Financial Technologies* (Cambridge, MA, USA) *(AFT '22)*. Association for Computing Machinery, New York, NY, USA, 61–73. doi:10.1145/3558535.3559794

[29] Sadamori Kojaku and Naoki Masuda. 2018. A generalised significance test for individual communities in networks. *Scientific Reports* 8 (2018). doi:10.1038/s41598-018-25560-z

[30] Mikolaj Kulakowski and Flavius Frasincar. 2023. Sentiment Classification of Cryptocurrency-Related Social Media Posts. 38, 4 (July 2023), 5–9. doi:10.1109/MIS.2023.3283170

[31] Vito Latora and Massimo Marchiori. 2001. Efficient Behavior of Small-World Networks. *Physical Review Letters* 87, 19 (2001), 198701. doi:10.1103/PhysRevLett.87.198701

[32] Philip Leifeld. 2013. Reconceptualizing Major Policy Change in the Advocacy Coalition Framework: A Discourse Network Analysis of German Pension Politics. *Policy Studies Journal* 41, 1 (2013), 169–198. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/psj.12007 doi:10.1111/psj.12007

[33] Renee Li, Pavitthra Pandurangan, Hana Frluckaj, and Laura Dabbish. 2021. Code of Conduct Conversations in Open Source Software Projects on Github. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 19 (April 2021), 31 pages. doi:10.1145/3449093

[34] J. Lin. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 37, 1 (1991), 145–151. doi:10.1109/18.61115

[35] Fabrice Lumineau, Wenqian Wang, and Oliver Schilke. 2021. Blockchain Governance—A New Way of Organizing Collaborations? *Organization Science* 32, 2 (March 2021), 500–521. doi:10.1287/orsc.2020.1379

[36] Chang Ma, Junlei Zhang, Zhihao Zhu, Cheng Yang, Yujiu Yang, Yaohui Jin, Zhenzhong Lan, Lingpeng Kong, and Junxian He. 2024. AgentBoard: an analytical evaluation board of multi-turn LLM agents. In *Proceedings of the 38th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS '24)*. Curran Associates Inc., Red Hook, NY, USA, Article 2365, 38 pages.

[37] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 3, 29 (2018), 861. doi:10.21105/joss.00861

[38] MiniMax. 2026. MiniMax M2.5: Built for Real-World Productivity. MiniMax. https://www.minimax.io/news/minimax-m25 Accessed: Apr. 11, 2026.

[39] Audris Mockus, Roy T. Fielding, and James D. Herbsleb. 2002. Two case studies of open source software development: Apache and Mozilla. *ACM Trans. Softw. Eng. Methodol.* 11, 3 (July 2002), 309–346. doi:10.1145/567793.567795

[40] Mosa Motea and Pius Oba. 2026. Who governs the ledger: rethinking blockchain governance through democratic innovation. *Humanities and Social Sciences Communications* 13, 1 (2026), 770. doi:10.1057/s41599-026-06980-z

[41] Alex Murray, Scott Kuban, Matthew Josefy, and Jon Anderson. 2021. Contracting in the Smart Era: The Implications of Blockchain and Decentralized Autonomous Organizations for Contracting and Corporate Governance. *Academy of Management Perspectives* 35, 4 (2021), 622–641. doi:10.5465/amp.2018.0066

[42] Mark E. J. Newman. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 23 (2006), 8577–8582. doi:10.1073/pnas.0601602103

[43] Tingrui Qiao, Caroline Walker, Chris Cunningham, and Yun Sing Koh. 2025. Thematic-LM: A LLM-based Multi-agent System for Large-scale Thematic Analysis. In *Proceedings of the ACM on Web Conference 2025* (Sydney NSW, Australia) *(WWW '25)*. Association for Computing Machinery, New York, NY, USA, 649–658. doi:10.1145/3696410.3714595

[44] Yutong Quan, Xintong Wu, Wanlin Deng, and Luyao Zhang. 2024. Decoding Social Sentiment in DAO: A Comparative Analysis of Blockchain Governance Communities. In *2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*. 216–224. doi:10.1109/QRS-C63300.2024.00037

[45] Hatim A. Rahman, Arvind Karunakaran, and Lindsey D. Cameron. 2024. Taming Platform Power: Taking Accountability into Account in the Management of Platforms. *Academy of Management Annals* 18, 1 (2024), 251–294. doi:10.5465/annals.2022.0090

[46] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 3973–3983. doi:10.18653/v1/D19-1410

[47] Philipp Reineke, Riitta Katila, and Kathleen M. Eisenhardt. 2025. Decentralization in Organizations: A Revolution or a Mirage? *Academy of Management Annals* 19, 1 (2025), 298–342. doi:10.5465/annals.2022.0206

[48] Camille Roth and Jean-Philippe Cointet. 2010. Social and semantic coevolution in knowledge networks. *Social Networks* 32, 1 (2010), 16–29. doi:10.1016/j.socnet.2009.04.005 Dynamics of Social Networks.

[49] Andrew L. Russell. 2006. 'Rough Consensus and Running Code' and the Internet-OSI Standards War. *IEEE Annals of the History of Computing* 28, 3 (July 2006), 48–61. doi:10.1109/MAHC.2006.42

[50] Zachary Kimo Stine and Nitin Agarwal. 2020. Comparative Discourse Analysis Using Topic Models: Contrasting Perspectives on China from Reddit. In *International Conference on Social Media and Society* (Toronto, ON, Canada) *(SMSociety'20)*. Association for Computing Machinery, New York, NY, USA, 73–84. doi:10.1145/3400806.3400816

[51] Ali Sunyaev, Michel Avital, and Mary C. Lacity. 2026. From Ideology to Design: Toward Purposeful Decentralization of Information Systems Governance. *Journal of the Association for Information Systems* 41, 1 (2026). doi:10.1177/02683962261430919

[52] Qin Wang, Guangsheng Yu, Yilin Sai, Caijun Sun, Lam Duc Nguyen, and Shiping Chen. 2025. Understanding DAOs: An Empirical Study on Governance Dynamics. *IEEE Transactions on Computational Social Systems* 12, 5 (2025), 2814–2832. doi:10.1109/TCSS.2025.3539889

[53] World Economic Forum and Accenture. 2026. Organizational Transformation in the Age of AI: How Organizations Maximize AI's Potential. World Economic Forum. https://www.weforum.org/publications/organizational-transformation-in-the-age-of-ai-how-organizations-maximize-ais-potential/

[54] Robert K. Yin. 2018. *Case Study Research and Applications: Design and Methods* (6th ed.). SAGE, Thousand Oaks, CA.

[55] Rafael Ziolkowski, Gianluca Miscione, and Gerhard Schwabe. 2020. Decision Problems in Blockchain Governance: Old Wine in New Bottles or Walking in Someone Else's Shoes? *Journal of Management Information Systems* 37, 2 (2020), 316–348. arXiv:https://doi.org/10.1080/07421222.2020.1759974 doi:10.1080/07421222.2020.1759974

A Data Expansion and Robustness Verification

This appendix tests the robustness of the main-text findings across the three sub-research questions. For RQ1 (Decision Architecture), we expand the DAO corpus from a single ERC to a 34-standard cluster. For RQ2 (Discourse Composition), we replicate the Thematic-LM analysis with four independent annotator models. For RQ3 (Relational Networks), we reconstruct co-participation and discourse networks at expanded scale and compare structural metrics against the main-text baseline.

We treat the 34 ERC standards as a single interconnected deliberative space rather than averaging separate networks, for three reasons. First, cross-standard interconnection is the phenomenon. The expanded ERC network coalesces into a dominant component (GCR 0.917, see table 10) precisely because the same contributors participate across multiple standards, creating cross-referencing ties. Averaging 34 separate networks would erase these inter-standard edges and miss the ecosystem-level finding that DAO governance externalizes deliberation into an interconnected public record. Second, matched-case comparability: the main text compares one standard against one corporate protocol, and the robustness check asks whether the same patterns hold when the DAO side is expanded to its full contemporaneous ecosystem. Averaging 34 small, sparse networks would not be comparable to the single dense A2A network. Third, the observability differential depends on clustering: the key divergence—ERC becomes more connected at scale while A2A remains fragmented—only emerges when the 34 standards are treated as one interconnected deliberative space. If averaged, one

would likely see 34 fragmented networks and falsely conclude that DAO governance is equally fragmented at scale.

To verify that main-text findings survive annotator choice and case scope, we expanded the DAO corpus from a single ERC-8004 thread to a 34-ERC agent-standardization cluster, re-annotated both cases with multiple independent models, and compared results against the original MiniMax-M2.5 labels.

A.1 Design

Data Expansion. Data were scraped on 2026-06-14. Beyond ERC-8004, we added 33 contemporaneous (≥2025-08) agent-standardization ERCs from the Ethereum Magicians forum, totaling 34 unique ERC standards (Table 5). The cluster spans identity and reputation, verification and proof, discovery and registry, commerce and payments, governance and security, and tooling and extensions. The expanded ERC corpus contains 1,664 annotated records, up from 142 in the main text. The A2A corpus was reused.

Table 5: ERC Standards in the Expanded Cluster ($N = 34$). Forum post counts from the Ethereum Magicians forum, scraped 2026-06-14. Total annotated records (forum + GitHub) = 1,664.

ERC	Title / Description	Posts	Thread
8001	Secure Intents: Cryptographic Framework for Agent Coordination	2	24989
8004	Trustless Agents (anchor case)	227	25098
8033	Agent Council Oracles	6	25638
8041	Fixed-Supply Agent NFT Collections	3	25656
8107	ENS Trust Registry for Agent Coordination	1	27200
8118	Agent Authorization	12	27402
8122	Minimal Agent Registry	1	27405
8126	AI Agent Verification	39	27445
8150	Zero-Knowledge Agent Payment Verification	8	27665
8160	Primary Agent Registry: Address-to-Agent Resolution	1	27727
8162	Agent Subscription Protocol	3	27751
8165	Agentic On-Chain Operation Interface	1	27773
8166	Shared Sequencer Interface for Agent L2s	1	27772
8171	Token Bound Account (Agent Registry)	4	27967
8181	Self-Sovereign Agent NFTs	3	27512
8183	Agentic Commerce	307	27902
8184	Payment Channels with Signed Vouchers	4	28012
8196	AI Agent Authenticated Wallet	10	27987
8203	Agent Off-Chain Conditional Settlement Interface	20	28041
8210	Agent Assurance	44	28097
8217	Agent NFT Identity Bindings	25	28339
8220	Onchain AI Governance	7	28162
8226	Regulated Agent Mandate	8	28208
8239	Agent Skill Registry	18	28335
8240	Trust Infrastructure for Agents and Assets	40	28322
8242	H3 Spatial Identity Extension for On-Chain Agents	1	28394
8257	Agent Tool Registry	15	28457
8259	AI Agent Identity, Reputation & Threat Registry	14	28521
8263	Onchain Proof Layer for AI Agents	20	28577
8264	AI Agent Memory Access Rights	1	28584
8273	Attestation-Gated Agentic Actions	12	28617
8274	AI Inference Proof Verification	115	28083
8275	Agent Service Discovery and Escrow Payments	13	28622
8294	Validation Network Interface for ERC-8004	14	28669

Thread URLs: `https://ethereum-magicians.org/t/{thread_id}`. ERC-8004 forum posts aggregated across 8 threads.

Cross-Model Annotation. Three models (DeepSeek-V4-Flash, GLM-4-Plus, and Moonshot-v1-auto) independently annotated every record on all five fields (stakeholder_institution, argument_type, stance, consensus_signal, key_point). Majority voting (2 out of 3) yielded consensus labels for 1,664 ERC and 4,187 A2A records. Together with the original MiniMax-M2.5, four independent annotators provide inter-annotator agreement estimates.

Cross-Round Annotation. Each of the three models independently re-annotated every record three times on the three core fields (argument_type, stance, consensus_signal). Majority vote across rounds within each model produced a per-model consensus; majority vote across models then produced the final cross-consensus. Per-field cross-consensus agreement rates are reported in Table 6. When all three models disagree for a given field (no majority; < 2%–3.6% of records), the label is assigned by plurality with an effectively arbitrary tie-break and is treated as low-confidence in subsequent analyses.

Table 6: Three-model cross-consensus agreement rates (majority voting across 3 rounds within each model, then 2-of-3 across models). "No maj." = all three models disagree on that field.

	ERC (N=1,664)			A2A (N=4,187)		
Field	3/3	2/3	No maj.	3/3	2/3	No maj.
Argument type	68.0%	30.1%	1.9%	66.4%	31.6%	1.9%
Stance	58.5%	39.4%	2.2%	55.5%	40.9%	3.6%
Consensus signal	59.6%	37.9%	2.5%	61.1%	37.3%	1.6%

Table 7: Scale Bridge: Main Text (Single-Standard) vs. Appendix (Expanded / Re-annotated)

Case	ERC-8004	34 protocols cluster
ERC records	142	1,664
ERC actors (SNA)	67	194
ERC actors (DNA)	66	194

A.2 Robustness Verification

Cross-model reliability. Four-model Fleiss' κ (a chance-corrected inter-rater agreement statistic ranging from −1 to 1) was computed (Table 8). All three fields reach Moderate agreement in both cases, with argument type—the core field driving the χ^2, BERTopic, and Thematic-LM analyses—at $\kappa = 0.545$ (ERC) and $\kappa = 0.541$ (A2A). The strongest pairwise agreement occurs between GLM-4-Plus and Moonshot-v1-auto (argument type $\kappa = 0.671$ ERC, 0.566 A2A), suggesting these two models share similar interpretive biases relative to DeepSeek-V4-Flash and MiniMax-M2.5.

Table 8: Four-model Fleiss' κ (MiniMax-M2.5, DeepSeek-V4-Flash, GLM-4-Plus, Moonshot-v1-auto). ERC $N = 144$, A2A $N = 4{,}185$.

Field	ERC	A2A
Argument type	0.545	0.541
Stance	0.579	0.525
Consensus signal	0.485	0.495

Cross-round self-consistency. Within-model test-retest reliability (Table 9) reveals clear model-level differences. GLM-4-Plus and Moonshot-v1-auto achieve almost perfect self-consistency (Fleiss' $\kappa > 0.81$ on all three fields in both cases), while DeepSeek-V4-Flash yields Moderate-to-Substantial values ($\kappa = 0.50$–0.63). The gap confirms that annotator model choice dominates stochastic variation: two models are near-deterministic in their labeling

behavior, while one exhibits higher intrinsic variance [26]. Notably, Moonshot-v1-auto, a generic auto-routing model with no special reasoning configuration, outperforms both alternatives in self-consistency, challenging the intuition that reasoning models are inherently more reliable annotators.

Table 9: Within-model cross-round Fleiss' κ (3 rounds, majority vote). ERC N=1,664, A2A N=4,187.

	ERC			A2A		
Model	**AT**	**St**	**CS**	**AT**	**St**	**CS**
Moonshot-v1-auto	0.976	0.956	0.942	0.966	0.952	0.929
GLM-4-Plus	0.925	0.904	0.861	0.917	0.880	0.835
DeepSeek-V4-Flash	0.634	0.554	0.507	0.587	0.556	0.504

AT = argument type; St = stance; CS = consensus signal.

A.3 Substantive Replication

We re-ran the annotation pipeline with three additional models (Moonshot-v1-auto, GLM-4-Plus, DeepSeek-V4-Flash; ICR $\kappa \approx 0.7$–0.9 for the top two, § A.1) and replicated the Thematic-LM discourse analysis [43] using Moonshot-v1-auto as the LLM backbone, yielding a 12-theme codebook with 96.6% coverage. Table 10 reports the key combined metrics.

Table 10: Replication metrics: multi-model consensus annotation + Moonshot Thematic-LM codebook.

Metric	ERC (cluster)	A2A (re-anno.)
DNA actors	194	713
DNA congruence density	0.403	0.252
Giant component ratio	0.917	0.285
Mean actor entropy H	0.734	0.511
Dominant theme (% record)	Compliance (31.1%)	Documentation (31.4%)
Thematic JSD	0.092	

A.3.1 Replication vs. Divergence. Three main-text findings replicate across the expanded data: (1) Technical argument types dominate regardless of annotator, with A2A receiving roughly twice the Process share; (2) both networks remain steeply unequal (betweenness Gini ≈0.8); (3) ERC discourse congruence density exceeds A2A (0.403 vs. 0.252). Thematic content also replicates: ERC concentrates on constitutive themes (compliance, standards, verification), A2A on executive themes (documentation, coordination, community). However, **network connectivity reverses:** the expanded ERC network coalesces into one dominant component (GCR 0.328→0.917), while the re-annotated A2A network fragments further (GCR 0.534→0.285). The cross-case thematic JSD of 0.092 is lower than the main-text BERTopic value (0.288), partly reflecting Moonshot's coarser 12-theme granularity. We interpret the connectivity reversal as an *observability* differential: permissionless DAO governance externalizes deliberation into a genuinely interconnected public record, whereas corporate governance internalizes decisive coordination channels (TSC calls, private forums), leaving the public GitHub trace persistently fragmented. Participation *concentration* is comparable across both governance forms, but *observability* is not.

A.4 Implications for Expanded Scale

At expanded scale, the ERC cluster's higher congruence density and coalesced giant component suggest that open governance may foster greater thematic convergence as the number of standards grows. This strengthens the main-text implication for standards bodies: open entry rights should be paired with procedural accountability mechanisms, because open access alone does not redistribute deliberative authority.

B Comparative Case Study Design

We adopt a comparative case study design to examine how governance structure shapes participation dynamics in AI protocol standardization [54]. Interoperability is an inherited characteristic of decentralized applications (dApps) in DeFi [21], and ERC-8004 as a protocol specifically designed for AI agents extends this feature. Google A2A, on the corporation's side, realized similar functions in different contexts. Moreover, they originated within the same year (2025). Nevertheless, they differ sharply in governance architecture. ERC-8004 belongs to Ethereum, a permissionless and community-driven open-source DAO, and Google A2A is now governed by TSC. Therefore, the two protocols constitute a theoretically matched pair. Holding the technical domain constant while varying the governance form enables structured comparison of participation patterns, discourse composition, and network analysis.

C Data Provenance

All raw data files are versioned with SHA-256 checksums stored in `data/raw/CHECKSUMS.json` in the project repository. Table 11 lists the file names, record counts, and collection dates for the primary raw data files used in this study.

Table 11: Raw Data Files and Record Counts

File	Records	Collected
forum_posts.json	113	2026-03
github_comments_filtered.json	36	2026-03
a2a_issues.json	3,104	2026-03
a2a_prs.json	1,955	2026-03
a2a_discussions.json	822	2026-03
Total (raw)	6,030	
Total (retained)	4,323	

D ERC-8004 Lifecycle Phase Boundaries

Topic analysis within ERC-8004 was conducted across three consecutive two-month phases spanning the full proposal lifecycle (2025-08-13 to 2026-01-29):

- **Phase 1** (Aug 13 – Oct 13, 2025): Initial submission and early community review. n = 100 records.
- **Phase 2** (Oct 13 – Dec 13, 2025): Sustained review period prior to Last Call designation. n = 11 records.
- **Phase 3** (Dec 13, 2025 – Feb 13, 2026): Last Call through Final ratification (mainnet deployment 2026-01-29). n = 17 records.

Algorithm 1 EIP Lifecycle

```
Data: Stages = {IDEA, DRAFT, REVIEW, LAST_CALL,
    FINAL, STAGNANT, WITHDRAWN}
stage ← IDEA
while stage ≠ FINAL and stage ≠ WITHDRAWN do
    Proceed to NEXT(stage) based on community consensus;
    if No activity for a long period then
        stage ← STAGNANT;
    else if Author decides to cancel then
        stage ← WITHDRAWN;
    end if
end while
Deployment: Theoretically any party MAY deploy registry
independently of stage
```

Algorithm 2 A2A Decision Process (TSC Governance)

```
Input: GitHub Pull Request or Issue (pr)
Output: Decision Status (Merged or Rejected)
if pr is NOT Contested then
    // Apply Lazy Consensus
    Status ← MERGED;
else
    // Escalation to Technology Steering Committee (TSC)
    Initiate GitVote (Restricted to 1 seat per company);
    if Vote == PASS then
        Status ← MERGED;
    else
        Status ← REJECTED;
    end if
end if
return Status;
```

E Decision Architectures

There are 3 types of EIPs: standard track EIP, Meta EIP and Informational EIP. The standards track EIP can be broken down to 4 categories: core, networking, interface and ERC [13]. All EIPs, including ERC-8004 in our case, undergo 5 stages: idea, draft, review, last call and final. EIPs updated continuously are assigned to the special stage of "living". All EIPs are decided by rough consensus: Core EIPs by Ethereum core developers at the meeting called *AllCoreDevs*, and others by common developers and users.

This reflects a structural feature of ERCs as application-layer specifications: unlike Core EIPs, which modify the Ethereum protocol itself and requires hard-fork activation when deployment, ERCs define smart-contract interface standards that any party may implement independently, regardless of the proposal's formal lifecycle status [13].

Google A2A launched under `google/A2A` in April 2025 and was donated to the Linux Foundation in June 2025 [19], migrating to the vendor-neutral `a2aproject/A2A` organization. Governance authority rests with an eight-seat Technical Steering Committee (TSC), comprising representatives each from Google, Microsoft, Cisco, AWS, Salesforce, SAP, IBM, and ServiceNow. Independent contributors cannot join the TSC during an 18-month startup phase [2]. Routine pull requests are merged upon maintainer approval without a formal vote [2]; contested specification changes escalate to a GitVote [3], a GitHub-native ballot in which only TSC members cast binding votes at a 51% threshold.

For pseudocodes, see Algorithm 1 and 2.

F Institution Label Provenance Cascade

Institution labels were assigned using the following two-tier cascade, applied in priority order:

(1) **Manual investigation** (high confidence): A systematic review of GitHub profiles, LinkedIn pages, personal websites, and EIP commit metadata was conducted for the top 109 contributors across both cases (all 71 ERC-8004 participants plus the top 38 A2A contributors by post count). This review yielded 40 institutional upgrades, replacing LLM-inferred labels with verified affiliations. The original LLM-inferred label is preserved in a separate `institution_lm` field for all records to enable sensitivity checks.
(2) **LLM inference** (low confidence): For the remaining 517 authors, institution was inferred by MiniMax-M2.5 from contextual signals in the record text (e.g., repository ownership, self-identification, email patterns where visible).

G Actor Filtering Stages

The three network analyses operate on successively filtered actor sets (Table 12). Starting from the annotated records (ERC: 142 records / 71 contributors; A2A: 4,181 / 778), the co-participation network drops four ERC and seven A2A bot/short-text residuals, yielding $N = 67$ and $N = 771$ respectively. The DNA and socio-semantic analyses further (a) inner-join against the Thematic-LM `coded_records.json` (12 records lose a theme assignment) and (b) exclude records whose stance is *Off-topic* or *Unclassified*, dropping actors whose *only* contributions fall into those categories. This produces the final stance-bearing actor sets of $N = 66$ (ERC-8004) and $N = 710$ (Google A2A). All SNA statistics use $N = 67$ / $N = 771$; all DNA and socio-semantic statistics use $N = 66$ / $N = 710$.

Table 12: Actor and record counts through the three filtering stages.

	ERC-8004		Google A2A	
	Records	Actors	Records	Actors
Annotated (retained)	142	71	4,181	778
SNA (bot & short-text filter)	130	67	4,230	771
DNA / Socio-semantic	126	66	3,759	710

H Related Work Table

Table 13 positions fifteen representative works into three classes by paper type, each evaluated by criteria appropriate to that class. Panel (a) lists *foundational works* that originated the methods or concepts we inherit; panel (b) lists *perspective papers* that advance theoretical claims about DAO and corporate governance without conducting their own empirical tests; panel (c) lists *empirical and methodological studies*—the key methodological works evaluated additionally by openness. This study extends panel (c) with all data and code open.

Table 13: Positioning of this study relative to prior work, organized by paper type. Each panel applies an evaluation rubric appropriate to its class; the synthesis line at the bottom describes how this study bridges all three.

(a) Foundational / Seminal Works

Paper	Year	Core Contribution	Inheritance for this Study
Russell [49] *'Rough Consensus' & the Internet–OSI Standards War*	2006	"Rough consensus" as a standardization mode	Frames ERC-8004's decision rule
Roth & Cointet [48] *Social and Semantic Coevolution in Knowledge Networks*	2010	Coevolution of social and semantic ties	Root of our network–discourse layer
Leifeld [32] *Discourse Network Analysis of Policy Change*	2013	Discourse Network Analysis (DNA)	DNA layer used in §Methods
Beck et al. [5] *Governance in the Blockchain Economy*	2018	IS framework: rights / accountability / incentives	Defines the decentralization axis
Grootendorst [20] *BERTopic: Neural Topic Modeling*	2022	Neural topic modeling with c-TF-IDF	Used in our topic-discovery pipeline

(b) Perspective / Opinion Papers

Paper	Year	Central Claim	Domain	Backing
Murray et al. [41] *Contracting in the Smart Era*	2021	Smart contracts and DAOs reshape agency cost	DAO×Corp	Theory
Lumineau et al. [35] *Blockchain Governance: A New Way of Organizing*	2021	Blockchain as a new mode of organizing	DAO×Corp	Theory
Hui & Tucker [23] *Decentralization, Blockchain, AI*	2025	Decentralization is the right frame for AI governance	AI-Proto	Conceptual
Reineke et al. [47] *Decentralization: A Revolution or a Mirage?*	2025	Decentralization is contested between revolution and mirage	DAO×Corp	Review
Sunyaev et al. [51] *From Ideology to Design: Purposeful Decentralization*	2026	Decentralization should be design-driven, not ideological	DAO×Corp	Conceptual

(c) Empirical / Methodological Studies

Paper	Year	Method	Setting	Data	Code	Comparison
Stine & Agarwal [50] *Comparative Discourse Analysis via Topic Models*	2020	Comparative topic models	Online political discourse	×	×	✓
Qiao et al. [43] *Thematic-LM: LLM-Based Thematic Analysis*	2025	LLM-aided thematic analysis	Generic corpora	×	×	×
Ao et al. [4] *Is DeFi Actually Decentralized?*	2023	Social network analysis	Aave (DAO)	✓	✓	×
Leifeld [32] *Discourse Network Analysis of German Pension Politics*	2013	Discourse network analysis	German pension policy	×	×	×
Roth & Cointet [48] *Social and Semantic Coevolution in Knowledge Networks*	2010	Socio-semantic network analysis	Knowledge networks	×	×	×

I Glossary

Table 14 lists all specialized terms, abbreviations, and acronyms used in this paper, with their definitions, the section where they first appear, and their primary source or citation.

Table 14: Glossary of specialized terms and abbreviations.

Term / Abbreviation	Definition	First Appearance	Citation
DAO (Decentralized Autonomous Organization)	A blockchain-native organization governed by smart contracts and community consensus rather than corporate hierarchy.	§1 Introduction	[22], [5]
EIP (Ethereum Improvement Proposal)	A standards proposal document in the Ethereum ecosystem; the EIP process (defined by EIP-1) governs how such proposals are reviewed and ratified.	§1 Introduction	[13]
ERC (Ethereum Request for Comments)	A subclass of EIP defining application-layer smart-contract interface standards.	§1 Introduction	[13]
Rough Consensus	A decision rule requiring the absence of sustained substantive objection rather than a formal majority vote.	§1 Introduction	[49], [13]
TSC (Technical Steering Committee)	A governance body responsible for technical oversight; in A2A, an eight-seat committee of corporate representatives.	§1 Introduction	[2]
Lazy Consensus	A decision rule whereby proposals are merged by default unless an explicit objection is raised within a review window.	§3 Decision Architectures	[2]
GitVote	A GitHub-native ballot mechanism restricted to TSC members for resolving contested specification changes.	§3 Decision Architectures	[3]
SNA (Social Network Analysis)	Mapping actors as nodes and shared thread participation as undirected edges.	§2 Methodology	[4]
DNA (Discourse Network Analysis)	Projecting actor–theme stance matrices into congruence and conflict networks to capture discursive alignment and opposition.	§2 Methodology	[32]
BERTopic	A neural topic modeling framework combining sentence embeddings, UMAP dimensionality reduction, HDBSCAN clustering, and c-TF-IDF topic representation.	§2 Methodology	[20]
Thematic-LM	An LLM-assisted inductive thematic analysis pipeline with open coding, aggregation, codebook review, and theme assignment stages.	§2 Methodology	[43]
JSD (Jensen–Shannon Divergence)	A symmetric, smoothed measure of distributional distance bounded between 0 (identical) and 1 (disjoint).	§2 Methodology	[34]
UMAP (Uniform Manifold Approximation and Projection)	A non-linear dimensionality reduction technique for embedding high-dimensional data.	§2 Methodology	[37]
HDBSCAN (Hierarchical Density-Based Spatial Clustering)	A density-based clustering algorithm that extracts a hierarchy of clusters without requiring a flat cut.	§2 Methodology	[20]
c-TF-IDF (Class-based TF-IDF)	A term-weighting scheme that adjusts for topic-specific word distributions by treating each topic as a concatenated document.	§2 Methodology	[20]
Louvain algorithm	A community detection method that maximizes modularity via greedy hierarchical agglomeration.	§2 Methodology	[6]
BE coreness (Borgatti–Everett coreness)	A continuous measure of core-periphery structure computed as the correlation between the observed adjacency matrix and an ideal core-periphery pattern.	§2 Methodology	[7]
CP (Core-Periphery)	A network structure characterized by a densely connected core and a sparsely connected periphery.	§2 Methodology	[7]
GCR (Giant Component Ratio)	The fraction of nodes belonging to the largest connected component: GCR = $N_{\max}/N$.	§2 Methodology	[42]
Fleiss' κ	A multi-rater generalization of Cohen's κ measuring inter-annotator agreement beyond chance.	§A.1 (Appendix)	[17]
PSM-DID (Propensity Score Matching with Difference-in-Differences)	A quasi-experimental design combining matching on observed covariates with pre-post treatment comparison to estimate causal effects.	§1 Related Work	[9]
SoK (Systematization of Knowledge)	A structured literature review methodology that taxonomizes and evaluates a research area.	§1 Related Work	[28]
RfC (Requests for Comments)	A Wikipedia deliberation mechanism in which community members solicit structured feedback on contested topics.	§1 Related Work	[25]
DeFi (Decentralized Finance)	A blockchain-based financial ecosystem operating without traditional intermediaries through smart contracts.	§1 Related Work	[22]
dApps (Decentralized Applications)	Applications that run on a blockchain or peer-to-peer network rather than a single centralized server.	§A.2 (Appendix)	[21]
OSS (Open-Source Software / Society)	Software developed under licenses permitting public access, modification, and redistribution; also the norms and institutions of open collaborative production.	§4 Discussion	[18]