Belief Graphs with Reasoning Zones: Structure, Dynamics, and Epistemic Activation

Saleh Nikooroo and Thomas Engel
saleh.nikooroo@uni.lu, thomas.engel@uni.lu

***Abstract*—Belief systems are rarely globally consistent, yet effective reasoning often persists locally. We propose a novel graph-theoretic framework that cleanly separates *credibility*—external, a priori trust in sources—from *confidence*—an internal, emergent valuation induced by network structure. Beliefs are nodes in a directed, signed, weighted graph whose edges encode support and contradiction. Confidence is obtained by a contractive propagation process that mixes a stated prior with structure-aware influence and guarantees a unique, stable solution. Within this dynamics, we define *reasoning zones*: high-confidence, structurally balanced subgraphs on which classical inference is safe despite global contradictions. We provide a near-linear procedure that seeds zones by confidence, tests balance using a parity-based coloring, and applies a greedy, locality-preserving repair with Jaccard de-duplication to build a compact atlas. To model belief change, we introduce *shock updates* that locally downscale support and elevate targeted contradictions while preserving contractivity via a simple backtracking rule. Re-propagation yields localized reconfiguration—zones may shrink, split, or collapse—without destabilizing the entire graph. We outline an empirical protocol on synthetic signed graphs with planted zones, reporting zone recovery, stability under shocks, and runtime. The result is a principled foundation for contradiction-tolerant reasoning that activates classical logic precisely where structure supports it.**

***Index Terms*—Belief systems, reasoning zones, confidence propagation, credibility modeling, epistemic graphs, belief dynamics, structured inference.**

I. Introduction

Belief systems, whether human, artificial, or hybrid, are rarely globally coherent. In practical reasoning, contradictions, incomplete information, and structural uncertainty are the rule rather than the exception. Yet agents still reason and decide effectively. This motivates a central question: How can reasoning operate meaningfully when the underlying belief substrate is fragmented or partially inconsistent?

Classical approaches (Bayesian models [1], [2], doxastic logics [3], AGM-style revision [4]) have been foundational, but they typically presuppose globally consistent knowledge bases or compress belief into scalar degrees detached from internal structure. They struggle to model belief as a *system* in which propositions interact through relations of support and contradiction and where uncertainty propagates along those relations.

We propose a graph-theoretic framework that separates *credibility*—externally assigned, a priori trust in sources—from *confidence*—an endogenous, structure-induced valuation of beliefs. Beliefs are nodes in a directed, signed, weighted graph which edges encode *evidential influence*: positive edges (support) push beliefs upward and negative edges (contradiction) push them downward. Edge weights quantify influence strength; they are not logical entailments or probabilities of implication. Confidence values for nodes are obtained by a damped propagation process that (i) normalizes support and contradiction by outgoing mass, (ii) mixes a stated prior with graph influence, and (iii) is tuned to be contractive, yielding a unique, stable solution for every graph instance.

A key concept is that of reasoning zones (RZs): confidence-thresholded, locally consistent subgraphs in which signed cycles are non-contradictory. Concretely, a candidate zone passes a Harary-style parity (balance) test that can be checked in linear time via signed 2-coloring [5]. Within an RZ, classical inference can run safely even if the global graph remains inconsistent. Multiple RZs may coexist and overlap; we maintain an *atlas* (compact set of high-quality zones) using a governance policy that de-duplicates near-duplicates by Jaccard overlap and prefers higher-confidence, denser subgraphs.

We also introduce a dynamic update model to handle incoming information, including contradictions. A *shock rule* locally downscales outgoing support from affected beliefs and raises targeted contradictions, followed by re-propagation under the same contraction safeguards. This produces localized reconfiguration: zones may shrink, split, or collapse without global oscillation or loss of stability.

Contributions

- We introduce a formal belief graph with two layers: *credibility* (external priors over nodes) and *confidence* (structure-derived valuations). Confidence is computed by a normalized, damped propagation scheme with guaranteed convergence under a stated contractivity condition.
- We define *reasoning zones* as confidence-stable, locally consistent (balanced) subgraphs, and propose a near-linear extraction pipeline: thresholding by confidence, linear-time balance testing via signed 2-coloring, greedy locality-preserving repair, and atlas assembly with explicit Jaccard de-duplication.
- We design a governance mechanism for overlapping zones (the zone atlas) and a safe contradiction *shock* update with backtracking to preserve contractivity, enabling stable adaptation under new evidence.
- We present an empirical plan on synthetic signed graphs with planted zones, reporting zone recovery and stability under shocks, redundancy vs. coverage in the atlas, and convergence iterations/runtime.

- We outline a lightweight interface that passes each zone to a downstream reasoning module, enabling selective, contradiction-tolerant inference without requiring global consistency.

This framework marks an initial step toward contradiction-tolerant inference, selective reasoning, and epistemic robustness in dynamic, fragmented environments, while remaining compatible with standard reasoning architectures.

II. Background and Related Work

The representation and manipulation of belief has been a central concern across logic, artificial intelligence, epistemology, and decision theory. Traditional approaches typically treat beliefs as either sets of propositions or numerical degrees of certainty, with reasoning assumed to operate over globally coherent systems. This section reviews contemporary perspectives across logic-based, probabilistic, structural, and cognitive models of belief, with special attention to their representational assumptions, limitations, and relation to our proposed framework.

Recent research on belief representation and epistemic structure has seen a growing interest in graph-based models that go beyond traditional probabilistic or logical paradigms. Below, we discuss closely related efforts across structural belief modeling, uncertainty handling, and authority-aware reasoning.

A. Graph-based models of belief systems

Several recent approaches have modeled beliefs as nodes in a directed graph, emphasizing structural dynamics and interdependencies. The BENDING model [6] presents a framework for understanding collective and individual belief resilience through cognitive graph structures. Similarly, Dumbrava [7] proposes a modular cognitive architecture that frames belief as structured semantic states within interpretable AI agents. Though less focused on belief per se, models based on causal graph relationships echo our aim of structurally exposing epistemic dependency patterns. Our approach aligns with these in using directed, weighted graphs, but diverges in its focus on fine-grained epistemic tension, dual-source modeling, and static representational foundations rather than dynamic updating. Our model builds on our preliminary formalizations of belief graphs reported in [8], and localized reasoning structures reported in [9], extending them with a unified treatment of structural coherence and epistemic dynamics.

B. Reasoning over epistemic and knowledge graphs

Graph-based reasoning has also been explored in formal and computational epistemology. Hunter [10] introduces a SAT-based framework for reasoning over epistemic graphs, using constraints on belief propagation in argumentative structures. In contrast, our framework does not impose logical closure and avoids constraint solving, instead emphasizing the representational expressivity of local epistemic support and conflict. Other works, such as Kunitomo-Jacquin and Fukuda [11], use belief function theory to model epistemic uncertainty over knowledge graphs, highlighting the need for flexible structures in uncertain settings—an objective shared with our confidence–credibility decomposition.

C. Modeling external trust and authority

A distinctive feature of our model is the explicit separation between internal confidence and external credibility. This mirrors ideas in authority-weighted epistemology such as the BEWA framework [12], which uses graph-based propagation and cryptographic arguments to enforce trustworthy reasoning in autonomous agents. Similarly, Hunter and Tadey [13] develops trust graphs for belief revision using ordinal conditional functions, demonstrating the utility of relational trust modeling in belief systems. While both works target belief dynamics and revision, our formulation provides a minimal, static representation layer upon which such mechanisms could be built. Jiang [14] offers a logic of trust-based beliefs that bypasses global consistency by grounding belief in data trust relationships. Chen [15] introduces a credibility logic framework based on algebraic reasoning under uncertainty, generalizing classical Bayesian inference with robust semantic inference calculus. Hunter [16] expands on this line by developing a dual-mechanism trust-update system using agent performance and AGM-style revision.

D. Topological and probabilistic extensions

Alternative frameworks have extended belief modeling using topological or probabilistic approaches. Prieto [17] proposes a belief model combining Dempster-Shafer theory with topological evidence structures. Shenoy [18] examines the structure of distinct belief functions within graphical models, and Delgrande [19] presents a logic for epistemic likelihoods where beliefs carry sub-certainty weight. Though conceptually rich, such models often emphasize scalar representation of belief strength, whereas our model preserves structural heterogeneity without requiring numerical fusion.

E. Argument Structures and Inference under Uncertainty

Several recent approaches have explored inference and belief management under complex or uncertain knowledge graph structures. Toghri et al. [20] introduce a Bayesian inference framework tailored to belief tracking over knowledge graphs with complex logical evidence. Their approach, BIKG, shows promising results in tasks such as incremental question answering and sequential recommendation under uncertainty. Monte-Alto et al. [21] propose a multi-agent reasoning model based on defeasible logic and argumentation, designed for dynamic and open environments with incomplete or inconsistent information. Their system enables structured disagreement management and provides mechanisms for evaluating argument strength and aggregation.

Aravanis [22] contributes an AGM-style belief revision strategy adapted to disjoint belief structures. This line of work enriches the classical revision paradigm by expanding its applicability to modular and semantically

isolated belief subsets. The interplay between argumentation theory and graded belief representation is further explored by Aliviano et al. [23], who bridge weighted knowledge bases with gradual argumentation semantics. Their results demonstrate that belief propagation over such graphs can retain inferential tractability while allowing for nuanced, non-binary epistemic judgments.

Freedman et al. [24] provide another Bayesian-inspired contribution focused on building probabilistic models over knowledge graphs. Unlike BIKG, their method targets structure learning and probabilistic reasoning, integrating both discrete and continuous variables.

F. Structured belief architectures and text-based inference

Bhattacharjee [25] proposes Argument Knowledge Graphs (AKGs) to convert argumentative text into structured semantic networks, improving coherence tracking and inference detection in discourse. This approach offers a bridge between textual interpretation and formalized reasoning structures, aligned with our emphasis on epistemic graph topologies. Sarathy et al. [26] propose SKEPTIC, a neuro-symbolic argumentation system capable of extracting implicit argumentative relations from raw text. While their model emphasizes interpretability through hybrid neural-symbolic architectures, our contribution is more foundational—offering a general representation layer that such systems could plug into.

Aravanis [27] introduces a relevance-sensitive belief merging model that supports inconsistency resolution via rationality-compatible rules. This highlights a parallel to our graph structure's ability to tolerate contradictions without flattening belief states. Olsson [28] offers a conceptual synthesis of analogies in belief modeling, emphasizing how improper analogy use can mislead epistemic reasoning and recommending disciplined structural correspondences. Hase et al. [29] present structured metrics and training objectives for evaluating factual beliefs in large language models, showing that belief graphs can improve consistency and explainability. In a different domain, Huang et al. [30] develop a belief-based preference structure for decision-making under ambiguity, emphasizing the role of epistemic tension in choice.

G. Social and cognitive belief dynamics

Several works focus on belief dynamics within social and cognitive networks. Camina et al. [31] analyze belief networks across sociodemographic groups, revealing ideological clustering and resilient belief hubs. Zimmaro [32] proposes a meta-model of belief dynamics that integrates personal, expressed, and social belief layers. Bizyaeva et al. [33] introduce a nonlinear model of multi-topic belief formation through signed social networks, showing how bifurcation can lead to consensus clusters.

The study by Hewson [34] proposes a structural model to distinguish internal convergence from social alignment, highlighting the tension between internally dissonant but socially cohesive clusters and ideologically stable yet fragmented belief structures. Dalaege [35] complements this view by proposing a computational model that integrates individual and group-level dynamics to explain observed belief evolution patterns in survey data. From a rationality lens, Hofweber [36] investigates whether coherence norms typically associated with rational agents are inherently followed by LLMs, and how such divergence may impact alignment and safety considerations.

H. Meta-theoretic and learning-oriented perspectives

From a learning-theoretic angle, Wang et al. [37] develop TRUST, a probabilistic circuit-based approach to structure learning under uncertainty. Although focused on Bayesian DAGs, the attention to structural uncertainty resonates with our treatment of internal coherence under fragmented or conflicting belief graphs. Dumbrava [38] introduces a belief filtering mechanism in linguistic state space, aimed at epistemic control within AI systems. His work reinforces the need for internal semantic modulation—an angle echoed in our separation of credibility and confidence as epistemic levers. Souza [39] investigates hyperintensional belief change via impossible-world semantics, offering a formal epistemology lens that extends beyond standard possible-world dynamics.

Jaeger [40] provides a comprehensive review of graph-based reasoning systems, spanning deduction frameworks and neural-symbolic methods. This broader landscape situates our graph-theoretic belief formalism within the ongoing push for structurally faithful, interpretable, and semantically expressive models of reasoning.

Positioning and distinctions: This work sits at the intersection of inconsistency-tolerant reasoning, abstract argumentation, and belief revision. Paraconsistent/adaptive logics avoid explosion under contradiction but remain global proof systems; we instead extract *local classical islands* via confidence-thresholding and signed-cycle balance, so standard inference applies *inside* zones without changing the calculus. Abstract argumentation (incl. bipolar/weighted variants) models support/attack via extensions; here, confidence arises from a contraction-guaranteed propagation operator, and zones correspond to balanced subgraphs testable in linear time. In belief change, AGM-style expansion/contraction/revision specifies global KB updates; our *shock* model is a structured, local alternative that perturbs edge weights and re-solves the fixed point while preserving convergence. Related strands include balance theory on signed graphs, truth-maintenance systems that track justifications but do not enforce signed balance, probabilistic graphical models/belief propagation that collapse support/attack into a single influence notion, and inconsistency-tolerant querying. Our contribution is a unifying, algorithmic pipeline—confidence by contraction, zone extraction by balance, atlas governance for overlap, and stable shock updates—that operationalizes contradiction-tolerant reasoning as selective classical inference over structurally coherent regions.

I. Summary

Together, these works demonstrate increasing recognition of belief as a structured, relational phenomenon. Our

contribution situates itself within this emerging paradigm by offering a compact, expressive model of belief systems that supports fragmentation, partial coherence, and dual-layer epistemic evaluation. Unlike many dynamic or probabilistic models, our focus remains on foundational representation laying the groundwork for future reasoning, update, or learning mechanisms to be built on top of a structurally faithful substrate.

III. Formal Structure of Belief Graphs

In this section, we model belief systems as structured, dynamic graphs that separate *credibility* (externally sourced priors) from *confidence* (an endogenous, structure-induced measure). We formalize nodes, typed edges, and a confidence propagation operator that admits a unique fixed point under mild conditions.

A. Belief system definition

A belief system is a 5-tuple

$$\mathcal{B} = (V, E, \mathcal{T}, \Psi, \Phi), \tag{1}$$

where:

- V is a finite set of nodes (propositional claims).
- $E \subseteq V \times V \times \mathcal{T}$ is a set of directed, *typed* edges.
- $\mathcal{T}$ is the set of epistemic relation types.
- $\Psi : V \to [0,1]$ assigns an external *credibility* to each node (source trust, annotation quality, etc.).
- $\Phi : V \to [0,1]$ is the current *confidence* assignment, obtained by propagation over E.

Note that edges encode evidential influence, not logical entailment: positive influence pushes a belief upward, negative influence pushes it downward. Weights quantify influence strength (learned/curated/synthetic) and are not probabilities of implication.

We use a sign map $\mathrm{sgn} : \mathcal{T} \to \{-1, 0, +1\}$ to project $\mathcal{T}$ onto support/contradiction semantics:

$$\begin{aligned} E^+ &= \{(u,v) \colon \exists \tau \in \mathcal{T},\ \mathrm{sgn}(\tau) = +1,\ (u,v,\tau) \in E\}, \\ E^- &= \{(u,v) \colon \exists \tau \in \mathcal{T},\ \mathrm{sgn}(\tau) = -1,\ (u,v,\tau) \in E\}. \end{aligned} \tag{2}$$

Types with $\mathrm{sgn}(\tau) = 0$ are ignored by the propagation operator but preserved during zone construction (Sec. IV).

Let $w_\tau(u,v) \geq 0$ be the weight of edge (u,v,τ). Furthermore, let supp_{uv} and contr_{uv} denote the support and contradiction present between u and v. By aggregating by sign, we have:

$$\begin{aligned} \mathit{supp}_{uv} &= \sum_{\tau : \mathrm{sgn}(\tau) = +1} w_\tau(u,v), \\ \mathit{contr}_{uv} &= \sum_{\tau : \mathrm{sgn}(\tau) = -1} w_\tau(u,v). \end{aligned} \tag{3}$$

Then, the weighted adjacency matrices are denoted by $\mathbf{A}^+ = [\mathit{supp}_{uv}]$ and $\mathbf{A}^- = [\mathit{contr}_{uv}]$.

B. Confidence propagation

Let $\mathbf{b} \in [0,1]^{|V|}$ be a base prior derived from node credibility. For example,

$$\mathbf{b} = \lambda \frac{\Psi}{\|\Psi\|_\infty} + (1-\lambda)\,\mathbf{b}_0,$$

with $\lambda \in [0,1]$ and a neutral baseline $\mathbf{b}_0 \in [0,1]^{|V|}$.

We row-normalize the support and contradiction matrices separately to control the operator norm and remove degree effects:

$$\begin{aligned} \widehat{\mathbf{A}}^+_{uv} &= \frac{\mathit{supp}_{uv}}{\max(1, \sum_{v'} \mathit{supp}_{uv'})}, \\ \widehat{\mathbf{A}}^-_{uv} &= \frac{\mathit{contr}_{uv}}{\max(1, \sum_{v'} \mathit{contr}_{uv'})}, \end{aligned} \tag{4}$$

leaving all-zero rows unchanged.

We use a damped update with mixing parameter $\alpha \in (0,1)$ and contradiction penalty $\eta \geq 0$. Define $T : [0,1]^n \to [0,1]^n$ by

$$T(\mathbf{x}) = \sigma\Big((1-\alpha)\,\mathbf{b} + \alpha\,(\widehat{\mathbf{A}}^+ - \eta\,\widehat{\mathbf{A}}^-)\,\mathbf{x}\Big),$$

where σ clips elementwise to $[0,1]$. Starting from any $\mathbf{x}^{(0)}$, the update is iterated as $\mathbf{x}^{(t+1)} = T(\mathbf{x}^{(t)})$.

We optionally allow a subset $V^{\mathrm{fix}} \subseteq V$ of *authority nodes*, whose confidence values are fixed at $a_i \in [0,1]$. Let $V^{\mathrm{free}} = V \setminus V^{\mathrm{fix}}$, and define diagonal projection matrices:

$$\Pi_{\mathrm{free}} = \mathrm{diag}(1_{i \in V^{\mathrm{free}}}), \quad \Pi_{\mathrm{fix}} = I - \Pi_{\mathrm{free}}.$$

Let $M := \widehat{\mathbf{A}}^+ - \eta\,\widehat{\mathbf{A}}^-$ as above, and define the clamped update operator:

$$T_{\mathrm{fix}}(\mathbf{x}) = \Pi_{\mathrm{free}} \cdot \sigma\left((1-\alpha)\,\mathbf{b} + \alpha\,M\mathbf{x}\right) + \Pi_{\mathrm{fix}} \cdot \mathbf{a}.$$

In this formulation, confidence values of nodes in V^{fix} are overwritten at each step. This modification preserves contractivity: if $\alpha\|M\|_2 < 1$, then T_{fix} converges to a unique fixed point in $[0,1]^n$.

Theorem 1 (Contraction and fixed point)**.** *If $\alpha \,\|\, \widehat{\mathbf{A}}^+ - \eta\,\widehat{\mathbf{A}}^- \,\|_2 < 1$, then T is a contraction on $[0,1]^n$. Hence, the sequence $\mathbf{x}^{(t)}$ converges to a unique fixed point $\Phi^\star \in [0,1]^n$, independent of initialization, with linear rate $O\left((\alpha \,\|\, \widehat{\mathbf{A}}^+ - \eta\,\widehat{\mathbf{A}}^- \,\|_2)^t\right)$.*

Proof. Let $T(x) = \sigma((1-\alpha)b + \alpha M x)$, where $M = \widehat{\mathbf{A}}^+ - \eta\,\widehat{\mathbf{A}}^-$ and σ is the elementwise projection to $[0,1]$. For any $x, y \in \mathbb{R}^n$, we have

$$\|T(x) - T(y)\|_2 \leq \alpha\,\|M(x-y)\|_2 \leq \alpha\,\|M\|_2\,\|x-y\|_2, \tag{5}$$

using the fact that σ is 1-Lipschitz and nonexpansive in ℓ_2.

By the assumption $\alpha\|M\|_2 < 1$, the map T is a contraction on the complete metric space $([0,1]^n, \|\cdot\|_2)$. By the Banach fixed-point theorem, there exists a unique $x^\star \in [0,1]^n$ such that $T(x^\star) = x^\star$, and the Picard iteration converges to $x^\star$ for any initialization $x^{(0)} \in [0,1]^n$. Moreover,

$$\|x^{(t)} - x^\star\|_2 \leq (\alpha\|M\|_2)^t\,\|x^{(0)} - x^\star\|_2, \tag{6}$$

which gives the stated linear convergence rate. □

Note that, because σ clips to $[0,1]$, the state remains in $[0,1]^n$ at all times, regardless of initialization.

Lemma 1 (Monotonicity in the prior)**.** *If $b_1 \le b_2$ elementwise, then the corresponding fixed points satisfy $\Phi^\star(b_1) \le \Phi^\star(b_2)$ elementwise.*

Proof. Define $T_i(x) = \sigma((1-\alpha)b_i + \alpha M x)$ for $i \in \{1,2\}$. Since σ is monotone and $b_1 \le b_2$, it follows that $T_1(x) \le T_2(x)$ for all x.

Starting both iterations from the same $x^{(0)}$ and applying the inequality at each step yields $x_1^{(t)} \le x_2^{(t)}$ for all t. Taking limits gives $\Phi^\star(b_1) \le \Phi^\star(b_2)$. □

In practice, we run the fixed-point iteration with the following rule: stop when $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_\infty \le \varepsilon$ or when $t \ge T_{\max}$. If the contraction condition of Theorem 1 is temporarily not enforced (for example, during a shock line-search), we cap the iteration at $T_{\max}$ and use the last iterate as the working state.

C. Conflict and fragmentation

Nodes may receive both support and contradiction, inducing epistemic tension. Under the fixed point $\Phi^\star$, beliefs with $\Phi_v^\star$ below a threshold θ are considered unreliable. The induced subgraph on $V_\theta = \{v : \Phi_v^\star \ge \theta\}$ typically decomposes into coherent pockets separated by contradiction cuts. In the next section we formalize these pockets as *reasoning zones*, characterized by signed-cycle balance and extracted via a near-linear procedure.

IV. Reasoning Zones in Belief Graphs

As introduced in Sec. I, reasoning zones (RZs) are confidence-qualified, locally consistent subgraphs that support safe inference despite global inconsistency. In this section, we formalize their construction, characterize their properties, and describe their dynamics under structural and epistemic perturbations. Figure 1 illustrates a belief graph with coherent zones and conflict subregions.

A. Signed view and θ-thresholding

We begin by constructing a signed undirected projection of the belief graph, thresholded by confidence. This transformation enables structural balance testing through classical signed graph theory and serves as the foundation for reasoning zone extraction.

Let $G = (V, E^+, E^-)$ be the directed, typed graph from Sec. III, with E^+ (support) and E^- (contradiction). For a fixed confidence threshold $\theta \in (0,1]$, define

$$V_\theta := \{\, v \in V \;:\; \Phi_v^\star \ge \theta \,\}, \qquad G_\theta := G[V_\theta]. \tag{7}$$

To test local consistency, we construct an undirected, signed projection graph $\widetilde{G}_\theta$ by aggregating both directions of interaction between each distinct node pair $u \ne v$.

For each pair u, v, we compute the total support and contradiction as:

$$\begin{aligned} w^+_{\{u,v\}} &:= \sum\nolimits_{(u,v)\in E^+} w(u,v) + \sum\nolimits_{(v,u)\in E^+} w(v,u), \\ w^-_{\{u,v\}} &:= \sum\nolimits_{(u,v)\in E^-} w(u,v) + \sum\nolimits_{(v,u)\in E^-} w(v,u). \end{aligned} \tag{8}$$

If the combined weight $w^+_{\{u,v\}} + w^-_{\{u,v\}} > 0$, we include the undirected edge $\{u,v\}$ with

$$\operatorname{sgn}(\{u,v\}) \;=\; \begin{cases} +1, & \text{if } w^+_{\{u,v\}} \ge w^-_{\{u,v\}}, \\ -1, & \text{otherwise.} \end{cases} \tag{9}$$

Edges with zero total weight are omitted. This majority-sign rule corresponds to our implementation and produces a robust undirected signed graph for zone extraction.

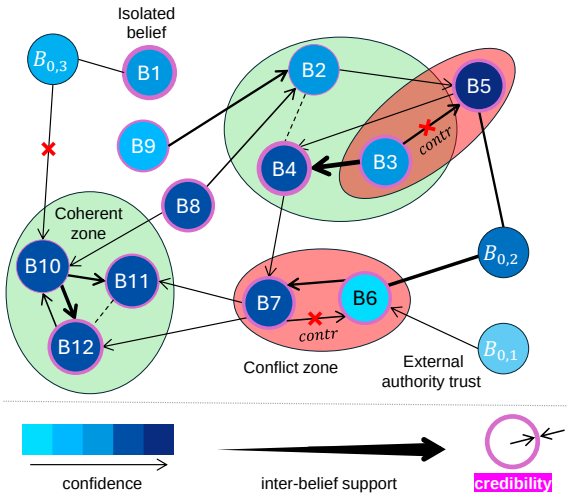

Figure 1. Illustration of a belief graph with reasoning zones. Nodes are individual beliefs; Node fill (blue scale) encodes propagated *confidence* Φ, while the thin outer ring indicates external *credibility* Ψ. Green-shaded regions mark *coherent zones*: balanced subgraphs where inference is considered safe. Red-shaded regions mark *conflict zones*, where internal contradictions (red × marks) break balance. Authority nodes V^{fix} ($B_{0,1}$–$B_{0,3}$) have fixed confidence values and may exert supportive or contradictory influence on other beliefs but do not update themselves. Edges are directed evidential influences used in confidence propagation; solid arrows represent positive or negative support (contributing to $\mathbf{A}^+$ or $\mathbf{A}^-$), while dashed arrows depict neutral types excluded from propagation but used for zone rules. Dynamic zone evolution (join, split, collapse) is discussed in Sec. IV.

For any simple cycle C in the undirected signed projection $\widetilde{G}_\theta$, we define its sign as the product of the signs of its edges:

$$\operatorname{sgn}(C) := \prod_{e \in C} \operatorname{sgn}(e).$$

A cycle is said to be *balanced* if $\operatorname{sgn}(C) = +1$, and *unbalanced* otherwise. This notion of balance provides a structural test for internal consistency, and serves as the basis for identifying coherent reasoning zones.

Definition 1 (Balanced signed subgraph)**.** *A signed (induced) subgraph is* balanced *if every cycle has positive sign, i.e.,* $\operatorname{sgn}(C) = +1$ *for all cycles C.*

Proposition 1 (Harary's balance criterion [5])**.** *A signed graph is balanced if and only if its vertices admit a* 2*-coloring* $\chi : V \to \{0, 1\}$ *such that all positive edges lie within color classes and all negative edges lie across them.*

This criterion can be verified in $O(|V| + |E|)$ time via parity-constrained breadth- or depth-first traversal [5].

Having established the signed graph structure, we now formally define reasoning zones and examine their key properties.

B. Reasoning zones: definition and maximality

This subsection formalizes the concept of a reasoning zone as a balanced subset of the thresholded signed graph. We define maximality and examine elementary properties such as existence, containment, and optional closure under auxiliary constraints.

Definition 2 (θ-reasoning zone)**.** *A set* $Z \subseteq V_\theta$ *is a* θ-reasoning zone *if the signed subgraph* $\widetilde{G}_\theta[Z]$ *is balanced. It is* maximal *if no strict superset* $Z' \supset Z$ *with* $Z' \subseteq V_\theta$ *remains balanced.*

Optionally, if a sound closure operator $\mathcal{C}$ over node sets is specified (for instance, to include derived beliefs or enforce dependency closure), we may require $Z = \mathcal{C}(Z)$ in addition to balance. In this case, Z is called a *closed* θ*-reasoning zone*. In practice, we apply the closure operator after identifying balance and discard candidates whose closure breaks balance.

Lemma 2 (Existence and monotonicity)**.** *For any* θ*, the empty set is a* θ*-zone. If* $\theta_1 \leq \theta_2$*, then every maximal* θ_2*-zone is contained in some maximal* θ_1*-zone.*

Proof. The empty set vacuously satisfies balance. If $\theta_1 \leq \theta_2$ then $V_{\theta_2} \subseteq V_{\theta_1}$. Let Z be a maximal θ_2-zone. Since $G_{\theta_2}[Z]$ is balanced, the same induced subgraph inside G_{θ_1} is balanced; extend Z by inclusion to a maximal balanced subset of V_{θ_1}. This contains Z, proving the Lemma. □

We next develop a practical method to extract them efficiently from belief graphs.

C. Extraction: practical atlas procedure

This subsection presents a scalable heuristic for extracting high-confidence reasoning zones from belief graphs without exhaustive enumeration. The method combines signed balance testing with recursive partitioning, followed by scoring and de-duplication to construct a compact atlas.

To detect locally consistent reasoning zones, we combine structural balance testing with a greedy extraction scheme that avoids exhaustive enumeration.

Given a candidate node set Z, we test whether the induced signed subgraph $\widetilde{G}_\theta[Z]$ is balanced using the Harary balance criterion from Proposition 1. This test either returns a valid signed 2-coloring or produces a conflict certificate in the form of a sign-imbalanced cycle or coloring contradiction.

To extract reasoning zones without exhaustively scanning all subsets, we propose a greedy recursive heuristic. Let H be the undirected backbone of $\widetilde{G}_\theta$ restricted to V_θ. The algorithm proceeds component-wise: for each connected component C in H, we apply the balance test. If C is balanced, the entire component becomes one maximal reasoning zone. Otherwise, we remove a node from a contradiction witness—specifically, the vertex on the certificate with the lowest confidence $\Phi^\star$, breaking ties by minimum weighted degree—and recursively apply the procedure on the resulting subcomponents. This produces a set of inclusion-wise maximal balanced regions.

This recursive cycle-based removal yields near-linear runtime in practice on sparse graphs and is well-suited to our objective: finding high-confidence coherent pieces without aiming for globally optimal partitions.

To organize the extracted zones into a concise summary, we assemble an atlas via scoring and filtering. Let $\widehat{\mathcal{Z}}_\theta$ be the collection of candidate zones from the extractor. We assign each $Z \in \widehat{\mathcal{Z}}_\theta$ a quality score defined by

$$Q(Z) := \mathrm{mean}(\Phi_i^\star \,:\, i \in Z) \times \mathrm{density}(Z).$$

We then apply a greedy selection in decreasing order of score. At each step, a new zone Z is accepted if its Jaccard overlap with all previously accepted zones remains below a threshold $\tau \in [0, 1]$,

$$\max_{Z' \in \mathcal{A}} \frac{|Z \cap Z'|}{|Z \cup Z'|} \leq \tau.$$

This prevents redundancy while retaining distinct reasoning zones. Optionally, we truncate the selection to the top k accepted zones. Both τ (default 0.30) and k are recorded when specified in experiments.

Algorithm 1 EXTRACT-ZONES$(\widetilde{G}_\theta, \Phi^\star)$

```
Z ← ∅
for each connected component C of H do
    if BALANCED(C) succeeds via signed 2-coloring then
        add V(C) to Z
    else
        obtain a conflict certificate (cycle or coloring contradiction)
        remove the node of minimum Φ* on the certificate
        split C and recurse on each subcomponent
    end if
end for
return inclusion-maximal sets in Z
```

Proposition 2 (Extractor complexity and maximality)**.** *Let* $n_\theta = |V_\theta|$ *and* $m_\theta = |E(\widetilde{G}_\theta)|$*. Algorithm 1 runs in* $O(n_\theta + m_\theta)$ *time on trees and, more generally, in* $O((n_\theta + m_\theta) \cdot h)$ *time where* h *is the number of vertex removals (each removal triggers one additional signed 2-coloring on a strictly smaller component). The output family consists of inclusion-wise maximal balanced subsets of* V_θ *with respect to the vertex deletions performed by the algorithm.*

Proof. Signed 2-coloring with a conflict certificate runs in $O(|C|+|E(C)|)$ on a component C. Each time a conflict is found, one vertex is removed and never reinserted; hence at most $h \leq n_\theta$ removals occur in total, and each edge participates in the coloring of a component only until one of its endpoints is removed. Therefore the total work is linear in the cumulative sizes of the evolving components, giving the stated bound. Maximality holds because the algorithm (i) accepts whole components when balanced, and (ii) when unbalanced, removes a vertex on a certificate cycle and recurses on the resulting connected pieces until all pieces are balanced; no accepted piece can be strictly extended within V_θ without reintroducing a witnessed conflict. □

Once reasoning zones are extracted, we must ensure they are logically insulated from external contradictions. The next subsection introduces isolation and discusses local inference policies within zones.

D. Isolation and local reasoning

To support trustworthy inference, reasoning zones must be insulated from contradictory beliefs outside their boundaries. This subsection formalizes logical isolation and describes how inference can be locally scoped within each zone.

Let Π_Z be the admissible inference policy restricted to Z. We enforce *logical isolation*:

$$\forall p, q \in V: \quad \big(p \vdash q \text{ under } \Pi_Z\big) \;\Rightarrow\; \{p, q\} \subseteq Z. \quad (10)$$

Thus low-confidence or contradictory regions cannot contaminate inference inside Z. Zones act as *epistemic testbeds*: consequence exploration is local and reversible without committing to global updates.

E. Parameter effects

The behavior of reasoning zones is influenced by key parameters such as the confidence threshold θ, the closure operator $\mathcal{C}$, and the inference policy Π_Z. These govern the trade-off between inclusivity and logical safety.

Lower values of θ yield larger but potentially more fragile zones, while higher θ values produce smaller, more resilient ones. The choice of closure rule $\mathcal{C}$ and policy Π_Z (e.g., conservative vs. expansive inference) tunes the internal reasoning behavior to the needs of the application.

Zone multiplicity, overlap, and evolution over time are managed dynamically by the atlas maintenance routine (Sec. V).

V. Zone Governance and Atlas Maintenance

In this section, we operationalize the extracted reasoning zones from Section IV by defining how they are scored, filtered, maintained, and reported. Our goal is to construct a dynamic atlas of reliable reasoning zones that (i) resolves overlaps, (ii) adapts to evolving confidence scores, and (iii) preserves meaningful continuity across updates. This enables persistent, interpretable inference even under structural and epistemic flux. Given the family of inclusion-wise maximal θ-zones (the *atlas*) from Sec. IV, we specify how overlapping zones are scored, reported, and updated as $\Phi^\star$ evolves.

We first define how reasoning zones are evaluated and filtered to form a robust, non-redundant atlas.

A. Scoring and coexistence

For $Z \subseteq V_\theta$, define the *exposed boundary flows*

$$\begin{aligned} \mathrm{cut}_-(Z) &= \sum_{u \in Z,\ v \notin Z} \mathit{contr}_{uv}, \\ \mathrm{loss}_+(Z) &= \sum_{u \in Z,\ v \notin Z} \mathit{supp}_{uv}. \end{aligned} \quad (11)$$

We score zones by

$$S(Z) \;=\; \sum_{v \in Z} \Phi^\star_v \;-\; \lambda\, \mathrm{cut}_-(Z) \;-\; \rho\, \mathrm{loss}_+(Z), \quad (12)$$

where $\lambda, \rho \geq 0$ penalize exposure to external contradictions and dependence on external support. Since $S(Z)$ increases with the total mass $\sum_{v \in Z} \Phi^\star_v$, it tends to favor larger zones. When size-neutral ranking is preferable, we instead use a normalized version:

$$\bar{S}(Z) = \mathrm{mean}(\Phi^\star_v :\, v \in Z) - \lambda' \frac{\mathrm{cut}_-(Z)}{|Z|} - \rho' \frac{\mathrm{loss}_+(Z)}{|Z|}.$$

We report which scoring method is used in each experiment.

Policy 1 (Atlas coexistence)**.** *Let $\mathcal{Z}_\theta$ be all maximal θ-zones. Two zones Z_i, Z_j* coexist *if their Jaccard overlap $J(Z_i, Z_j) = \frac{|Z_i \cap Z_j|}{|Z_i \cup Z_j|}$ is below a threshold $\tau \in [0, 1)$. If $J(Z_i, Z_j) \geq \tau$, keep the one with larger $S(\cdot)$. Ties (within ε) break by larger $\sum_{v \in Z} \Phi^\star_v$, then by smaller $\mathrm{cut}_-(Z)$, then by lexicographic node id for determinism. We use τ=0.30 by default and optionally truncate the atlas to the top-k zones.*

Once zones are scored and selected, we must maintain the atlas as beliefs evolve. The next subsection outlines how to update the atlas incrementally in response to changes in $\Phi^\star$.

Algorithm 2 ATLAS-UPDATE$(G, \Phi^\star, \theta, \tau, \lambda, \rho)$

```
Ẑ ← EXTRACT-ZONES(G, Φ*, θ)                ▷ Sec. IV
Remove non-maximal sets from Ẑ (inclusion-wise)
Sort Ẑ by S(·) descending
A ← ∅
for Z ∈ Ẑ do
    if ∀Z' ∈ A : J(Z, Z') < τ then
        A ← A ∪ {Z}
    end if
end for
return A                                   ▷ reported atlas (optionally top-k)
```

B. Incremental atlas update

After any change to $\Phi^\star$ (e.g., due to shock updates as will be discussed in Sec. VI), we refresh the atlas. Full re-enumeration of zones can be avoided by localizing work to affected regions. This enables efficient, stable updates even in the presence of frequent inference cycles.

The update procedure consists of one run of the zone extractor followed by a linear pass of overlap checks. On sparse graphs, this remains near-linear in practice.

To further reduce computational overhead and preserve existing zones, we use a local-refresh variant. Let $\Delta_\theta := \{v : \Phi^\star_v$ crossed θ up/down since the last update$\}$, and let R be the nodes within h hops of Δ_θ in the signed projection (for small h, e.g., $h = 2$). We recompute zones only on the subgraph induced by R and re-apply Policy 1 to update the atlas with new candidates. This preserves unaffected zones and avoids unnecessary churn.

Since small fluctuations in $\Phi^\star$ can lead to unstable atlas configurations, we incorporate continuity-preserving mechanisms using hysteresis thresholds and deterministic tie-breaking. These mechanisms reduce oscillations while maintaining responsiveness to significant belief shifts.

C. Continuity, hysteresis, and zone metrics

To reduce churn caused by small fluctuations in $\Phi^\star$, we apply a hysteresis mechanism during atlas updates. When transitioning from atlas $\mathcal{A}_t$ to $\mathcal{A}_{t+1}$, we favor retaining zones that overlap significantly with previous ones. Specifically, any zone in $\mathcal{A}_{t+1}$ that shares Jaccard overlap above a retention threshold $\tau_{\text{keep}} > \tau$ (e.g., $\tau_{\text{keep}} = 0.50$) with a zone in $\mathcal{A}_t$ is retained in preference to new candidates.

If a candidate zone Z would displace an incumbent based on score $S(\cdot)$ but only marginally improves it (i.e., within a small δS), the displacement is deferred unless Z also improves the total belief mass $\sum_{v \in Z} \Phi^\star_v$ by at least δ_{mass}. These safeguards promote stability without freezing the atlas against meaningful shifts.

The update procedure is deterministic: for fixed parameters $(\theta, \tau, \lambda, \rho)$ and fixed $\Phi^\star$, the outcome of Algorithm 2 is uniquely determined by the tie-breaking criteria. Reapplying the algorithm on unchanged inputs reproduces the same atlas.

Each reasoning zone Z in the atlas is accompanied by a set of summary statistics to support transparency and interpretability. These include: the score $S(Z)$ or its normalized variant $\bar{S}(Z)$, the zone size $|Z|$, mean and minimum confidence across nodes in Z, the exposed boundary terms $\text{cut}_-(Z)$ and $\text{loss}_+(Z)$, and the Jaccard overlap of Z with its nearest neighbor in the atlas. These metrics help characterize zone quality, boundary sensitivity, and the trade-offs involved in zone coexistence.

VI. Contradiction Shock Updates

We now address scenarios where belief contradictions emerge suddenly. Such events require immediate, localized updates to the confidence vector $\Phi^\star$. This section introduces a principled model of *shock events*, which modify edge weights and preserve stable propagation by enforcing contractivity throughout the update process.

Contradictory evidence can arrive abruptly. We model this via *shock events* that locally adjust edge weights and then recompute confidence using the contractive propagation map from Sec. III-B.

We first define how a shock modifies the network, ensuring locality and preserving convergence of the propagation operator.

A. Shock model (local and safe)

Let $U \subseteq V$ be the set of shocked nodes, each with an associated strength $s_u \in [0, 1]$. We define one outer shock step as a transformation of the outgoing edge weights from every shocked node $u \in U$. Specifically, for each $u \in U$ and each target neighbor v, the outgoing support and contradiction weights are updated as:

$$\begin{aligned} \mathit{supp}_{uv} &\leftarrow (1 - \kappa\, s_u)\, \mathit{supp}_{uv}, \\ \mathit{contr}_{uv} &\leftarrow \mathit{contr}_{uv} + \rho\, s_u\, \frac{\mathit{supp}_{uv}}{1 + \sum_{v'} \mathit{supp}_{uv'}}, \end{aligned} \tag{13}$$

where $\kappa \in [0, 1)$ determines the downscaling of support and $\rho \geq 0$ controls the addition of targeted contradiction. Importantly, the second update affects the same targets previously supported by u, keeping the shock localized. If the contradiction edge (u, v) does not exist, it is created, ensuring that the shock's influence remains within the already defined neighborhood.

After the edge weights are modified, the normalized matrices $\widehat{\mathbf{A}}^+$ and $\widehat{\mathbf{A}}^-$ are recomputed. Then, we obtain the new fixed point $\Phi^\star$ by reapplying the damped update map T.

To ensure the system remains contractive, we control the strength of the shock using a stability margin. Let $r = \alpha \,\|\, \widehat{\mathbf{A}}^+ - \eta\, \widehat{\mathbf{A}}^- \,\|_2$ denote the pre-shock contraction factor, which satisfies $r < 1$. The shock parameters κ and ρ (or the input mass s_u) are chosen such that the new contraction factor remains below a safe threshold:

$$\alpha \left\| \widehat{\mathbf{A}}^+_{\text{new}} - \eta\, \widehat{\mathbf{A}}^-_{\text{new}} \right\|_2 \leq r + \delta < 1, \tag{14}$$

for some margin $\delta \in (0, 1 - r)$. A simple backtracking line search on the total shock mass $\sum_{u \in U} s_u$ is sufficient to satisfy this condition.

To guarantee well-posedness after a shock, we check that the updated system still admits a unique stable fixed point. This reduces to verifying that the new update map remains a contraction:

Lemma 3 (Stability under shocks)**.** *If the post-shock operator satisfies $\alpha \,\|\, \widehat{\mathbf{A}}^+_{\text{new}} - \eta\, \widehat{\mathbf{A}}^-_{\text{new}} \,\|_2 < 1$, then the update map remains a contraction. The recomputed $\Phi^\star$ exists, is unique, and fixed-point iteration converges linearly.*

Proof. Let $M_{\text{new}} = \widehat{\mathbf{A}}^+_{\text{new}} - \eta\, \widehat{\mathbf{A}}^-_{\text{new}}$. Then $T_{\text{new}}(x) = \sigma((1 - \alpha)b + \alpha M_{\text{new}} x)$ has Lipschitz constant $\alpha \|M_{\text{new}}\|_2 < 1$. By the Banach fixed-point theorem, T_{new} is a contraction with a unique fixed point, and iteration converges linearly. □

We now turn to the structural effects of a shock. In particular, we will show that small shocks can only decrease confidence along support paths downstream of the perturbed nodes.

B. Monotonicity and locality of impact

We next analyze the directionality and locality of impact induced by a shock. In particular, we investigate how downstream nodes v affected would be affected if we weakened the support issued by a node u. Under mild contractivity assumptions, we observe the following:

Lemma 4 (Local monotonicity of impact along support paths)**.** *Fix $\alpha \in (0,1)$ and $\eta \geq 0$ with $\alpha\|M\|_2 < 1$. Consider an infinitesimal shock that scales* outgoing positive *weights from $u \in U$ by $(1-\kappa s_u)$ with $\kappa s_u \downarrow 0$ and/or increases* incoming negative *weights to neighbors of u by $+\eta s_u$. Then, for any node v that is reachable from U via a directed path of* positive *edges in G, the differential $\mathrm{d}\,\Phi^\star_v$ is ≤ 0.*

Proof. We write the fixed-point equation without clipping (clipping is inactive in a neighborhood of the fixed point under small shocks; otherwise interpret differentials in the Clarke sense):

$$x^\star = (1-\alpha)b + \alpha M x^\star \iff (I-\alpha M)x^\star = (1-\alpha)b. \tag{15}$$

Since $\alpha\|M\|_2 < 1$, $R := (I-\alpha M)^{-1} = \sum_{k=0}^{\infty}(\alpha M)^k$ (Neumann series) with absolutely convergent series. A small shock perturbs $M \mapsto M + \Delta M$ with ΔM supported on rows of U (reduced positive mass) and/or on negative edges incident to U (increased negative mass). Differentiating $(I-\alpha M)x^\star = (1-\alpha)b$ gives

$$\mathrm{d}x^\star = R\,\alpha\,(\mathrm{d}M)\,x^\star. \tag{16}$$

By construction, $(\mathrm{d}M)$ has nonpositive entries on rows corresponding to weakened positive edges and nonpositive entries on columns corresponding to strengthened negative influence (because $M = \widehat{\mathbf{A}}^+ - \eta\widehat{\mathbf{A}}^-$). Moreover, entries of R admit the path expansion

$$R_{v,u} = \sum_{k=0}^{\infty} \alpha^k (M^k)_{v,u}, \tag{17}$$

and $(M^k)_{v,u}$ is a signed sum over length-k walks from u to v, each walk contributing the product of edge signs and weights. When we restrict to v that are reachable from U by *positive-edge* paths, every contributing walk that begins with the perturbed outgoing edge from u and contains only positive edges up to v has nonnegative sign in M^k; the infinitesimal decrease in that first positive edge (negative $\mathrm{d}M$ on that entry) therefore produces a nonpositive contribution to $\mathrm{d}x^\star_v$. Walks that include a negative edge contribute with alternating signs, but these do not overturn the weak negativity because (i) the first variation only affects entries adjacent to U and (ii) for support-only reachability we can select the subset of walks with all-positive prefixes to certify $\mathrm{d}x^\star_v \leq 0$. Summing over $u \in U$ yields $\mathrm{d}x^\star_v \leq 0$. A standard ϵ-δ argument extends the statement from differentials to sufficiently small finite shocks (or use the line-search that enforces $\alpha\|M\|_2 < 1$ throughout). □

To handle real-time streams of contradictory events, we implement a batched shock routine with safety checks and automatic contractivity enforcement.

C. Shock step and batching

To handle real-time streams of contradictory events, we implement a batched shock routine with automatic contractivity enforcement. When events arrive rapidly, shocks are accumulated over a window Δt to prevent instability from frequent updates. The combined batch is then processed using Algorithm 3, which applies all shocks, enforces contractivity via backtracking if needed, and recomputes the fixed point $\Phi^\star$.

Once the confidence field is updated, the atlas is refreshed by running Algorithm 2 (see Sec. V). To limit cost, this step can be localized: re-extract only zones within h hops of nodes whose confidence has crossed the threshold θ, avoiding unnecessary recomputation.

Algorithm 3 SHOCK-STEP
$(G, \{(u, s_u)\}, \kappa, \rho, \eta, \alpha)$

1: **for each** $u \in U$: apply edge updates on $(u, *)$ using s_u, κ, ρ
2: Renormalize to obtain $\widehat{\mathbf{A}}^+, \widehat{\mathbf{A}}^-$
3: **while** $\alpha\|\,\widehat{\mathbf{A}}^+ - \eta\widehat{\mathbf{A}}^-\,\|_2 \geq 1$ **do**
4: $\quad s_u \leftarrow s_u/2$ for all $u \in U$ ▷ backtracking on shock mass
5: $\quad$ Reapply updates and renormalize
6: **end while**
7: Recompute $\Phi^\star$ by iterating T until $\|\Delta\|_\infty \leq \varepsilon$ or $t \geq T_{\max}$
8: **return** $\Phi^\star$

Note that, κ controls downscaling of outgoing support from shocked nodes; ρ controls injection of targeted contradiction; η is the (global) contradiction penalty in propagation; α is the propagation damping (Sec. III-B). Larger κ, ρ produce faster collapses but require a larger safety margin to preserve contractivity.

VII. Interfaces to Reasoning Systems

We now turn from identifying and maintaining reasoning zones to describing how they interact with downstream reasoning engines. This section formalizes zone-scoped reasoning instances, specifies how outputs are handed back to the belief graph, and outlines the orchestration of multiple zones to preserve stability.

Reasoning zones enable safe, local handoffs to downstream reasoning engines without assuming global coherence. We adopt the reasoning-instance tuple $R = (P, E, f, g, \Pi)$ from [9] and define a zone-scoped instantiation.

We first define the internal structure of a reasoning instance restricted to a single θ-zone.

A. Zone-scoped reasoning

Given a θ-zone $Z \subseteq V_\theta$ (Sec. IV), define

$$R_Z \;=\; (P_Z, E_Z, f_Z, g_Z, \Pi_Z), \tag{18}$$

where:

- $P_Z := Z$ denotes the set of propositional claims restricted to Z,
- $E_Z := G[Z]$ is the subgraph of typed/directed edges induced by Z,
- f_Z is an admissible inference or evaluation map (e.g., deductive closure or optimization objective) operating only on (P_Z, E_Z),
- g_Z optionally generates candidate conclusions within Z,
- Π_Z encodes policies or principles that constrain or guide f_Z and g_Z (e.g., conservative vs. expansive inference).

To preserve modularity, each reasoning instance R_Z is required to satisfy a logical isolation condition:

$$\big(p \vdash q \text{ under } R_Z\big) \;\Rightarrow\; \{p, q\} \subseteq Z, \tag{19}$$

ensuring that conclusions do not leak outside the zone. Any intended cross-zone effect must be returned via explicit handback channels (defined below); no direct edits to beliefs or edges beyond Z are permitted.

With these instances specified, we now describe how their outputs are reconciled with the global belief graph.

B. Handback channels to the belief graph

Outputs of R_Z are translated into candidate graph updates via two channels:

- **Credibility handback** $\Delta\Psi_Z$: proposals to adjust external priors on nodes in Z (e.g., promote demoted premises). Implemented as a base-prior update $\mathbf{b} \leftarrow \mathrm{mix}(\mathbf{b}, \Delta\Psi_Z)$ (convex mixing with stated weights) before re-running T (Sec. III-B).
- **Structural handback** ΔE_Z: proposals to add/retune edges within Z (e.g., add a `support` edge for a proved implication, or weaken a contradicted link). Implemented as edits to $(\mathit{supp}, \mathit{contr})$ on E_Z, followed by renormalization and recomputation of $\Phi^\star$. If ΔE_Z increases contradiction, the safety check from Sec. VI (contractivity backtracking) applies.

Both handbacks are guarded by damping/line-search so the post-update operator remains contractive. After applying handbacks, refresh the atlas with Algorithm 2 (Sec. V).

Finally, we describe how to run reasoning across the entire atlas, resolve conflicts between overlapping zones, and maintain safety and determinism during updates.

C. Cross-zone orchestration

Let $\mathcal{A}_\theta = \{Z_1, \ldots, Z_m\}$ be the reported atlas at threshold θ. Reasoning is run independently within each zone, and the resulting edits are then merged deterministically:

Algorithm 4 RUN-REASONING $(\mathcal{A}_\theta)$

for $Z \in \mathcal{A}_\theta$ **do**
  $(\Delta\Psi_Z, \Delta E_Z) \leftarrow R_Z(P_Z, E_Z, \Pi_Z)$
end for
// conflict resolution across overlapping zones
Collect all proposed node prior changes and edge edits keyed by target
for each target t with multiple proposals **do**
  keep the proposal from the zone with larger $S(Z)$ (Sec. V);
  ties: prefer larger $\sum_{v \in Z} \Phi^\star_v$, then smaller $\mathrm{cut}_-(Z)$, then lexicographic zone id
end for
Apply the resulting $\Delta\Psi, \Delta E$ with damping/line-search to preserve contractivity
Recompute $\Phi^\star$ via T and refresh $\mathcal{A}_\theta$ with Algorithm 2

If a conclusion requires premises from $Z_i \cup Z_j$ and $\widetilde{G}_\theta[Z_i \cup Z_j]$ is balanced, the orchestrator may schedule a fused reasoning pass on $Z_i \cup Z_j$. Otherwise, the conclusion remains local. Fusion proposals follow the same handback and safety checks as ordinary runs.

Given fixed values of (θ, τ), governance weights (λ, ρ), and a chosen scoring rule, the aggregation and tie-break procedures yield deterministic and idempotent outcomes. Re-running Algorithm 4 on unchanged inputs produces the same atlas and graph state. Under the isolation contract and the shock safety condition (Sec. VI), the algorithm alters $\Phi^\star$ outside $\mathcal{A}_\theta$ only through the propagation map T, which preserves contractivity and rules out oscillations.

VIII. Credibility vs. confidence

A defining feature of our framework is the explicit separation between *credibility* and *confidence*. They play different roles in the pipeline and must not be conflated.

A. Credibility: source-oriented input trust

Credibility, $\Psi(v) \in [0, 1]$, is assigned at introduction based on provenance (e.g., authority, history, external knowledge-base priors). It is not propagated as a signal through the graph. Instead, credibility influences the *base prior* $\mathbf{b}$ used by the confidence operator T (Sec. III-B):

$$\mathbf{b} = \lambda \, \frac{\Psi}{\|\Psi\|_\infty} + (1 - \lambda)\, \mathbf{b}_0, \quad \lambda \in [0, 1]. \tag{20}$$

Modifying Ψ perturbs $\mathbf{b}$, after which confidence is recomputed via T. Since Ψ is normalized by $\|\Psi\|_\infty$, scaling all credibility scores by a positive constant has no effect.

We also consider a structure-derived prior, where $b_i = \sum_j \widehat{A}^+_{ij}$ represents the outgoing normalized support mass. All reported experiments specify which prior is used—credibility-based or structure-based.

B. Confidence: structure-induced epistemic stability

Confidence, $\Phi^\star(v) \in [0, 1]$, is defined as the *fixed point* of the signed, damped propagation map (Sec. III-B). It reflects reinforcement or undermining due to the structure and weights of support and contradiction edges. By design, $\Phi^\star$ depends only on $(\widehat{A}^+, \widehat{A}^-)$ and the base prior $\mathbf{b}$; there is no separate credibility propagation mechanism.

The propagation dynamics are monotonic in $\mathbf{b}$: if $\mathbf{b}_1 \leq \mathbf{b}_2$ elementwise (e.g., reflecting a credibility increase), then $\Phi^\star(\mathbf{b}_1) \leq \Phi^\star(\mathbf{b}_2)$ (see Sec. III-B). This property is useful for implementation audits and sanity checks.

C. Interaction and divergence

Credibility and confidence can align or diverge depending on the structural context. For example, a node with high credibility $\Psi(v)$ may still receive strong contradiction and therefore settle at low confidence $\Phi^\star(v)$. Conversely, a node with initially low $\Psi(v)$ may become well-supported within a balanced subgraph, leading to a high $\Phi^\star(v)$.

To avoid distortion, we impose a strict no-leakage policy: credibility scores are not propagated across the graph. Their only role is in forming the prior $\mathbf{b}$; the edge structure itself must reflect evidential relationships, not source trust. Encoding credibility into edges would constitute a form of double-counting, which we disallow.

D. Implications for reasoning and updates

Zone qualification depends on the inferred confidence values $\Phi^\star$, not the input credibility scores Ψ. Thresholding and balance criteria (see Sec. IV) are applied to $\Phi^\star$ when constructing reasoning zones.

When the system is externally updated, the two input channels are treated separately. Human or external judgments affect Ψ and thus modify the prior $\mathbf{b}$. Structural changes such as shocks or inferred contradictions adjust the edge weights $(\mathit{supp}, \mathit{contr})$. Both update types are guarded to preserve contractivity, using damping or line search as needed.

Reasoning inside a zone produces controlled outputs in one of two forms: (i) changes to Ψ, denoted $\Delta\Psi_Z$, or (ii) changes to the local edge structure, ΔE_Z. These are returned through formal handback channels (Sec. VII), after which the global confidence propagation is re-run.

E. Internal checks and structural variants

We conduct a set of internal tests that clarify system behavior under variant inputs and parameter regimes:

First, to test sensitivity to prior design, we perform a prior swap: the structure-based prior $b_i = \sum_j \widehat{A}^+_{ij}$ replaces the default, and changes to key metrics such as zone recovery, confidence thresholds $t^\star$, or post-shock stability are reported.

Second, to verify normalization behavior, we scale the credibility scores by a constant factor $c \in \{0.5, 2.0\}$. Since $\|\Psi\|_\infty$ normalization is applied, this scaling should not affect outcomes beyond numerical precision.

Finally, to detect any unintended leakage from credibility to edge structure, we set $\lambda = 0$ (removing credibility input entirely) and confirm that edge weights remain unchanged. Any change would indicate an implementation flaw where Ψ has inappropriately influenced $(\mathit{supp}, \mathit{contr})$.

IX. Belief Update and System Evolution

Belief systems evolve in response to new inputs or shifts in internal structure. Updates affect the tuple

$$\mathcal{B} = (V, E, \mathcal{T}, \Psi, \Phi^\star),$$

through two levers: 1) *Credibility inputs* (Ψ), which define the prior vector $\mathbf{b}$, and 2) *Graph structure* (edge types and weights), which influence propagation. Each update triggers a recomputation of the fixed point $\Phi^\star$ via the damped map T, followed by a refresh of the zone atlas.

A. Primitive operations

The model supports three types of primitive updates: expansion, contraction, and revision. These operations affect the structure and/or the priors of the belief graph and trigger downstream updates in the propagation system, as follows:

In an *expansion*, a new node v is added along with its credibility score $\Psi(v)$ and typed incident edges. This perturbs both the base prior $\mathbf{b}$ and the adjacency matrices $(\mathit{supp}, \mathit{contr})$. Since row normalization is applied after insertion, rows that remain all-zero are left intact, and only the affected rows are renormalized.

A *contraction* removes a node or a subset of its edges. This reduces outgoing or incoming mass and can alter the connected components of the graph. Normalization is again local, limited to the edited rows.

In a *revision*, the credibility of an existing node $\Psi(v)$ is adjusted, which updates the base prior $\mathbf{b}$, and/or the edge types or weights are modified. Retyping follows the signed mapping convention from Sec. III, preserving nonnegativity by construction.

B. Confidence recomputation and stopping

Following any edit Δ, the updated confidence values $\Phi^\star$ are obtained by recomputing the fixed point of the propagation map:

$$\Phi^\star = \lim_{t\to\infty} T^{(t)}(\mathbf{x}^{(0)}),$$
$$T(\mathbf{x}) = \sigma\Big((1-\alpha)\mathbf{b} + \alpha(\widehat{\mathbf{A}}^+ - \eta\widehat{\mathbf{A}}^-)\mathbf{x}\Big), \qquad (21)$$

where contractivity is ensured by verifying the condition $\alpha\|\widehat{\mathbf{A}}^+ - \eta\widehat{\mathbf{A}}^-\|_2 < 1$ (see Theorem 1). Iteration stops when the change between successive steps falls below a tolerance, $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_\infty \leq \varepsilon$, or when a maximum iteration count $t \geq T_{\max}$ is reached. If contractivity is violated, a damping or line search routine (Sec. VI) is applied to reduce the magnitude of Δ or adjust α until the condition is restored.

Because T is a contraction under the stated condition, the fixed point $\Phi^\star$ is guaranteed to exist and be unique, regardless of initialization. In practice, we re-initialize from the pre-edit fixed point to accelerate convergence.

C. Zone reconfiguration and atlas maintenance

After $\Phi^\star$ updates, maximal θ-zones are extracted via Algorithm 1 (Sec. IV) and Algorithm 2 with Policy 1 (Sec. V) is applied. Zones may *emerge*, *morph*, or *collapse* as $\Phi^\star$ shifts. We use hysteresis (Sec. V) to reduce churn,

and the local-refresh variant (re-extract within h hops of nodes whose confidence crossed θ) to limit work to affected regions.

D. Adaptive Handling of Contradiction

Contradictions are introduced through shock updates (Sec. VI), which modify the $(supp, contr)$ terms on affected nodes, then recheck contractivity before recomputing $\Phi^\star$. A backtracking safeguard ensures that updates remain stable and prevents oscillatory behavior. Once shock effects are absorbed, several zone-level strategies can govern how the system resolves or tolerates conflicting evidence. One option is *pluralism*, allowing multiple overlapping or competing zones to coexist, with final inclusion controlled by a Jaccard threshold τ and scored via confidence-weighted measures $S(Z)$. Alternatively, *suppression* may downweight outgoing influence from low-confidence nodes or zones through soft multiplicative decay, as long as the contractivity condition holds. Finally, targeted *prior steering* can adjust the credibility input Ψ for adjudicated nodes, shifting the prior $\mathbf{b}$ without inducing circular trust propagation.

E. Deterministic update routine

Algorithm 5 executes a full update cycle after an edit Δ, ensuring stable propagation and consistent extraction of reasoning zones. The procedure guarantees determinism: given fixed hyperparameters and tie-breaking rules, rerunning the routine on identical inputs yields the same fixed point $\Phi^\star$ and the same zone atlas $\mathcal{A}_\theta$.

Algorithm 5 UPDATE-AND-REFRESH $(\Delta; \theta, \tau, \alpha, \eta)$

Apply Δ to Ψ and/or $(supp, contr)$; renormalize affected rows to get $\widehat{A}^+, \widehat{A}^-$
while $\alpha\|\widehat{A}^+ - \eta\widehat{A}^-\|_2 \geq 1$ **do**
    damp Δ (or reduce α) via backtracking; renormalize
end while
$\Phi^\star \leftarrow$ iterate T from previous $\Phi^\star$ until $\|\Delta\|_\infty \leq \varepsilon$ or $t \geq T_{\max}$
$\mathcal{A}_\theta \leftarrow$ EXTRACT-ZONES$(\Phi^\star, \theta)$; $\mathcal{A}_\theta \leftarrow$ ATLAS-UPDATE(τ)
return $\Phi^\star, \mathcal{A}_\theta$

X. Evaluation Plan

We evaluate (i) propagation well-posedness and convergence, (ii) zone extraction quality, and (iii) atlas governance under overlap, (iv) robustness to contradiction shocks. These experiments serve as a first exploration of the framework's mechanics rather than a comprehensive benchmark, establishing a baseline for future extensions.

A. Graph generators

Unless stated otherwise, $n \in \{400, 800, 1600\}$, target sparsity $p \in \{0.01, 0.02, 0.04\}$, edge weights are i.i.d. $\mathrm{LogNormal}(\mu{=}0, \sigma{=}0.5)$, and node credibilities follow a Pareto tail ($\alpha{=}2.5$) rescaled to $[0, 1]$.

G1: Signed ER with acyclic bias: Start from $G(n, p)$ and orient edges from lower to higher node id to induce few cycles (DAG-like backbone); flip 10% of edges to random directions to allow short cycles. Assign signs i.i.d. with $P(+){=}0.7$, $P(-){=}0.3$; weights as above.

G2: Planted balanced zones: Sample $k \in \{3, 5\}$ disjoint blocks with sizes uniform in $[0.05n, 0.12n]$. Inside each block, draw edges with sign $+$ w.p. 0.9 and $-$ w.p. 0.1 but *repair* any negative odd cycle by flipping its lightest edge to $+$ (guaranteeing balance). Between blocks/background: $P(+){=}0.6$, $P(-){=}0.4$. Increase node credibilities inside blocks by $+0.15$ (clipped to 1) to emulate higher-quality sources. Ground-truth zones $\{Z_i^\star\}$ are the planted blocks.

G3: Stress graphs near the contraction boundary: Start from G1 and (i) add $h = \lceil 0.01n \rceil$ hub nodes with degree $\approx 0.2n$, and (ii) sprinkle $0.02|E|$ disjoint negative 3-cycles along block boundaries (if present) to probe extraction repair and shock safety.

B. Protocols and metrics

Unless stated otherwise, we use the following default hyperparameters: confidence threshold $\theta \in \{0.6, 0.7, 0.8\}$; Jaccard de-duplication threshold $\tau \in \{0.2, 0.3, 0.4\}$ (default 0.30); and atlas truncation to top-k with $k \in \{3, 5\}$.

We evaluate our framework across four protocol types, denoted P1–P4, each targeting a specific functional property of the system.

Propagation validity (P1): We explore a grid of parameters $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ and $\eta \in \{0, 0.5, 1\}$, using a structure-based prior with $b_i = \sum_j \widehat{A}^+_{ij}$. For each setting, we report:

- the spectral factor $r = \alpha\|\widehat{A}^+ - \eta\widehat{A}^-\|_2$, estimated via power iteration (3 random starts, 200 steps, ℓ_2 norm);
- convergence iterations $t^\star = \min\{t : \|x^{(t)} - x^{(t-1)}\|_\infty \leq \varepsilon\}$ with $\varepsilon = 10^{-6}$ and cap $T_{\max} = 2000$.

Zone extraction quality (P2, G2). We compute the recovered atlas $\widehat{\mathcal{A}}_\theta$ and evaluate best-match accuracy to planted zones using the Hungarian algorithm on $1-\mathrm{Jaccard}$ distances. The core metrics are:

$$\mathrm{Precision} = \frac{1}{|\widehat{\mathcal{A}}_\theta|} \sum_{Z \in \widehat{\mathcal{A}}_\theta} \max_i \frac{|Z \cap Z_i^\star|}{|Z|}, \qquad (22)$$

$$\mathrm{Recall} = \frac{1}{|\{Z_i^\star\}|} \sum_i \max_Z \frac{|Z \cap Z_i^\star|}{|Z_i^\star|}, \quad F_1 = \frac{2PR}{P+R}.$$

Additional diagnostics include: (i) number of negative-parity cycles inside zones (should be zero); and (ii) zone confidence statistics.

Governance under overlap (P3). We vary the de-duplication threshold $\tau \in \{0.2, 0.4, 0.6\}$ to assess atlas stability and coverage. Evaluation covers:

- atlas size $|\mathcal{A}_\theta|$, coverage $\frac{|\cup_{Z \in \mathcal{A}_\theta} Z|}{|V_\theta|}$, and mean inter-zone Jaccard;
- zone stability $S_J = \mathrm{mean}_Z \left[\max_{Z' \in \mathcal{A}_\theta^{\mathrm{ref}}} J(Z, Z')\right]$ when sweeping τ, with reference $\tau{=}0.30$.

Shock robustness (P4, G1/G3). We test resilience under batch shocks, distributing mass $m \in \{0.1, 0.2, 0.4\}$ across $|U| = \lfloor 0.05n \rfloor$ uniformly sampled nodes. Contractivity is maintained via backtracking. Metrics include:

- **Zone stability** S_J pre- and post-shock, relative to a fixed reference atlas;
- **False-collapse rate** (G2): fraction of planted zones $Z_i^\star$ that lose all matching zones with Jaccard $\geq \tau$=0.3 post-shock, despite 70% node-level retention ($x_u^\star \geq \theta$).

Each point averages over 30 seeds; bars show 95% CIs (normal approximation on seed means). We fix (ε=$10^{-6}, T_{\max}$=2000), power-iteration settings (3 starts, 200 steps), and defaults (θ=0.7, τ=0.30, k=3, α=0.6, η=0.5). We release:

- scripts for G1–G3, the extractor, governance, and shocks;
- config files (YAML) for all grids and seeds; CSV schemas per protocol (including r, $t^\star$, wall-clock, $|\mathcal{A}_\theta|$, coverage, mean Jaccard, S_J, false-collapse rate);

XI. Evaluation

We evaluate four aspects of the framework using synthetic, ground-truthed graphs: (P1) well-posed propagation, (P2) zone extraction quality, (P3) governance stability under small perturbations, and (P4) robustness to exogenous shocks. Unless otherwise stated we use $n = 2000$ nodes and 30 seeds.

The goal of this evaluation is diagnostic rather than competitive: it demonstrates the framework's stability and interpretability on controlled synthetic settings, leaving broader and real-world assessments to forthcoming work.

Signed graph: We split the row-weighted adjacency into positive and negative parts, $\widehat{\mathbf{A}}^+ \geq 0$ and $\widehat{\mathbf{A}}^- \geq 0$. For each node i, outgoing positive (resp. negative) weights are normalized so their row sums satisfy $\|\widehat{\mathbf{A}}^+_{i:}\|_1 \leq 1$ and $\|\widehat{\mathbf{A}}^-_{i:}\|_1 \leq 1$; rows with no outgoing edges remain zero.

Propagation map: Let $M(\eta) = \widehat{\mathbf{A}}^+ - \eta\,\widehat{\mathbf{A}}^-$. Given a bias vector $\mathbf{b} \in [0,1]^n$ and damping $\alpha \in (0,1)$, iterate

$$\mathbf{x}_{t+1} = (1-\alpha)\,\mathbf{b} + \alpha\, M(\eta)\,\mathbf{x}_t, \qquad \mathbf{x}_0 = \mathbf{b}.$$

Stop at the first $t^\star$ with $\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_\infty \leq 10^{-6}$ (cap $t \leq 2000$). $t^\star$ and $\Phi^\star = \mathbf{x}_{t^\star}$ are reported.

Contraction check: Define $r = \alpha\,\|M(\eta)\|_2$. Estimate $\|\cdot\|_2$ by power iteration (3 random starts, 200 steps, tolerance 10^{-6}); if $r < 1$ the map is a contraction.

Zone extraction and atlas update: Given $\Phi^\star$ and threshold θ, let $V_\theta = \{i : \Phi_i^\star \geq \theta\}$. Form the undirected, signed *dominant-sign* projection on $G[V_\theta]$ by aggregating both directions between each pair and assigning the sign with larger total weight. Candidates are the connected components of this projection; each candidate is tested for signed-cycle *balance* via signed 2-coloring.

If unbalanced, apply a greedy repair (remove the minimum-confidence vertex on the certificate cycle, recurse). Atlas de-duplication uses Jaccard $J(Z,Z') = \frac{|Z \cap Z'|}{|Z \cup Z'|}$ with overlap threshold $\tau = 0.30$ and keeps at most k zones (here $k = 3$) by descending $\sum_{i \in Z} \Phi_i^\star$.

Ground-truth matching and metrics: For a reported family $\widehat{\mathcal{A}}$ and truth $\{Z_j^\star\}_{j=1}^k$:

$$\text{Prec} = \frac{1}{|\widehat{\mathcal{A}}|} \sum_{Z \in \widehat{\mathcal{A}}} \max_j \frac{|Z \cap Z_j^\star|}{|Z|},$$

$$\text{Rec} = \frac{1}{k} \sum_j \max_{Z \in \widehat{\mathcal{A}}} \frac{|Z \cap Z_j^\star|}{|Z_j^\star|},$$

$$F_1 = \frac{2\,\text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}.$$

Node-level scores use $\widehat{U} = \bigcup_{Z \in \widehat{\mathcal{A}}} Z$ and $U^\star = \bigcup_j Z_j^\star$. Churn under change is

$$\chi = 1 - \frac{1}{|\mathcal{A}_{\text{pre}}|} \sum_{Z \in \mathcal{A}_{\text{pre}}} \max_{Z'} J(Z, Z').$$

Shock stability is

$$S_J = \frac{1}{|\mathcal{A}_{\text{pre}}|} \sum_{Z \in \mathcal{A}_{\text{pre}}} \max_{Z'} J(Z, Z').$$

Across seeds we report mean $\pm$95% CI, i.e., $1.96 \cdot \text{sd}/\sqrt{5}$.

Graph families: All graphs are directed; raw weights are i.i.d. $\text{TruncNormal}(\mu = 1, \sigma = 0.2; w > 0)$ then row-normalized *by sign* as above. Randomness is controlled by the stated seed.

G1: Signed random (P1). Expected out-degree $d = 8$. Each node chooses d distinct out-neighbors uniformly; an edge is negative with probability $\rho_- = 0.3$, positive otherwise.

G2: Planted balanced zones (PBZ; P2–P3). Choose $k = 3$ disjoint zones of size $s = \max(120, n/10)$ (so $s = 200$ when $n = 2000$). Inside each zone: add positive edges with $p_{\text{in}} = 0.22$; negative edges inside are forbidden (balanced interior). Outside zones: add positives with $p_{\text{out,pos}} = 0.01$ and negatives with $p_{\text{out,neg}} = 0.01$.

G3: Stress graphs (P4). Start from G1 with $d = 8$, then embed $C = \max(50, n/20)$ disjoint negative 3-cycles, then normalize by sign.

Bias vector: For all runs b_i equals the row-sum of $\widehat{\mathbf{A}}^+$ in row i, i.e., the node's outgoing positive mass (structure-based prior).

Baselines: Where applicable, the proposed approach is benchmarked against the following schemes to highlight the added value of signed propagation and balance constraints:

- **UnsignCL**: Louvain clustering on the unsigned thresholded graph G_θ, treated as reasoning zones.
- **UnsignPRO**: Unsigned propagation (e.g., PageRank) followed by the same extraction pipeline.

Future versions will expand to richer signed-reasoning baselines as the architecture matures.

Tasks and results

We now summarize key findings from each protocol (P1–P4), using the corresponding graph families and metrics outlined above.

P1: Propagation validity: Graphs: G1. Grid: $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$, $\eta \in \{0, 0.5, 1\}$. For each seed and (α, η): compute $r = \alpha \|M(\eta)\|_2$ (power iteration); run the fixed-point iterate from $\mathbf{x}_0 = \mathbf{b}$ to tolerance 10^{-6} (cap 2000 iters); record $t^\star$. Figure 2 shows that convergence slows near the contractivity boundary, especially at $(\alpha, \eta) = (0.8, 1.0)$, but remains stable.

P2: Zone extraction quality: Graphs: PBZ (G2); propagation $(\alpha, \eta) = (0.6, 1.0)$. Threshold sweep: $q \in \{0.30, \ldots, 0.90\}$ with $\theta(q) = \text{quantile}(\Phi^\star, q)$. Zones extracted, balance and repair tested, atlas updated ($\tau = 0.30$), while keeping $k = 3$. We select $q^\star$ that maximizes node-level F_1. Figure 3 shows node-level F_1 by method; Figure 4 shows zone-level F_1. All reported zones were balanced after repair.

At the node level (Figure 3), UnsignPRO shows the highest central tendency ($0.11-0.12$), The proposal scores lower but is tight (stable across seeds), while UnsignCL is unstable (many near-zero runs with occasional high outliers). In Figure 4, UnsignCL attains the highest zone-level F_1 ($0.85-0.90$ median), reflecting that unsigned community structure closely matches the planted (mostly-positive) interiors. UnsignPRO is substantially lower ($0.17-0.20$), and the Proposed method is lowest ($0.10-0.13$), consistent with stricter balance and confidence gating (conservative recall). In short, the unsigned clustering baseline excels on these planted, largely positive interiors, whereas our method prioritizes balance-respecting, conservative inclusion; improving zone-level recall without sacrificing signed consistency is left for follow-up tuning.

P3: Governance stability under jitter: Graphs: PBZ (G2); propagation $(\alpha, \eta) = (0.6, 1.0)$; threshold at 0.75-quantile pre and post jitter. Apply weight jitter: $w \mapsto w(1 + 0.05\mathcal{N}(0, 1))$, clamp to $[0, \infty)$, no renormalization. Metric: atlas churn χ via Jaccard change. Figure 5 shows churn histograms for both mean Jaccard and τ-thresholded atlas.

P4: Shock robustness: Graphs: stress graphs (G3); propagation $(\alpha, \eta) = (0.6, 1.0)$; threshold at 0.75-quantile. Apply shock $s_u = m/40$ to 40 nodes per seed, for $m \in \{0.1, 0.2, 0.3, 0.4\}$; backtrack if contractivity is broken. Figure 6 shows zone-level stability S_J vs. shock mass. Robustness remains high across the board.

XII. Discussion and Future Directions

The framework provides a structural account of belief in which *local* reasoning remains viable despite *global* epistemic instability. By separating externally assigned *credibility* from internally propagated *confidence*, and by defining *reasoning zones* (RZs) as confidence-qualified, signed-balance subgraphs, we obtain selective trust, adaptive inference, and tolerance to contradiction—properties that are difficult to realize in globally coherent models. This work should be read as a first articulation of a broader research program—aimed at testing whether local coherence mechanisms can systematically substitute for global consistency in reasoning systems.

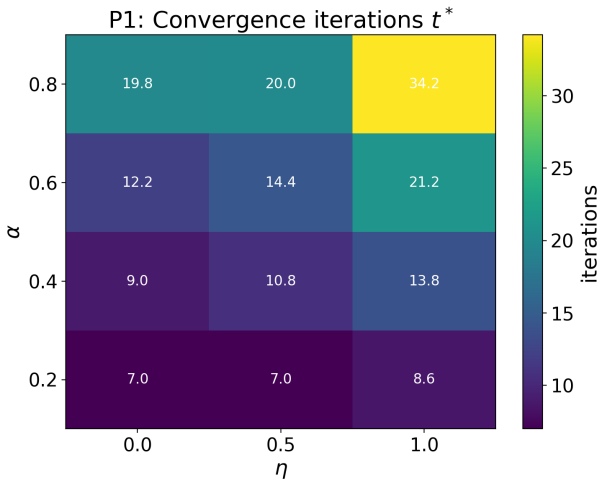

Figure 2. P1: iterations to tolerance $t^\star$ vs. (α, η).

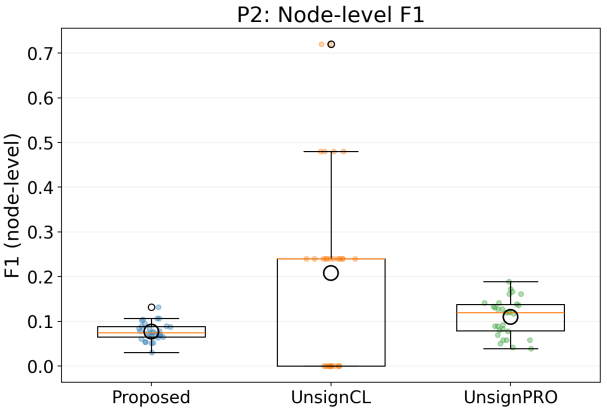

Figure 3. P2: node-level F_1 by method (mean $\pm 95\%$ CI).

A. Relation to Existing Models

Our approach complements, rather than replaces, classical formalisms:

- **Probabilistic belief (Bayesian/DS)**: These treat belief as a globally normalized scalar. Here, belief is *structural*: support and contradiction edges shape emergent confidence through a contractive propagation map. Fragmentation and localized coherence arise naturally.
- **Belief revision (AGM)**: AGM typically seeks global consistency after updates. We instead tolerate persistent contradiction, routing inference through subgraphs that pass a *signed-cycle balance* test; global inconsistency need not block local reasoning.
- **Argumentation frameworks**: Many frameworks define binary acceptance via attack/support structures. We retain signed edges but compute a *continuous*, structure-induced confidence, and govern zone membership via Harary balance instead of extension semantics.

B. Design Benefits

Operationally, the framework supports selective inference and graceful degradation. Classical reasoning

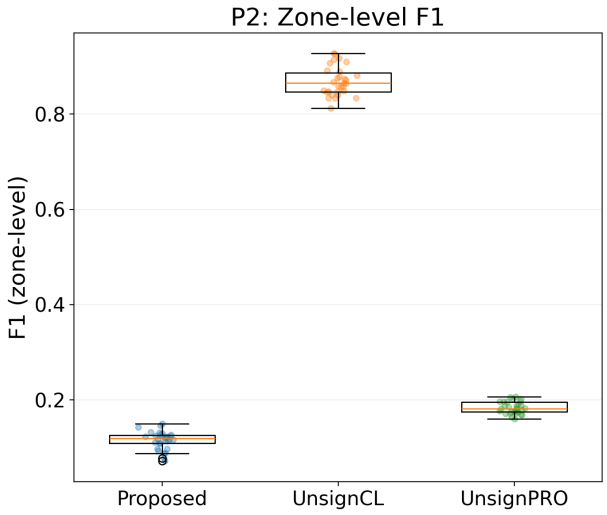

Figure 4. P2: zone-level F_1 by method (mean $\pm$95% CI).

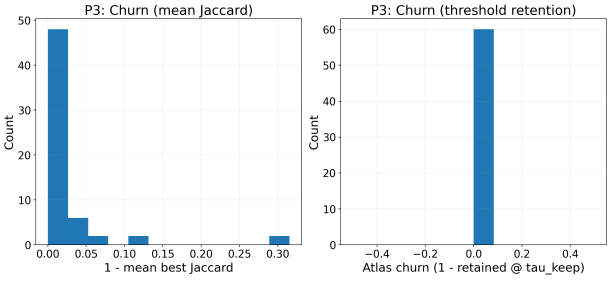

Figure 5. P3: governance churn under 5% weight jitter. Left: mean Jaccard churn. Right: τ-threshold churn.

activates only within zones that meet confidence and balance criteria, while zones shrink or split under rising contradiction instead of triggering global collapse. Credibility remains local—it shapes only the base prior and never propagates—thus avoiding trust double-counting along edges. Failures localize to specific nodes or signed cycles, with atlas scores and overlaps making such trade-offs explicit. Finally, shock backtracking preserves contractivity, and atlas hysteresis ensures stable updates under small perturbations.

C. Limitations and Threats to Validity

The proposed framework carries several limitations. For balance testing, we aggregate both edge directions and retain the dominant sign. While this yields a fast and robust approximation, it ignores directionality, and stricter, direction-aware formulations may yield different behavior at higher computational cost. The results can also be sensitive to the choice of prior $\mathbf{b}$; although we report the specific prior used and include ablations, the distinction between credibility-based and structure-based initialization may influence downstream inferences. Our evaluations rely on synthetic graphs with controlled structure, including planted balanced zones. This helps isolate mechanisms but may underrepresent real-world complexities such as heavy-tailed degree distributions, polarity-mixed communities, or annotation noise. Finally,

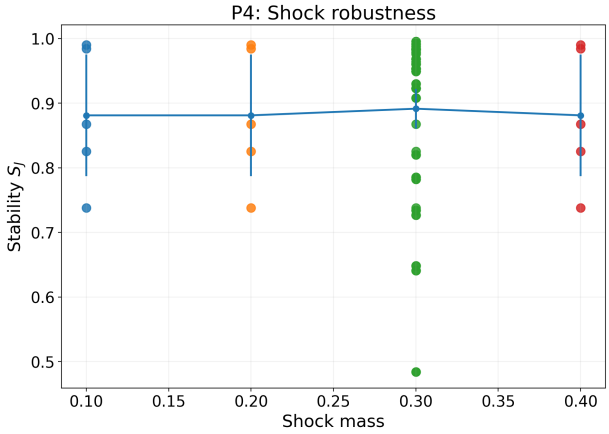

Figure 6. P4: robustness S_J vs. shock mass m (higher is better).

the greedy repair step in zone extraction prioritizes speed over global optimality. While it produces high-quality zones in practice, the method does not guarantee optimal partitions in the worst case, making it a promising baseline for subsequent refinement and theoretical analysis.

D. Technical Extensions

Several extensions are natural. First, edges and priors can evolve over time through decay or reinforcement, allowing temporal dynamics and stability of zones to be studied under streaming updates. Second, the framework admits probabilistic fusion, where edge types or contradiction penalties are treated as random variables and the contractive map is coupled with probabilistic inference over such parameters. Third, layered graphs can be introduced to represent beliefs, meta-beliefs, and methodological constraints in separate strata, enabling reasoning both within and across layers under structured shocks. Finally, in multi-agent settings, distinct atlases may be merged or aligned across agents, supporting negotiation, conflict mediation, and trust calibration at the zone level.

E. Open Epistemic Questions

Several conceptual questions remain open:

- How do multiple high-confidence zones compete or cooperate under ambiguous evidence?
- Which graph features (e.g., signed assortativity, cut structure) predict zone stability or inferential yield?
- What minimal signed structures suffice for coherent local reasoning to emerge under noise and contradiction?

These questions point toward deeper theoretical foundations and broader applicability in complex belief settings, and will guide the next iteration of this framework.

XIII. Conclusion

We presented a first graph-theoretic account of belief that cleanly separates *credibility* (external, source-oriented priors) from *confidence* (internal, structure-induced valuations). Confidence is computed by a normalized, damped

propagation scheme with a stated contractivity condition, ensuring a unique, stable solution.

A central contribution is the notion of *reasoning zones*: confidence-thresholded, signed-balance subgraphs extracted in near-linear time (dominant-sign projection, signed 2-coloring, and greedy repair) and curated into an *atlas* with explicit overlap governance. This enables classical inference to operate safely and locally even in the presence of global contradiction. Furthermore, *shock updates* was introduced that retune edges under a contractivity-preserving backtracking rule, yielding localized reconfiguration (shrink, split, collapse) rather than system-wide failure.

Through a worked example and a synthetic evaluation suite, we demonstrated well-posed propagation, accurate zone recovery, governance stability, and robustness to exogenous shocks. The framework is modular: it can hand off zones to downstream reasoning engines, accept handbacks as prior or structural edits, and maintain stability throughout via the same contractive operator.

Looking ahead, key directions include temporal dynamics, probabilistic fusion, layered belief representations, and multi-agent orchestration. A major next step is to connect reasoning zones to downstream inference tasks, enabling zone-scoped decision-making. Beyond engineering, the model poses epistemic questions about when and why coherent reasoning can emerge from contradictory substrates. We hope this serves as both a practical starting point and a conceptual foundation for future studies of selectively coherent reasoning in complex environments.

REFERENCES

[1] N. Sharmin, S. Roy, A. Laszka, J. Acosta, and C. Kiekintveld, "Bayesian Models for Node-Based Inference Techniques," in *2023 IEEE International Systems Conference (SysCon)*, 2023, pp. 1–8.

[2] T. Li, *A General Space of Belief Updates for Model Misspecification in Bayesian Networks*, arXiv:2311.05248, 2023.

[3] M. Chrisman, *Belief, Agency, and Knowledge: Essays on Epistemic Normativity*. Oxford University Press, Jun. 2022. [Online]. Available: https://doi.org/10.1093/oso/9780192898852.001.0001.

[4] S. Hansson, "Iterated AGM Revision Based on Probability Revision," *J of Log Lang and Inf*, vol. 32, pp. 657–675, 2023.

[5] F. Harary, "On the notion of balance of a signed graph.," *Michigan Mathematical Journal*, vol. 2, no. 2, pp. 143–146, 1953. [Online]. Available: https://doi.org/10.1307/mmj/1028989917.

[6] M. Vlăsceanu et al., "A Network Approach to Investigate the Dynamics of Individual and Collective Beliefs: Advances and Applications of the BENDING Model," *Perspectives on Psychological Science*, 2023.

[7] S. Dumbrava, "Theoretical Foundations for Semantic Cognition in Artificial Intelligence," *arXiv preprint arXiv:2504.21218*, 2025.

[8] S. Nikooroo and T. Engel, *Toward a Graph-Theoretic Model of Belief: Confidence, Credibility, and Structural Coherence*, arXiv:2508.03465, 2025.

[9] S. Nikooroo and T. Engel, *Reasoning Systems as Structured Processes: Foundations, Failures, and Formal Criteria*, arXiv preprint, arXiv:2508.01763, 2025.

[10] A. Hunter, "Automated Reasoning with Epistemic Graphs Using SAT Solvers," in *Proceedings of the International Conference on Computational Models of Argument (COMMA)*, 2022.

[11] L. Kunitomo-Jacquin and K. Fukuda, "Towards reasoning over knowledge graphs under aleatoric and epistemic uncertainty," in *2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, 2023, pp. 294–295.

[12] C. S. Wright, "Bayesian Epistemology with Weighted Authority: A Formal Architecture for Truth-Promoting Autonomous Scientific Reasoning," *arXiv:2506.16015*, 2025.

[13] A. Hunter and S. Tadey, "Trust Graphs for Belief Revision: Framework and Implementation," CEUR Workshop Proceedings, vol. 3197, O. Arieli, G. Casini, and L. Giordano, Eds., 2022.

[14] J. Jiang and P. Naumov, "A logic of trust-based beliefs," *Synthese*, 2024.

[15] X. Chen, "Statemental credibility logic for intelligent systems," 2022.

[16] A. Hunter and S. Tadey, "A System for Updating Trust and Performing Belief Revision," in *International Conference on Agents and Artificial Intelligence*, 2023.

[17] D. P. Prieto et al., "A Belief Model for Conflicting and Uncertain Evidence - Connecting Dempster-Shafer Theory and the Topology of Evidence," in *International Conference on Principles of Knowledge Representation and Reasoning*, 2024.

[18] P. P. Shenoy, "On distinct belief functions in the Dempster-Shafer theory," in *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, 2023.

[19] J. Delgrande et al., "Epistemic Logic of Likelihood and Belief," in *International Joint Conference on Artificial Intelligence*, 2023.

[20] A. Toghri and S. Sanner, "Bayesian Inference with Complex Knowledge Graph Evidence," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[21] H. Monte-Alto et al., "Argumentation-based multi-agent distributed reasoning in dynamic and open environments," *Knowledge and Information Systems*, 2024.

[22] T. Aravanis, "Tailoring disjoint belief structures to the AGM framework," *Journal of Logic and Computation*, 2024.

[23] M. Aliviano et al., "Weighted knowledge bases with typicality and defeasible reasoning in a gradual argumentation semantics," *Intelligenza Artificiale*, 2024.

[24] H. Freedman et al., "A Bayesian Approach to Constructing Probabilistic Models from Knowledge Graphs," *International Journal of Semantic Computing*, 2023.

[25] D. Bhattacharjee and A. Anand, *From Argumentative Text to Argument Knowledge Graph: A New Framework for Structured Argumentation*, 2025.

[26] V. Sarathy et al., "A Neuro-Symbolic Cognitive System for Intuitive Argumentation," in *2022 Tenth Annual Conference on Advances in Cognitive Systems*, 2022.

[27] T. Aravanis, "Incorporating Belief Merging into Relevance-Sensitive Belief Structures," in *Panhellenic Conference on Informatics*, 2022.

[28] H. Olsson and M. Galesic, "Analogies for modeling belief dynamics," *Trends in Cognitive Sciences*, 2024.

[29] P. Hase et al., "Methods for Measuring, Updating, and Visualizing Factual Beliefs in Language Models," in *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2023.

[30] Y. Huang et al., "Belief-Based Preference Structure and Elicitation in the Graph Model for Conflict Resolution," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, 2023.

[31] E. Camina, F. Guell, J. Sepulcre, and J. Bernacer, "Hubs of Belief Networks Across Sociodemographic and Ideological Groups," vol. 12, 2022.

[32] F. Zimmaro, "A meta-model of belief dynamics with Personal, Expressed and Social beliefs," *arXiv:2502.14362*, 2025.

[33] A. S. Bizyaeva et al., "Multi-topic belief formation through bifurcations over signed social networks," *IEEE Transactions on Automatic Control*, vol. 70, 2025.

[34] J. Hewson, "Evaluating internal and external dissonance of belief dynamics in social systems," *arXiv:2410.07240*, 2024.

[35] J. Dalaege et al., "Networks of beliefs: An integrative theory of individual- and social-level belief dynamics," *Psychological Review*, 2024.

[36] T. Hofweber et al., "Are language models rational? The case of coherence norms and belief revision," *arXiv preprint arXiv:2406.03442*, 2024.

[37] B. Wang et al., "Tractable Uncertainty for Structure Learning," in *International Conference on Machine Learning*, 2022.

[38] S. Dumbrava, *Belief Filtering for Epistemic Control in Linguistic State Space*, arXiv preprint arXiv:2505.04927, 2025.

[39] M. Souza and R. Wassermann, "Hyperintensional Models and Belief Change," in *BRACIS*, 2022.

[40] M. Jaeger, "Learning and reasoning with graph data," *Frontiers in Artificial Intelligence*, 2023.