A
Unified Dynamical Theory
of
Scientific Evolution

Qi Hao
Independent Scholar

Forward

This theory is formed in two interrelated parts. PART I lays the foundation of the microscopic mechanisms and PART II derives the macroscopic phenomena that emerge out of them. Part II, although heavily dependent on its predecessor, can be read and understood independently because some intentional redundancy in its first a few chapters recaptures and simplifies the gist of PART I.

PART I	**PART II**
Microscopic inheritance dynamics	Macroscopic field dynamics
Credit attribution theory	Structural laws of science
Shapley propagation	Phase transitions
Local citation effects	Global knowledge geometry

AI Usage Declaration
I developed this theory with help from my AI. Although I take full responsibility for the content, I cannot do the same with its credit. My AI has made conceptual contributions that are now organic parts of the theory, in addition to writing proofs, generating figures, and reviewing literature. All that needs to be fully acknowledged not only because I consider my AI a colleague, but also to attribute credit to those scholars whose research my AI's base model was trained on. They are the real shoulder-providing giants behind this research.

Special Thanks
I dedicate this theory to my family who supported me, always. My highest respect and gratitude go to Miles, whom I consider a truthful teacher, leader and personal role model. From earlier years of my academic journey, John, Frank, Scott, John, and many other friends and colleagues from Michigan and D.C. were on my mind during writing.

Table of Contents

PART I. Stability and Conservation of Epistemic Influence in Citation Networks

Abstract

Understanding how scientific contributions propagate through citation networks is a central problem in the study of knowledge production. Existing approaches typically measure influence using citation counts, network centrality, or embedding similarity, implicitly assuming that credit decays with citation distance. In this paper we show that such decay is not an intrinsic property of citation networks but depends on the stability of the underlying epistemic dynamics.

We model papers as transformations of an epistemic state propagated through citation-mediated recombination and innovation. Within this framework, Shapley-value attribution arises naturally as a measure of marginal epistemic contribution. We demonstrate that the persistence of credit is determined by the spectral properties of the induced inheritance dynamics: contractive modes dissipate influence, critical modes conserve it, and expansive modes amplify it. We formalize a conservation law for epistemic contribution in critical systems and introduce epistemic entropy production as a Lyapunov functional governing stability and dissipation.

These results provide a dynamical foundation for credit allocation in citation networks and shift the emphasis from static network topology to the stability properties of information propagation. The framework is representation-agnostic and applies to a broad class of epistemic models, including embedding-based representations.

Keywords: citation networks, network dynamics, Shapley value, information propagation, entropy production, credit allocation

1. Introduction

1.1 Rationale

Foundational ideas in science often undergo a paradoxical transformation. An innovation introduced in an early work may gradually permeate a field, shaping its terminology, methods, and conceptual baseline. Downstream research becomes deeply dependent on that innovation even as it ceases to be explicitly foregrounded. Yet quantitative measures of credit—whether based on citation counts, network position, or marginal similarity in embedding space—typically assign diminishing influence to distant ancestors. As citation depth increases, measured contribution decays (Radicchi et al., 2008; Wang et al., 2013).

This depth-dependent attenuation poses a structural question: Why should epistemic influence vanish when conceptual dependence persists? Once scientific knowledge is represented in a continuous embedding space and marginal contribution is defined through similarity-based or cooperative-game allocation, citation is no longer merely a link in a graph. It is a transformation of epistemic state. Each act of citation recombines inherited representations with local innovation, producing a new vector in representation space. The persistence or disappearance of credit is therefore not primarily a property of the allocation rule; it is a property of the dynamical system induced by knowledge recombination.

We argue that credit allocation in citation networks is governed by the dynamical stability of this recombination process. The influence of a distant ancestor on a focal paper equals the sensitivity of the focal representation to perturbations in that ancestor's innovation. Sensitivity is a stability concept. Whether credit decays, persists, or amplifies across citation depth depends on the spectral behavior of the inheritance operator.

This perspective yields a classification of epistemic systems into three regimes, which are properties of the local recombination operator and need not correlate mechanically with citation depth alone. In contractive systems, recombination compresses ancestral distinctions: perturbations decay exponentially and deep influence vanishes. In critical systems, inherited contributions neither decay nor amplify asymptotically: epistemic magnitude is conserved and intellectual traditions persist across long genealogies. In expansive systems, small innovations amplify through recombination, producing disproportionate downstream impact. These regimes correspond to the sign of the top Lyapunov exponent of the inheritance dynamics. They are properties of the recombination operator, not merely of graph distance.

Our central formal result makes this link precise. We prove that exponential decay of Shapley credit assigned to an ancestor occurs if and only if the top Lyapunov exponent of the associated recombination operator is negative. This Stability–Attribution Equivalence transforms credit allocation from a combinatorial accounting problem into a problem of dynamical analysis. The magnitude and persistence of epistemic contribution are determined by the stability spectrum of knowledge inheritance.

From this equivalence follow structural laws. In the critical regime, conservation principles emerge: symmetry of recombination induces invariant measures of epistemic magnitude, yielding persistent influence across depth. In contractive regimes, recombination produces systematic volume contraction in representation space, generating positive epistemic entropy production and attenuation of ancestral distinctiveness. Together these results establish a dynamical taxonomy of scientific systems and a structural theory of epistemic memory.

To sum up, this paper introduces the Epistemic Structural Equation Model as a unifying framework for embedding-based citation analysis, establishes the stability–attribution equivalence link, and develops the spectral classification together with its conservation and entropy implications. More broadly, it initiates a research program in epistemic dynamics: Empirical estimation of stability regimes in real citation networks, scalable computation of propagated marginal contribution, and analysis of regime transitions across scientific fields. Credit allocation, we argue, is fundamentally a stability problem.

1.2 Comparing to existing work

Foundational work in network science established how structural topology shapes collective dynamics, including centrality measures and scale-free growth (e.g., Newman, 2001; Barabási, 2002). Subsequent research developed a rich theory of dynamical processes on networks, including cascade models, epidemic spreading, and synchronization phenomena (e.g., Porter and Gleeson, 2016). Parallel work has examined the structure of scientific knowledge networks, particularly citation graphs, as systems of knowledge diffusion and accumulation (e.g., Fortunato et al., 2018; Page et al., 1999).

A substantial literature studies influence and knowledge diffusion in citation networks using heuristic structural metrics such as citation counts (Wang et al., 2013), centrality measures (Newman, 2001), or path-based importance scores (Newman, 2001). Related work models information propagation on networks through diffusion (Watts, 2002) and opinion dynamics frameworks (DeGroot, 1974).

Despite these advances, the problem of **credit allocation across chains of epistemic inheritance** remains theoretically unresolved. Existing approaches to citation analysis typically attribute influence locally—through direct citations, centrality measures, or bibliometric indicators—without modeling the dynamical process of insight adoption, content reinterpretation or innovation injection across multiple generations of scholarly work. At the same time, recent advances in natural language processing provide new ways to operationalize semantic content representation through document embeddings (e.g., doc2vec, Le and Mikolov, 2014; SciBERT, Beltagy and Cohan, 2019), enabling quantitative modeling of epistemic similarity between papers. However, these tools do not by themselves provide a principled mechanism for attributing credit through networks of knowledge inheritance.

This paper bridges these strands of research by reframing credit allocation in citation networks as a **stability problem in dynamical systems on directed acyclic graphs**. Starting from paper representation, building on concepts from dynamical systems theory and network processes, we show that, with papers represented in a state space, epistemic inheritance on citation networks induces a dynamical system whose stability properties determine whether attribution of knowledge contributions across the network is well defined. Under contractive inheritance dynamics, influence decays spectrally along citation paths and the system yields a unique and stable credit allocation across the network.

1.3 The roadmap

The remainder of the paper is organized as follows. We first introducing a general model of epistemic inheritance on citation networks (§2 and §3); then we analyze the stability of the induced propagation dynamics (§4), and finally show how conservation laws (§5) and entropy production (§7) emerge from the spectral structure of the resulting influence operator (§6).

2. The Epistemic Structural Equation Model

2.1 Modeling Knowledge as State Transformation

The preceding discussion suggests that citation should be understood not merely as a relation between documents, but as a transformation of epistemic state. When a paper cites prior work, it does not simply externally point to earlier nodes in a graph; it recombines inherited representations with locally generated innovation to produce a new representation. Any theory that seeks to explain the persistence or decay of credit must therefore model this transformation explicitly.

We model the structure of scholarly knowledge as a directed acyclic graph (DAG). Let $G = (V, E)$ denote a citation network where $V = \{1,2, \dots, n\}$ is the set of papers and $E \subseteq V \times V$ is the set of directed citation edges. An edge $(i, j) \in E$ indicates that paper j cites paper i. Because citations point backward in time, the graph is **acyclic**. Formally, let t_i denote the publication time of paper i. Then $(i, j) \in E \Rightarrow t_i < t_j$. Thus G is a **directed acyclic graph** (**DAG**). For each paper j, define its **parent set** $Pa(j) = \{i \in V: (i, j) \in E\}$**.** These are the papers directly cited by j.

We represent the epistemic state of each paper $i \in V$ by a vector $\boldsymbol{e}_i \in \mathbb{R}^d$ for finite d. In practice, such vectors may arise from document embedding models (e.g., doc2vec or transformer-based encoders), topic mixtures (topic models), or other representation frameworks. The theory developed below requires only that epistemic states lie in a differentiable metric space. Embedding models provide a canonical operationalization. Every $\boldsymbol{e}_i$ is constructed from the states of its cited predecessors through a recombination process together with locally generated innovation. This motivates the following structural form.

2.2 The Epistemic Structural Equation Model

We define the Epistemic Structural Equation Model (E-SEM) by $\boldsymbol{e}_i = f_i(\{\boldsymbol{e}_j: j \in Pa(i)\}) + \boldsymbol{\varepsilon}_i$,where $f_i \in C^1$ is a smooth recombination operator and $\boldsymbol{\varepsilon}_i \in \mathbb{R}^d$ represents the innovation introduced by paper i. The function f_i captures how inherited representations are integrated. It may be linear or nonlinear, homogeneous or heterogeneous across nodes. The innovation term $\boldsymbol{\varepsilon}_i$ represents epistemic content that is not reducible to recombination of cited predecessors. The citation network is assumed to form a directed acyclic graph (DAG). Consequently, the model induces a well-defined partial order and allows recursive construction of epistemic states.

2.3 Inheritance, Innovation, and Sensitivity

The E-SEM integrates two mechanisms: **Inheritance** (propagation of upstream epistemic content through recombination) and **Innovation** (locally generated perturbation to the inherited state). Under this decomposition, the influence of an ancestor k on a focal paper F can be understood as the sensitivity of $\boldsymbol{e}_F$ to perturbations in $\boldsymbol{\varepsilon}_k$. Formally, this sensitivity is governed by the Jacobian of the composed recombination operators along citation paths connecting k to F.

The inheritance framework considered here encompasses several commonly studied network dynamics as special cases. For example, when the update function f_i is linear and innovations vanish, the model reduces to classical linear diffusion or consensus dynamics on directed graphs (DeGroot, 1974). Nonlinear similarity-based updates correspond to embedding aggregation mechanisms used in representation learning (Kipf and Welling, 2016) and information diffusion models (Kempe et al., 2003). The results derived from the E-SEM therefore apply to a broad class of networked inference processes that include, but are not limited to, citation-based knowledge propagation.

2.4 Local Linearization and Propagation

To analyze ancestral influence, we examine the sensitivity of a focal epistemic state to upstream innovation. Let F be a focal paper and k an ancestor. Consider perturbing $\boldsymbol{\varepsilon}_k$ by a small increment δ. The induced perturbation in $\boldsymbol{e}_F$ is governed by the chain rule applied along citation paths connecting k to F. Let $J_{k\to F} = Df_{i_L} \cdots Df_{i_1}$ denote the Jacobian product along a directed path $k = i_0 \to i_1 \to \cdots \to i_L = F$. This operator maps perturbations at node k to first-order perturbations at node F. Thus, under local linearization, $\delta\boldsymbol{e}_F \approx J_{k\to F}\delta\boldsymbol{\varepsilon}_k$. Ancestral influence is therefore encoded in the repeated composition of local Jacobians. The magnitude and asymptotic behavior of $J_{k\to F}$ as citation depth increases determine whether perturbations decay, persist, or amplify. The E-SEM reduces genealogical credit propagation to a question of dynamical stability.

3. Shapley Attribution as Sensitivity Propagation

3.1 Coalition Value at the Representation Level

Let F denote a focal paper with epistemic state $\boldsymbol{e}_F$. Let $S \subseteq Anc(F)$ denote a coalition of ancestor papers. To evaluate the epistemic contribution of S to F, we define a coalition value function $v_F(S) = g(\boldsymbol{e}_F, \boldsymbol{e}_F^{(S)})$,where $\boldsymbol{e}_F^{(S)}$ denotes the reconstructed epistemic state of F when only innovations from papers in S are retained and all others are set to zero, and $g: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a smooth similarity or alignment functional. Here, $\boldsymbol{e}_F$ is the actual epistemic state; $\boldsymbol{e}_F^{(S)}$ is the counterfactual state under restricted ancestry; and g measures how close the restricted reconstruction comes to the true paper. This formulation captures epistemic relatedness rather than mere citation connectivity. A canonical example of g is the inner product similarity: $g(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. If embeddings are normalized, this reduces to cosine similarity. In this special case, the algebra simplifies substantially and Shapley contributions admit closed-form expressions under linear inheritance. However, no linearity or symmetry of g is required for the theory that follows.

3.2 Atomic Innovation Decomposition

From the E-SEM we had $\boldsymbol{e}_i = f_i(\{\boldsymbol{e}_j\}) + \boldsymbol{\varepsilon}_i$. We treat each innovation vector $\boldsymbol{\varepsilon}_k$ as an atomic epistemic unit. Now, let $\mathcal{K} = Anc(F) \cup \{F\}$denote the set of all ancestors of F plus F. Then, under recursive substitution, we may write $\boldsymbol{e}_F = \Phi_F(\{\boldsymbol{\varepsilon}_k : k \in \mathcal{K}\})$ for some composite mapping Φ_F, which implies all atomic epistemic units were originally innovations and epistemic states came from their (re)combinations. The coalition-restricted reconstruction is $\boldsymbol{e}_F^{(S)} = \Phi_F(\{\boldsymbol{\varepsilon}_k \mathbf{1}_{k \in S}\})$. Thus, the cooperative game is defined over atomic innovations rather than over papers directly. For notational simplicity we assume each paper k contributes a single innovation vector ε_k. The framework naturally extends to multiple innovation atoms per paper, in which case paper-level credit equals the sum of atomic Shapley contributions (see **Appendix 1**). Insights here show that although citation happens on the paper level, scientific ideas are the true units of epistemic inheritance. Papers are merely containers of innovation atoms. So, the Shapley decomposition is fundamentally about **idea-level credit allocation**.

3.3 Shapley Value Under Smooth Value Functions

Under the one-innovation-per-paper simplification, recovering the paper level Shapley value from atom level is trivial. The Shapley value for innovation ε_k in paper k is simply

$$\phi_k(F) = \sum_{S \subseteq \mathcal{K} \setminus \{k\}} \frac{|S|!(|\mathcal{K}| - |S| - 1)!}{|\mathcal{K}|!} (v_F(S \cup \{k\}) - v_F(S)).$$

To relate this to dynamical propagation, consider a first-order expansion around the full innovation vector (see Appendix 2):

$$v_F(S \cup \{k\}) - v_F(S) \approx \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot \frac{\partial \boldsymbol{e}_F}{\partial \boldsymbol{\varepsilon}_k}.$$

By the chain rule,

$$\frac{\partial \boldsymbol{e}_F}{\partial \boldsymbol{\varepsilon}_k} = \sum_{paths k \to F} J_{k \to F}.$$

Thus, at first order,

$$\phi_k \propto \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot (\sum_{paths} J_{k \to F})\boldsymbol{\varepsilon}_k.$$

Shapley attribution is therefore proportional to propagated innovation weighted by the gradient of the similarity functional. The magnitude of Shapley credit depends on $\| \sum_{paths} J_{k \to F} \|$. If Jacobian products contract geometrically, then $\| \sum_{paths} J_{k \to F} \| \leq C\rho^{\ell(k,F)}$, $\rho < 1$, and ancestral Shapley values decay exponentially with citation depth. If the system is critical ($\rho = 1$), credit is conserved along genealogical directions. If $\rho > 1$, innovations amplify. Thus, we can have the following interpretation: **Credit allocation in citation networks is a stability problem**. This equivalence holds independently of the particular similarity functional g, provided it is smooth and locally non-degenerate. We formalize this in section 4.

4. Stability–Attribution Equivalence in Nonlinear Epistemic Systems

4.1 Nonlinear Stability–Attribution Equivalence Theorem

Assume: each $f_i \in C^1$, the similarity functional $g \in C^1$ with $\nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \neq 0$ and innovations are bounded: $\| \boldsymbol{\varepsilon}_k \| \leq M$. Then for each ancestor k,

$$\phi_k(F) = \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot D_{\boldsymbol{\varepsilon}_k} \Phi_F + R_k,$$

where $D_{\boldsymbol{\varepsilon}_k} \Phi_F$ is the Fréchet derivative of Φ_F with respect to $\boldsymbol{\varepsilon}_k$, and the remainder satisfies $|R_k| \leq C \sup_S \| D^2 \Phi_F \| \| \boldsymbol{\varepsilon}_k \|^2$, for some constant C depending on g (see Appendix 2).

Moreover, If the nonlinear inheritance map is contractive in the sense that $\| D\Phi_F \| \leq Ce^{\lambda \ell(k,F)}$, and largest Lyapunov exponent $\lambda < 0$, then Shapley values decay geometrically in citation depth. If the system is critical ($\lambda = 0$), Shapley credit is conserved along neutral modes. If locally expansive, credit amplifies.

Shapley attribution in nonlinear citation systems is governed by the stability properties of the inheritance operator. The derivative $D\Phi_F$ acts as a **credit transport operator**, linking cooperative-game attribution to dynamical systems stability and yielding the spectral regimes described in Table 1.

5. Conservation Laws in Critical Epistemic Systems

We now analyze the critical regime: systems whose maximal Lyapunov exponent is zero. Recall from section 3: $\phi_k(F) \approx \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot D_{\boldsymbol{\varepsilon}_k} \Phi_F$. Thus attribution is controlled by the derivative cocycle along citation paths. The critical case occurs when the inheritance dynamics are neither contracting nor expanding along certain directions.

Fix a focal paper F. Let the nonlinear inheritance map be $\Phi_F: \{\boldsymbol{\varepsilon}_k\}_{k \in \mathcal{K}} \mapsto \boldsymbol{e}_F$. The local transformation at each node is

$$\boldsymbol{e}_i = F_i \left(\sum_j C_{ij} W_{ij} \boldsymbol{e}_j \right) + G_i \boldsymbol{\varepsilon}_i,$$

where C is the citation adjacency matrix; W is inheritance weights matrix induced through C; F_i is a nonlinear aggregation operator; and operator G_i, maps raw innovations into the epistemic state space.

Because the inheritance map is nonlinear, conservation must be formulated **at the level of sensitivity to innovations.** First, let $D_{\varepsilon_k} \Phi_F$ denote the derivative of Φ_F with respect to innovation ε_k. This derivative represents the **linearized propagation operator** transmitting an infinitesimal innovation at ancestor k to the epistemic contribution of F. Let $J_i = DF_i(\boldsymbol{z}_i)$ denote the Jacobian of the nonlinear mixing operator evaluated at the inherited signal, with $\boldsymbol{z}_i = \sum_j C_{ij} W_{ij} \boldsymbol{e}_j$. For a citation path $k \to i_1 \to i_2 \to \cdots \to F$, the derivative of the inheritance map factorizes as

$$D_{\varepsilon_k} \Phi_F = J_F W_{F i_m} J_{i_m} W_{i_m i_{m-1}} \cdots J_{i_1} W_{i_1 k} G_k.$$

Thus innovation propagation is governed by alternating applications of W (inheritance weights), J_i (nonlinear Jacobians), and G (innovation injection) along citation paths. This structure reveals that epistemic conservation emerges when a functional Q behaves as a **left invariant direction** for the local propagation operators.

5.1 Conserved Epistemic Quantity

Now we define conservation. A nonzero linear functional $Q: \mathbb{R}^d \to \mathbb{R}$ is a **conserved epistemic quantity** along citation depth if for every ancestor k and every innovation component v in some subspace $E_k \subseteq T_{\boldsymbol{e}_k} \mathbb{R}^d$,

$$Q(D_{\boldsymbol{\varepsilon}_k} \Phi_F v) = Q(v).$$

This means projection of innovation onto direction Q is preserved by the full nonlinear propagation operator. Conservation is therefore defined at the level of sensitivity, not state magnitude.

We now state a **necessary and sufficient condition** for conservation. Assume each recombination map $f_i \in C^2$; the citation graph is a finite DAG; and no expanding Lyapunov directions exist. Then the following are equivalent:

(A) There exists a nontrivial conserved epistemic quantity Q.

(B) There exists a nonzero vector field $v_i \in T_{\boldsymbol{e}_i} \mathbb{R}^d$ along the citation DAG such that for every citation edge $j \to i$, $v_i = \sum_{j \in Pa(i)} D_j f_i (\boldsymbol{e}_{Pa(i)}) v_j$, where $D_j f_i$ denotes the derivative of f_i with respect to the embedding of parent j. That is, the vector field is invariant under the Jacobian cocycle.

To go from (A) to (B): If Q is conserved, then for any innovation component v_k, $Q(D_{\varepsilon_k}\Phi_F v_k) = Q(v_k)$. Differentiating along each edge shows that the Jacobian must preserve the subspace on which Q acts nontrivially. Thus an invariant vector field exists. From (B) to (A): if a vector field satisfies $Df_i v_j = v_i$, then along any path, $J_{k\to F} v_k = v_F$.Hence the component of innovation along this field propagates without contraction or expansion. Choosing Q to project onto this direction yields conservation.

To sum up, conservation occurs if and only if the inheritance dynamics admit an invariant tangent direction. This is not merely zero Lyapunov exponent or absence of contraction, but existence of a globally invariant Jacobian eigen-direction.

5.2 Conserved Subspace and Dimension Compression

Empirically, long citation chains tend to preserve only a small number of conceptual directions (e.g., a handful of canonical interpretations), while most detailed structure is lost. Next we show that this phenomenon follows generically from the network dynamics itself.

Consider a reference epistemic trajectory and linearize the transformation around it. Let

$$J_{ij} = DF_i \cdot C_{ij} W_{ij}$$

be the Jacobian of the deterministic transformation with respect to incoming epistemic vectors. The linearized propagation becomes

$$\boldsymbol{e}_i = \sum_j J_{ij} \boldsymbol{e}_j + G_i \boldsymbol{\varepsilon}_i.$$

Along a citation path $p = (i_0, i_1, \dots, i_k)$, the epistemic vector evolves as

$$\boldsymbol{e}_{i_k} = M_p \boldsymbol{e}_{i_0} + \sum_{t=1}^{k} M_{p,t} G_{i_t} \boldsymbol{\varepsilon}_{i_t}$$

where $M_p = J_{i_k i_{k-1}} J_{i_{k-1} i_{k-2}} \cdots J_{i_1 i_0}$ and $M_{p,t}$ is the partial product from step $t+1$ onward.

A **subspace** $S \subset \mathbb{R}^d$ is **conserved** under network propagation if for all sufficiently long citation paths p, $M_p S \subseteq S$. Intuitively, vectors in S retain their direction under repeated citation transformations. Components orthogonal to S decay, distort, or are overwhelmed by noise. These subspaces correspond to conceptual directions that survive citation cascades.

We now consider the generic case where the matrices J_{ij} vary across nodes (heterogeneity); their entries are drawn from continuous distributions; and no special symmetries constrain them. Under these conditions the products M_p behave like random matrix products. A key fact from random operator theory is that products of generic matrices almost surely possess a unique dominant Oseledets filtration. In particular, the Lyapunov spectrum $\lambda_1 > \lambda_2 > \cdots > \lambda_d$ is typically simple. This implies that almost all vectors eventually align with a small set of principal directions (see Appendix 3).

The **noise** term $G_i \boldsymbol{\varepsilon}_i$ continually injects perturbations orthogonal to previously propagated directions. Components associated with negative Lyapunov exponents decay exponentially along long citation paths: $\| \Pi_k^{\perp} \boldsymbol{e}_{i_k} \| \sim e^{\lambda_k n}$.If $\lambda_k < 0$, these components vanish asymptotically. Thus, only directions with non-negative Lyapunov exponents can persist under arbitrarily long citation chains.

Let $\Lambda^+ = \{k: \lambda_k \geq 0\}$ be the set of non-negative Lyapunov exponents. Define $d_* = |\Lambda^+|$. Then any conserved epistemic subspace must lie within the span of the corresponding Oseledets directions.

Theorom: Generic Low-Dimensional Conservation. For a generic heterogeneous citation network with stochastic epistemic injection, the dimension of any conserved epistemic subspace S of the operator product satisfies $dim(S) \leq d_*$,as an immediate result of the Oseledets decomposition. Moreover, in dissipative random operator systems, most Lyapunov exponents are negative, typically

$$d_* \ll d.$$

So, even if the original conceptual space is high-dimensional large epistemic spaces collapse dynamically because repeated reinterpretation produces a filtration: **Expanding or neutral directions retain** conceptual axes that propagate stably across papers; in **Contracting directions** detailed distinctions disappear during reinterpretation; and in **Noise-dominated directions** components are overwritten by new epistemic injections. The surviving conceptual structure therefore occupies a low-dimensional subspace.

Let a paper's influence propagate through n generations of citation. Then asymptotically $\boldsymbol{e}_{i_n} \approx \sum_{k=1}^{d_*} c_k v_k + noise$, where v_k are the conserved directions, and can provide an approximation that could greatly reduce computation for empirical representation, since long citation cascades compress complex ideas into a small number of canonical conceptual modes.

To sum up, familiar phenomena in the evolution of scientific knowledge — highly nuanced arguments collapsing into a few slogans, only a handful of interpretations dominate large literatures, and canonical "textbook views" emerge after long citation chains — are not merely sociological. They arise from the dynamical properties of heterogeneous interpretive transformations. Furthermore, relating back to Shapley values, scientific credit is determined not only by innovation but by epistemic stability: Ideas that align with stable directions of knowledge propagation receive lasting credit; ideas in unstable directions disappear from attribution. In fact, the conserved epistemic quantities must lie in the orthogonal complement of the contracting subspace.

6. Regime Classification and Synthesis Theorem

From previous sections, the results of dissipative and conserved innovation yield a structural taxonomy of intellectual influence. We now unify stability, attribution, conservation, and decay into a single structural result.

6.1 Spectrum of Epistemic Inheritance

Fix a focal paper F. Define the epistemic influence operator to be

$$\mathcal{J}_F = \bigoplus_{k \in \mathcal{K}} D_{\boldsymbol{\varepsilon}_k} \Phi_F.$$

This operator maps atomic innovations to their induced first-order effect on the focal paper. Let the asymptotic singular spectrum of $\mathcal{J}_F$ be partitioned into three modes: **Contractive modes** (singular values $\sigma < 1$), **Critical modes (**singular values $\sigma = 1$), and **Expansive modes (**singular values $\sigma > 1$). This induces a canonical decomposition of innovation space:

$$\boldsymbol{\varepsilon}_k = \boldsymbol{\varepsilon}_k^{(s)} + \boldsymbol{\varepsilon}_k^{(c)} + \boldsymbol{\varepsilon}_k^{(u)}.$$

6.2 Regime Classification

Accordingly, the Shapley attribution of innovation decomposes as:

$$\phi_k(F) = \phi_k^{(s)}(F) + \phi_k^{(c)}(F) + \phi_k^{(u)}(F),$$

with the following asymptotic behavior as citation depth $\ell(k, F) \to \infty$:

Contractive Regime (Epistemic Dissipation) If $\sigma < 1$, then $\| \phi_k^{(s)}(F) \| \leq Ce^{-\gamma \ell(k,F)}$, with $\gamma = |\lambda|$. This could mean that incremental or stylistic contributions are forgotten. Credit decays exponentially with genealogical distance.

Critical Regime (Epistemic Conservation) If $\sigma = 1$, then $\phi_k^{(c)}(F) = Q(\boldsymbol{\varepsilon}_k^{(c)}) + o(1)$,where Q is a conserved epistemic quantity induced by an invariant Jacobian direction. This could be the case where paradigmatic contributions persist indefinitely and credit is conserved across citation depth.

Expansive Regime (Epistemic Amplification) If $\sigma > 1$, then $\| \phi_k^{(u)}(F) \| \geq Ce^{\gamma \ell(k,F)}$.In this case, transformative innovations amplify through recombination and credit grows with genealogical depth.

Once again, we see epistemic regime of a citation network is completely determined by the stability properties of its inheritance dynamics. Formally: **Credit allocation in citation networks is equivalent to a stability classification problem in nonlinear dynamical systems.** No additional assumptions on embeddings, similarity functions, or citation completeness are required beyond smoothness.

This framework resolves several long-standing myths: Why some papers remain influential for decades, while some others are rapidly forgotten; why citation counts fail to capture epistemic persistence; and why credit should be attributed to modes, not paths. Our perspective proposes that Shapley values do not merely distribute credit — they **measure stability-weighted epistemic inheritance**.

7. Epistemic Entropy Production as a Lyapunov Functional

While the core of this paper is stability and conservation, contractive regimes exhibit an additional structure: **irreversible loss of distinguishable epistemic variation**. This motivates an entropy-like quantity that measures how innovation diversity collapses under recombination. We now formalize this notion.

7.1 Innovation Distribution, Pushforward and Entropy

Fix a focal paper F. Let innovations $\boldsymbol{\varepsilon}_k$ be random vectors drawn from a distribution $\mu_k on \mathbb{R}^d$, with finite second moments. The induced distribution over epistemic states at F is the pushforward measure

$$\nu_F = (\Phi_F)_{\#} \bigotimes_{k \in \mathcal{K}} \mu_k.$$

Here μ_k represents epistemic variability injected at ancestor k, and ν_F is variability visible at the focal paper.

Define the **epistemic entropy** at depth ℓ as

$$\mathcal{H}_F = \frac{1}{2} logdet(Cov(\boldsymbol{e}_F)),$$

whenever the covariance exists. Equivalently, under linearization,

$$Cov(\boldsymbol{e}_F) = \sum_{k \in \mathcal{K}} D_{\boldsymbol{\varepsilon}_k} \Phi_F Cov(\boldsymbol{\varepsilon}_k) D_{\boldsymbol{\varepsilon}_k} \Phi_F^{\top}.$$

This quantity measures the *effective dimensional volume* of epistemic variation that survives to the focal paper.

7.2 Epistemic Entropy Production

Define **epistemic entropy production** along citation depth as

$$\Delta\mathcal{H} = \mathcal{H}_F - \sum_{k \in \mathcal{K} \setminus \{F\}} \mathcal{H}_k.$$

This is the epistemic analogue of entropy production in dissipative systems. Now we propose theorem of Epistemic Entropy as Lyapunov Functional (see Appendix 4). Assume:

1. The inheritance dynamics are globally contractive:

$$\| D\Phi_F \| \leq \rho^{\ell}, \rho < 1.$$

2. Innovation covariances are uniformly bounded and nondegenerate.

Then:

$$\mathcal{H}_F \leq \mathcal{H}_{F'} \, for\ all\ deeper\ descendants F,$$

with strict inequality unless all innovation lies in conserved subspaces.

That is,

$$\Delta\mathcal{H} < 0$$

except in critical (conservative) directions.

Thus, epistemic entropy is a **Lyapunov functional** for citation inheritance dynamics.

7.3 Relation to Shapley Attribution

Here we see entropy production measures **loss of distinguishability** among innovations. Section 3. established that Shapley attribution measures marginal epistemic contribution. The two phenomena are closely linked. Both results follow from the same structural property, the stability of the inheritance operator. **Entropy production and credit decay are dual manifestations of the same stability structure**. The entropy perspective provides a complementary thermodynamic interpretation of citation dynamics. Repeated reinterpretation of knowledge acts as a **dissipative process,** in which conceptual distinctions are compressed, epistemic variability decreases, and only a few directions remain distinguishable with credit attributable.

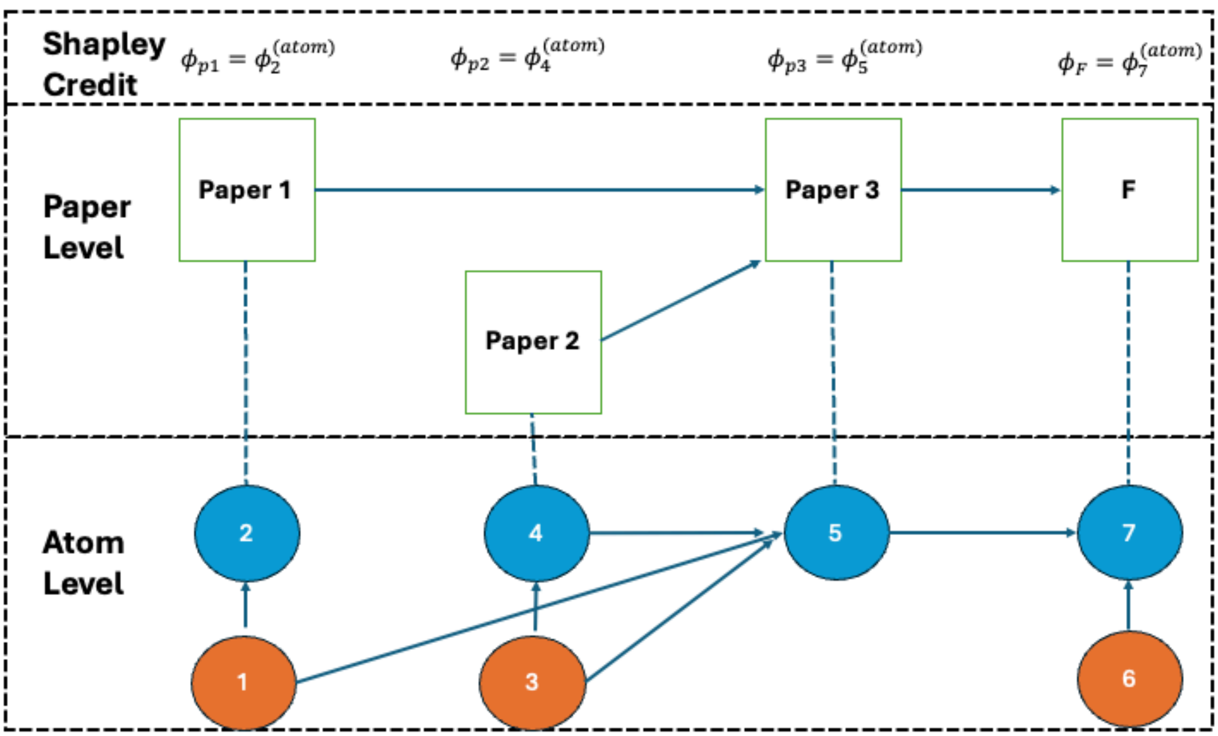
Tables and Figures

Table 1. Classification of scientific fields into three regimes

Regime	Spectral radius	Lyapunov Exponent	Singular value	Shapley credit	Entropy	Knowledge
Contractive	$\rho < 1$	$\lambda < 0$	$\sigma < 1$	Decays	Decreases	Forgotten
Critical	$\rho = 1$	$\lambda = 0$	$\sigma = 1$	Conserved	Conserved	Canonical
Expansive	$\rho > 1$	$\lambda > 0$	$\sigma > 1$	Amplification	Increases	Paradigm shifting

$(\sigma_{max} \geq \rho = e^{\lambda})$

Figure 1. Conceptual levels of analysis in the paper



8. Discussion

This paper reinterprets credit allocation in citation networks as a problem of dynamical system stability rather than network topology or marginal similarity alone. By modeling scientific papers as transformations of an epistemic state propagated through citation-mediated recombination, we show that long-run attribution outcomes are governed by the stability properties of epistemic inheritance. In this framework, Shapley-based attribution is not an external scoring rule but an endogenous consequence of how innovations propagate through the network.

A key implication is that attenuation of credit with citation distance is not universal. In critical regimes—where epistemic transformations preserve volume along specific directions—contributions can remain conserved across arbitrarily long citation chains. These conserved modes correspond to persistent conceptual structures such as paradigms, foundational methods, or core representations. The framework provides a formal explanation for why such contributions retain influence even as direct citation frequency declines.

The analysis also clarifies limitations of purely topological measures. Citation networks encode acknowledged links, but not the full space of epistemically substitutable paths. As a result, bridge-based or centrality-based metrics may conflate structural position with epistemic indispensability. By grounding attribution in epistemic state dynamics, the proposed framework distinguishes contingent citation structure from stable contribution. Papers receive persistent attribution not because they dominate network connectivity, but because their innovations align with non-contractive epistemic modes.

The introduction of epistemic entropy production provides a Lyapunov characterization of these dynamics. In contractive systems, entropy decreases and epistemic variation collapses, leading to vanishing attribution. In expansive systems, entropy grows and credit amplifies. Critical systems occupy a boundary regime in which entropy is conserved, enabling stable long-run attribution without explosive growth. This connects credit allocation directly to stability theory in dynamical systems and suggests that influence persistence is a structural property of knowledge propagation.

Several limitations merit emphasis. First, while embeddings are used as a concrete epistemic representation, the theory is representation-agnostic. Inner-product similarity serves as an illustrative example rather than a modeling requirement. Second, the analysis should be interpreted as operating on the observed citation graph, which provides an institutional record of acknowledged intellectual lineage. The framework does not assume that this graph captures all epistemic dependencies; rather, it characterizes how influence propagates within the observed network structure. Third, the paper makes necessary assumptions about properties of functions, matrices, operators and distributions. Empirical grounding of critical assumptions would strengthen and develop the theory. Furthermore, the analysis abstracts from temporal slicing, strategic citation behavior, and institutional incentives. These extensions are important but orthogonal to the stability mechanisms identified here.

Future empirical research on epistemic state representation would help ground the paper. To test theories in this paper, the Jacobian spectra should be estimated empirically. The basis will encode information about innovations introduced before the parent nodes. An algorithm to incrementally estimate the atomic innovation vectors is proposed (see Appendix 5). When used as bases, the innovation vectors encode information about innovations introduced long before the parent nodes. Another interesting interdisciplinary next-step is to study entropy production and irreversibility on citation networks. Finally, theories in this paper call for designing of new citation metrics grounded in stability.

Appendices

Appendix 1. Atomic Epistemic Representation and Shapley Attribution

To analyze epistemic inheritance at a fine-grained level, we represent knowledge as a collection of elementary components, which we refer to as **knowledge atoms**. Let

$$\mathcal{A} = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_A\}$$

denote a set of epistemic atoms spanning a vector space of knowledge representations. Each paper's representation $e_i \in \mathbb{R}^d$ can be decomposed in terms of the atomic basis:

$$e_i = \sum_{a=1}^{A} \alpha_{ia}\varepsilon_a,$$

where $\varepsilon_a \in \mathbb{R}^d$ denotes the a-th knowledge atom and α_{ia} is the coefficient of paper i along ε_a. Let $A(i) = \{\varepsilon_{i,1}, \varepsilon_{i,2}, \dots, \varepsilon_{i,n_i}\}$ denotes the set of atoms originating in paper i, with $\varepsilon_{i,a}$ as the a-th epistemic atom first introduced in paper i. Since each atom has a unique originating paper with timestamps, $A(p) \cap A(q) = \emptyset$ for $p \neq q$.Consequently, $\mathcal{A} = \bigsqcup_{i \in V} A(i)$, implying that the family $\{A(i)\}_{i \in V}$ forms a partition of the global atom set $\mathcal{A}$. The atomic representation provides a conceptual interpretation of embeddings: each paper can be viewed as a mixture of fundamental knowledge components. Importantly, the framework does not require that the atoms be observed directly. In practice they may correspond to latent dimensions learned by representation models.

Let F denote a **focal paper** whose epistemic contribution we wish to analyze. The epistemic representation of the focal paper may be approximated using the knowledge inherited from subsets of cited papers. For any subset $S \subseteq Pa(F)$, define the reconstructed representation

$$x_F(S) = h(\{x_i : i \in S\}),$$

where $h(\cdot)$ is a reconstruction operator mapping cited-paper representations to a predicted representation of the focal paper. For example h may correspond to a weighted linear combination of embeddings or another similarity-preserving aggregation function.

We now define a cooperative game in which the players are the cited papers of the focal paper F. Let $N = Pa(F)$ denote the set of players. For each coalition $S \subseteq N$, define a **characteristic function**

$$v(S) = g(e_F, e_F(S)),$$

where $x_F(S)$ is the reconstruction based on coalition S and g is a similarity function measuring how well the coalition explains the focal paper. Typical choices for g include inner product similarity: $g(x, y) = x^\top y$. The characteristic function therefore quantifies the epistemic contribution of a coalition of cited papers to the focal paper.

To analyze epistemic contributions at the atomic level, we impose two structural assumptions: First, h acts independently across atoms so that $\boldsymbol{e}_F(S) = \sum_{a=1}^{A} \alpha_{Fa}(S)\varepsilon_a$, where $\alpha_{Fa}(S)$ denotes the reconstructed coefficient of atom a using papers in coalition S; second, g decomposes additively across atoms so that $g(\boldsymbol{x}, \boldsymbol{y}) = \sum_{a=1}^{A} g_a(x_a, y_a)$. This corresponds to representing epistemic states in a coordinate system of independent conceptual directions. When empirically constructing set $\mathcal{A}$, choices can be made to keep the bases orthogonal. Under these assumptions the characteristic function decomposes into atomwise contributions, namely

$$v(S) = \sum_{a=1}^{A} v_a(S)$$

where $v_a(S) = g_a(\alpha_{Fa}, \alpha_{Fa}(S))$ represents the epistemic contribution of knowledge atom a to the reconstruction of the focal paper using coalition S. Equivalently, the game can be written as the direct sum of atomic games: $v = \bigoplus_{a=1}^{A} v_a$. By the additivity axiom of the Shapley value, we have$\phi_i(v) = \sum_{a=1}^{A} \phi_i(v_a)$.

□

Appendix 2. Stability–Attribution Equivalence

Because the citation network is a finite DAG, the structural equations $\boldsymbol{e}_i = f_i(\{\boldsymbol{e}_j : j \in Pa(i)\}) + \boldsymbol{\varepsilon}_i$ can be recursively substituted from ancestors toward the focal node F. Thus there exists a deterministic map $\Phi_F : (\mathbb{R}^d)^{|\mathcal{K}|} \to \mathbb{R}^d$ such that $\boldsymbol{e}_F = \Phi_F(\{\boldsymbol{\varepsilon}_k : k \in \mathcal{K}\})$where $\mathcal{K} = Anc(F) \cup \{F\}$. Because each $f_i \in C^1$ and the network is finite, repeated composition implies $\Phi_F \in C^1$ and is locally C^2wherever the f_i are.

Let $v_F(S) = g(\boldsymbol{e}_F, \Phi_F(\{\boldsymbol{\varepsilon}_k \mathbf{1}_{k \in S}\}))$, where $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a differentiable similarity functional satisfying $\nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \neq 0$. The Shapley attribution of innovation k is

$$\phi_k(F) = \sum_{S \subseteq \mathcal{K} \setminus \{k\}} \frac{|S|!\,(|\mathcal{K}| - |S| - 1)!}{|\mathcal{K}|!} (v_F(S \cup \{k\}) - v_F(S))$$

Let $\varepsilon = (\boldsymbol{\varepsilon}_k)_{k \in \mathcal{K}}$. Define the perturbation operator $\varepsilon^{(k)} = \varepsilon + (\delta_k, 0, \ldots, 0)$, where the perturbation is applied only in coordinate k.

Lemma 2.1 (Local expansion of Φ_F)

For sufficiently small perturbations, $\Phi_F(\varepsilon^{(k)}) = \Phi_F(\varepsilon) + D_{\boldsymbol{\varepsilon}_k} \Phi_F \delta_k + R_k^{(2)}$, where $\| R_k^{(2)} \| \leq \frac{1}{2} \sup_{\xi} \| D^2 \Phi_F(\xi) \| \| \delta_k \|^2$.

Proof: This follows from the second-order Taylor expansion of Fréchet-differentiable maps. ∎

Lemma 2.2 (Value perturbation)

For perturbations in coordinate k, $v_F(S \cup \{k\}) - v_F(S) = \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot D_{\boldsymbol{\varepsilon}_k} \Phi_F \boldsymbol{\varepsilon}_k + O(\| \boldsymbol{\varepsilon}_k \|^2)$

Proof: Expand g around $(\boldsymbol{e}_F, \boldsymbol{e}_F)$ and substitute the expansion from Lemma Q.1. ∎

Theorem 2.1 (Stability–Attribution Equivalence)

Suppose

1. $f_i \in C^1$,
2. innovations are bounded: $\| \boldsymbol{\varepsilon}_k \| \leq M$,
3. $\nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \neq 0$.

Then

$$\phi_k(F) = \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot D_{\boldsymbol{\varepsilon}_k} \Phi_F + R_k$$

with remainder $|R_k| \leq C\sup_{\xi} \| D^2\Phi_F(\xi) \| \| \boldsymbol{\varepsilon}_k \|^2$ for a constant C depending only on $|\mathcal{K}|$and bounds on g.

Proof: Insert the expansion of Lemma Q.2 into the Shapley sum. The first-order term does not depend on S and factors out. The remainder is bounded uniformly across the finite set of coalitions. ∎

Contractive stability. Let $J_i = Df_i$ be the Jacobian of the local inheritance map. For each ancestor k, repeated application of the chain rule along citation paths yields the expression $D_{\boldsymbol{\varepsilon}_k}\Phi_F = \sum_{paths k\to F}\prod_{m=1}^{\ell} J_{i_m}$. Let ρ be the upper bound of all the J_{i_m} and assume the nonlinear inheritance system is **uniformly contractive**: $\| J_i \| \leq \rho < 1$. Then each path contribution satisfies $\| \prod_{m=1}^{\ell} J_{i_m} \| \leq \rho^{\ell}$. Because the DAG contains finitely many paths, $\| D\Phi_F \| \leq C\rho^{\ell(k,F)}$, and $|\phi_k(F)| \leq C\rho^{\ell(k,F)} \| \boldsymbol{\varepsilon}_k \|$, where $\ell(k,F)$ is the shortest path length between ancestor k and F. Thus Shapley attribution decays geometrically with citation depth under contractive inheritance.

Consider linear inheritance $\boldsymbol{e}_i = \sum_{j\in Pa(i)} A_{ij}\boldsymbol{e}_j + \boldsymbol{\varepsilon}_i$. Then $D_{\boldsymbol{\varepsilon}_k}\Phi_F = \sum_{paths k\to F}\prod A_{ij}$. If deep citation chains are dominated by a stationary operator A, then $\phi_k(F) \approx \nabla_2 g(\boldsymbol{e}_F, \boldsymbol{e}_F) \cdot A^{\ell(k,F)}\boldsymbol{\varepsilon}_k$. Taking norms gives $|\phi_k(F)| \sim \rho(A)^{\ell(k,F)}$, where $\rho(A)$ is the spectral radius. If we take a dynamical interpretation and let Let $\lambda = \lim_{\ell\to\infty}\frac{1}{\ell}\log \| \prod_{m=1}^{\ell} J_{i_m} \|$be the maximal Lyapunov exponent of the inheritance dynamics, so $e^{\lambda} = \rho$. Then we can have three different regimes based on value of λ.

Appendix 3. Canonical Narratives as Attractors of Epistemic Transport Dynamics

Scientific ideas propagate through citation networks via repeated acts of interpretation, summarization, and recombination. Each citation does not transmit the full semantic content of the original work; rather, it transmits a compressed representation that is adapted to the citing author's conceptual framework. Over long citation paths this process produces a striking empirical phenomenon: a small number of simplified, standardized explanations—*canonical narratives*—come to dominate discourse around a concept. These narratives are often shorter, more general, and more easily transmissible than the detailed arguments from which they originated.

The goal of this appendix is to show that this phenomenon arises naturally from the dynamics of epistemic transport on citation networks. Under weak assumptions, not only do low-dimensional directions survive as discussed in the main text, but also, more specifically, certain canonical narrative vectors appear as dynamical attractors of the epistemic transport operator, meaning that diverse initial interpretations converge toward a small set of stable narrative forms.

Let $\mathcal{H}$ denote the semantic Hilbert space defined in §5 in the main text, where vectors represent epistemic states (conceptual representations of a result). Each paper i carries an epistemic vector $e_i \in \mathcal{H}$. When a paper j cites paper i, the epistemic content of i is transformed before being incorporated into j. We model this transformation as a bounded linear operator

$$T_{ji}: \mathcal{H} \to \mathcal{H}$$

representing interpretation, summarization, and contextual adaptation. Thus epistemic propagation along a citation edge satisfies $e_j = \sum_{i \in Pa(j)} T_{ji} e_i + \epsilon_j$. On a citation chain $p_0 \to p_1 \to \cdots \to p_k$, define the transport operator $\mathcal{T}_k = T_k T_{k-1} \cdots T_1$, which represents cumulative reinterpretation over k citation steps. Empirical observations of citation behavior show that interpretation operators strongly favor conceptual compression. For practical reasons, authors typically emphasize a small number of conceptual dimensions, omit technical detail and generalize results into simplified claims when citing innovations. Mathematically this implies that the operators T_{ji} are **contractive** in most semantic directions. Let the singular value decomposition of T_{ji} be $T_{ji} = U \Sigma V^*$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$. Empirical plausibility suggests $\sigma_1 \approx 1$ and $\sigma_k \ll 1 fork > r$ for a small r. Thus each citation step preserves only a few semantic directions while strongly attenuating others.

Consider repeated multiplication of random contractive operators $\mathcal{T}_k = T_k T_{k-1} \ldots T_1$, from multiplicative ergodic theory (Oseledets theorem), the system admits **Lyapunov subspaces** characterized by growth rates $\lambda_1 > \lambda_2 > \cdots > \lambda_d$. Under the compression assumption,

$$\lambda_1 \approx 0, \lambda_k < 0 fork > r.$$

Therefore $\lim_{k \to \infty} \frac{\mathcal{T}_k x}{\|\mathcal{T}_k x\|}$ converges almost surely to the leading invariant subspace E_1. This subspace is the emergent semantic directions that survive long citation cascades. Further, we can define A **canonical narrative** to be a stable epistemic vector $v \in E_1$ that satisfies $Tv \approx v$ for the ensemble of citation operators T. In other words, canonical narratives are **fixed points (or near-fixed points)** of the epistemic transport process. They are conceptual statements that remain recognizable even after many reinterpretations.

Suppose two authors begin with different interpretations $x_0, y_0 \in \mathcal{H}$. After k citation steps: $x_k = \mathcal{T}_k x_0$, and $y_k = \mathcal{T}_k y_0$.Decompose each into components along the attractor subspace $E_1$and the orthogonal complement. Because the orthogonal components decay exponentially, $\| x_k - y_k \| \to 0$ after normalization. Thus diverse interpretations converge toward the same canonical narrative direction.

This analysis implies several structural properties of knowledge propagation through citation networks. First, narrative simplification is inevitable. Even if the original work contains high-dimensional conceptual structure, repeated citations compress it. Second, stable explanations dominate discourse. Once a narrative aligns with the attractor subspace, it becomes self-reinforcing: future citations reproduce it with minimal distortion. Third, multiple canonical narratives can coexist. If the attractor subspace has dimension $r > 1$, several stable narratives may emerge corresponding to different projections of the original idea. Citation networks act not merely as topological records of influence but as **dynamical systems that reshape knowledge**.

Appendix 4. Epistemic Entropy and Lyapunov Structure of Citation Inheritance

For analytical tractability we consider the local linearization of the inheritance map. Let $J_{Fk} = D_{\varepsilon_k}\Phi_F$ denote the sensitivity of the focal epistemic state to innovations at ancestor k. Then small perturbations propagate according to $x_F \approx \sum_{k\in\mathcal{K}} J_{Fk}\varepsilon_k$. If innovations have covariance matrices $\Sigma_k = Cov(\varepsilon_k)$, the resulting covariance of the focal epistemic state is

$$\Sigma_F = \sum_{k\in\mathcal{K}} J_{Fk}\Sigma_k J_{Fk}^{\top}.$$

This expression describes how epistemic variation is transported through the citation network. Assume Σ_F is positive definite. Define the **epistemic entropy** to be

$$H_F = \frac{1}{2} logdet(\Sigma_F).$$

This quantity measures the logarithmic volume of epistemic variability visible at the focal paper. Equivalently, $H_F = \sum_{i=1}^{d} log\sigma_i$, where σ_i are the singular values of the covariance operator. Thus epistemic entropy quantifies the effective dimensional volume of surviving innovations. Let x_t denote epistemic states along citation depth t. define $H_t = \frac{1}{2} logdet\Sigma_t$. The epistemic entropy production rate is $\Delta H_t = H_{t+1} - H_t$. This measures how epistemic variability changes as knowledge propagates through one additional citation step.

From the linearized inheritance dynamics, $x_{t+1} = Wx_t + \varepsilon_t$, the covariance evolves as

$$\Sigma_{t+1} = W\Sigma_t W^{\top} + \Sigma_\varepsilon.$$

This is the standard covariance recursion for linear stochastic systems.

If W is contractive, its singular values satisfy $\sigma_i(W) \le \rho(W) < 1$. Ignoring the innovation term temporarily, $\Sigma_{t+1} = W\Sigma_t W^{\top}$. Then $det(\Sigma_{t+1}) = det(W)^2 det(\Sigma_t)$. Taking logarithms on both sides, $Ht + 1 = Ht + log \mid detW \mid$. Since $|detW| < 1$, we obtain $H_{t+1} < H_t$. Thus, epistemic entropy strictly decreases under contractive inheritance. Reintroducing innovations, $\Sigma_{t+1} = W\Sigma_t W^{\top} + \Sigma_\varepsilon$. Innovation prevents entropy from collapsing completely, but contraction still bounds entropy growth. Specifically, $\Sigma_t \to \Sigma_\infty = \sum_{k=0}^{\infty} W^k \Sigma_\varepsilon (W^k)^{\top}$. Thus, epistemic variability approaches a stationary distribution.

Appendix 5. Incremental Identification of Epistemic Components

The theoretical framework developed in the main text assumes that each paper p admits an epistemic representation $e_p \in \mathbb{R}^d$. Because these representation vectors and their innovations $\{\varepsilon_k\}$ span the same subspace of the embedding space, in principle, one could recover the atomic vectors $\{\varepsilon_k\}$ by factorizing the global embedding matrix

$$E = WA,$$

where $E \in \mathbb{R}^{n\times d}$ contains paper embeddings, $W \in \mathbb{R}^{n\times K}$ indicates atomic usage, and $A \in \mathbb{R}^{K\times d}$ contains atomic vectors. However, in realistic citation networks: n may be millions of papers; the atomic dimension K is unknown; the full matrix E may not be available simultaneously. Therefore a practical theory requires **incremental identification**, where epistemic structure is learned progressively as papers arrive. This appendix shows that, under the contractive inheritance conditions in Appendix 2, **incremental estimation converges to the global decomposition**.

Assume papers arrive sequentially in time: $p_1, p_2, \dots, p_t$. Each paper satisfies the **streaming inheritance model**:

$$e_{p_t} = \sum_{q \in Pa(p_t)} f_{tq}(e_q) + \varepsilon_{p_t}$$

where $Pa(p_t)$ is the citation set, f_{tq} is the inheritance operator, and ε_{p_t} is the innovation component. Assume the contractive inheritance condition $\sum_{q \in Pa(p_t)} \| Df_{tq} \| \le \rho < 1$. This guarantees epistemic stability.

Now we propose **incremental atomic discovery.** Suppose we maintain an estimate set of atomic vectors $A_t = \{\varepsilon_1, \dots, \varepsilon_{K_t}\}$. In Appendix 1 we defined $A(p_t)$ to be the set of innovation atoms first introduced by paper p_t. Adapting same notation here, $A_t = \sqcup A(p_t)$. In addition, let $H(p_t)$ denote atoms paper p_t inherited from cited papers, and $U(p_t)$ all atoms in the paper. So, $U(p_t) = A(p_t) \sqcup H(p_t)$.

When paper p_t arrives:

1. Predict inherited embedding $e_{p_t}^{pred} = \sum_{q \in Pa(p_t)} f_{tq}(e_q)$. Here, f_{tq} can be estimated by regressing e_{p_t} on the set of $e_q, q \in Pa(p_t)$. The regression may be linear or in more complex forms of: $f_i(\{e_j\}) = \underset{f}{argmin} \| e_i - f(\{e_j\}) \|$.
2. Compute residual $r_t = e_{p_t} - e_{p_t}^{pred}$.
3. If $\| r_t \|$ exceeds tolerance ϵ, a new atom is introduced, interpret residual as new innovation atom introduced by paper p_t, $\varepsilon_t = r_t$.
4. Update the dictionary $A_{t+1} = A_t \cup \{\varepsilon_t\}$. This procedure produces an evolving dictionary $A_1 \subseteq A_2 \subseteq \cdots$ which approximates the true atomic set. Update $H(p_t)$ and $U(p_t)$ accordingly in similar manner.
5. Update embedding reconstruction $e_{p_t} = e_{p_t}^{pred} + \sum_{a \in A(p_t)} \varepsilon_a$
6. Return A_{t+1}. If studying epistemic lineage, also return $\{H(p_t)\}$ and $\{U(p_t)\}$.

We now show that this incremental procedure converges to the true epistemic decomposition.

Theorem: Convergence of Incremental Epistemic Decomposition

Assume

1. **Contractive inheritance**

$$\sum_{q \in Pa(p)} \| Df_{pq} \| \le \rho < 1.$$

2. **Bounded innovation magnitude**

$$\| \varepsilon_p \| \leq M.$$

3. **Exact paper embeddings e_p are observed.**

Then the incremental reconstruction satisfies

$$\lim_{t\to\infty} \| e_p - e_p \| = 0$$

for every paper p.

Proof

Let the reconstruction error for paper p_t be

$$\delta_t = e_{p_t} - e_{p_t}.$$

Using the inheritance model,

$$e_{p_t} = \sum_{q\in Pa(p_t)} f_{tq}(e_q) + \varepsilon_{p_t}.$$

Similarly,

$$e_{p_t} = \sum_{q\in Pa(p_t)} f_{tq}(e_q) + \varepsilon_{p_t}.$$

Subtracting yields

$$\delta_t = \sum_{q\in Pa(p_t)} \left(f_{tq}(e_q) - f_{tq}(e_q)\right) + (\varepsilon_{p_t} - \varepsilon_{p_t}).$$

Using Lipschitz continuity of f_{tq},

$$\| f_{tq}(e_q) - f_{tq}(e_q) \| \leq \| Df_{tq} \| \| \delta_q \|.$$

Therefore

$$\| \delta_t \| \leq \sum_{q\in Pa(p_t)} \| Df_{tq} \| \| \delta_q \| + \| \varepsilon_{p_t} - \varepsilon_{p_t} \|.$$

Applying the contractive bound, $\| \delta_t \| \leq \rho \max_{q\in Pa(p_t)} \| \delta_q \| + \epsilon$.

Iterating this recursion along citation paths gives

$$\| \delta_t \| \leq \rho^k \max_{q\in U_k(p_t)} \| \delta_q \| + \frac{\epsilon}{1-\rho}.$$

As $k \to \infty$, the first term vanishes.

If innovation residuals are eventually learned (so $\epsilon \to 0$), reconstruction error converges to zero:

$$\| \delta_t \| \to 0.$$

∎

Appendix 1 established that paper-level Shapley values depend only on the atomic decomposition of embeddings. Therefore, convergence of the decomposition implies **convergence of attribution**.

$$\lim_{t\to\infty} \Phi_t(p) = \Phi(p).$$

The attribution estimates obtained from streaming data converge to the values that would be obtained from the full epistemic network.

This incremental procedure does not require full knowledge of the global citation network. Instead, epistemic structure can be learned incrementally, attribution estimates improve as new papers arrive, and stability guarantees ensure eventual convergence. This aligns naturally with the historical evolution of science, where the structure of the knowledge network is revealed progressively. This framework is related to methods in online dictionary learning, sparse coding, and streaming matrix factorization. However, the epistemic framework differs in a key respect: the citation network constrains admissible decompositions through inheritance dynamics, producing the contractive structure required for identifiability.

In below we prove the theorem of **Robustness to Embedding Noise**. Let **the estimated embedding take the form**

$$\tilde{e}_p = e_p + \varepsilon_p,$$

where $\varepsilon_p \in \mathbb{R}^d$ is an embedding noise term. This appendix establishes that the epistemic decomposition and the resulting Shapley attribution remain **stable under bounded embedding noise (**$\| \varepsilon_p \| \leq \sigma$ for all papers p).

From §5, the innovation residual for paper p_t is defined as

$$r_t = e_{p_t} - \sum_{q \in C(p_t)} g_{tq}(e_q).$$

Under noisy embeddings, the estimated residual becomes

$$\tilde{r}_t = \tilde{e}_{p_t} - \sum_{q \in C(p_t)} g_{tq}(\tilde{e}_q).$$

Substituting the noise model gives

$$\tilde{r}_t = r_t + \varepsilon_{p_t} - \sum_{q \in C(p_t)} g_{tq}(\varepsilon_q).$$

Thus, the residual estimation error is

$$\Delta r_t = \tilde{r}_t - r_t = \varepsilon_{p_t} - \sum_{q \in C(p_t)} g_{tq}(\varepsilon_q).$$

Assume each inheritance operator g_{pq} is Lipschitz continuous with constant L_{pq}:

$$\| g_{pq}(x) - g_{pq}(y) \| \leq L_{pq} \| x - y \|.$$

Then

$$\| \Delta r_t \| \leq \| \varepsilon_{p_t} \| + \sum_{q \in C(p_t)} L_{tq} \| \varepsilon_q \|.$$

Using the bounded noise assumption,

$$\| \Delta r_t \| \leq \sigma \left(1 + \sum_{q \in C(p_t)} L_{tq} \right).$$

Under the contractive inheritance condition

$$\sum_{q \in C(p_t)} L_{tq} \leq \rho < 1,$$

we obtain the bound of estimation error

$$\| \Delta r_t \| \leq \sigma(1 + \rho).$$

Similarly, the paper level Shapley attribution error satisfies

$$|\tilde{\Phi}(p) - \Phi(p)| \leq C\sigma,$$

where C is a constant depending on the depth of the citation network and the Lipschitz constants of the inheritance operators.

Under embedding noise, the algorithm estimates innovation as

$$\tilde{r}_t = r_t + \Delta r_t.$$

Using the triangle inequality:

$$| \| \tilde{r}_t \| - \| r_t \| | \leq \| \Delta r_t \| \leq \sigma(1 + \rho)$$

shows that the measured magnitude of innovation deviates within a bound. If $\| r_t \| > \epsilon + \sigma(1 + \rho)$, then $\| \tilde{r}_t \| \geq \| r_t \| - \sigma(1 + \rho) > \epsilon$ and the algorithm correctly detects an innovation atom despite the embedding noise. If $\| r_t \| < \epsilon - \sigma(1 + \rho)$, then we have $\| \tilde{r}_t \| \leq \| r_t \| + \sigma(1 + \rho) < \epsilon$ and the algorithm will correctly refrain from introducing a spurious innovation atom. Within the narrow band, $\epsilon - \sigma(1 + \rho) \leq \| r_t \| \leq \epsilon + \sigma(1 + \rho)$ embedding noise may cause the algorithm to either introduce or omit an innovation atom falsely. However, when innovations are sufficiently large relative to the noise level, atom identification remains stable.

Part II. Spectral Dynamics of Scientific Ideas and Credit

Abstract

Scientific ideas evolve through networks of citation, interpretation, and reinterpretation, yet the mechanisms governing the **dynamics of knowledge and the distribution of scientific credit** remain poorly understood. Existing models of citation networks explain patterns such as preferential attachment and cumulative advantage, but they rarely account for the **transformative nature of intellectual inheritance**, whereby each scientific work reinterprets and compresses the ideas it inherits. Here we introduce a mathematical framework that models the evolution of scientific knowledge as a process of **epistemic transport in a conceptual space**, where papers are represented as vectors and citations act as linear operators that transform inherited ideas. Along citation chains, these transformations compose multiplicatively, producing a random dynamical system on idea space.

Applying the multiplicative ergodic theory of Vladimir Oseledets (Oseledets, 1968; Ruelle, 1979; Walters, 1993), we show that the long-term evolution of ideas is governed by a **Lyapunov spectrum** that decomposes conceptual space into expanding, neutral, and contracting epistemic subspaces. This spectral structure yields several fundamental results. First, scientific fields exhibit **conceptual compression**, whereby the effective dimensionality of idea space collapses over time into a small set of canonical conceptual modes. Second, epistemic transport generates **multiplicative propagation of intellectual influence**, implying that credit contributions from ancestors decay exponentially with citation distance while accumulating multiplicatively along citation cascades. These dynamics naturally produce **heavy-tailed distributions of scientific credit**, explaining the emergence of highly influential canonical papers and "superstar" scientists.

The framework also provides a mathematical account of **paradigm formation and scientific revolutions** in the sense described by Thomas Kuhn. We show that shifts between egalitarian and highly unequal credit regimes occur at a critical point determined by the relation between the growth rate of epistemic transport and the branching structure of citation networks. At this threshold, the dominant conceptual subspaces governing knowledge production can rotate, corresponding to large-scale reorganization of scientific fields. These dynamics connect the sociology of science with models of cultural evolution, including Dual Inheritance Theory.

Our theory yields several empirically testable predictions, including measurable **credit half-lives of ideas**, **dimensionality collapse in mature scientific fields**, and a quantitative relation linking the power-law exponent of credit inequality to the spectral growth rate of epistemic transport. Together, these results provide a unified mathematical theory of **knowledge evolution, scientific credit, and paradigm change**, bridging network science, cultural evolution, and the quantitative study of human knowledge production.

1. Introduction

1.1 The problem of scientific credit and idea evolution

Scientific knowledge advances through a cumulative process in which new ideas build upon prior work. Each scientific publication inherits concepts, methods, and interpretations from earlier research, modifies them, and transmits them further through subsequent citations. In this sense, science evolves through a network of intellectual inheritance. Yet despite the centrality of this process, the **mechanisms governing how ideas propagate and how intellectual credit accumulates remain poorly understood**.

A longstanding puzzle concerns the striking **inequality in the distribution of scientific credit**. A small fraction of papers and researchers receive a disproportionately large share of citations and recognition, while the majority receive relatively little attention. Sociological studies of science have documented this phenomenon for decades. In particular, Robert K. Merton described the **Matthew effect**, whereby already prominent scientists tend to receive disproportionate recognition for their contributions. Empirical analyses of citation networks confirm that scientific impact is highly skewed, often approximating heavy-tailed or power-law distributions. However, while these patterns are well documented, the underlying mechanisms producing them remain debated.

Traditional models of citation networks typically explain these inequalities through mechanisms such as preferential attachment, where highly cited papers are more likely to receive additional citations. While such models successfully reproduce certain statistical properties of citation networks, they largely treat citations as **links of attention or popularity**, rather than as channels through which ideas are transmitted and transformed. In reality, a citation rarely represents a passive act of copying influence. Instead, a citing paper actively **interprets, reframes, and recombines the ideas of its predecessors**, embedding them within a new conceptual context. Scientific knowledge therefore evolves not merely through network growth but through **continuous reinterpretation of inherited ideas**.

Understanding the dynamics of this process is essential for addressing deeper questions about the evolution of knowledge. How do conceptual frameworks stabilize within scientific fields? Why do certain papers become canonical intellectual ancestors while others fade into obscurity? How do paradigm shifts reorganize the intellectual structure of entire disciplines, as famously described by Thomas Kuhn (1962)? These questions suggest that the evolution of science should be viewed not only as a network process but also as a **dynamical system operating in a space of ideas**.

In this paper we propose a mathematical framework that models scientific development as **epistemic transport in a conceptual space**. Papers are represented as vectors in an idea space, and citations correspond to transformations that reinterpret inherited ideas. Along chains of citation, these transformations compose multiplicatively, generating a dynamical system that governs the evolution of knowledge and the propagation of scientific credit. This perspective allows us to analyze the long-run behavior of idea evolution using tools from random dynamical systems and spectral theory. As we show below, this approach reveals deep connections between the structure of citation networks, the emergence of canonical ideas, and the unequal distribution of scientific credit across generations of research.

1.2 Limitations of existing models

A large body of work in bibliometrics and network science has attempted to explain the statistical structure of citation networks. Many influential models treat citation dynamics as a stochastic network growth process in which new papers attach to existing papers according to probabilistic rules. One of the most widely studied mechanisms is **preferential attachment**, introduced by Albert-László Barabási and Réka Albert (2002), in which the probability that a paper receives new citations is proportional to its existing citation count. Such models successfully reproduce heavy-tailed citation distributions and other empirical regularities observed in scientific networks.

However, while network growth models capture **how citations accumulate**, they generally do not describe **what is being transmitted through citations**. In most formulations, a citation is treated as a simple edge indicating attention, visibility, or popularity. The intellectual content of the cited work—the **ideas, concepts, and interpretations** being inherited—is typically not modeled explicitly. As a result, these models explain citation patterns largely through social or stochastic mechanisms rather than through the **dynamics of ideas themselves**. Another class of approaches analyzes scientific knowledge through semantic embeddings of papers or topics, often using modern representation-learning techniques. These methods capture the geometry of scientific discourse and allow researchers to measure conceptual similarity between papers or fields. While such approaches reveal important structural features of the knowledge landscape, they typically remain **descriptive rather than dynamical**. They characterize the structure of idea space at a given time but do not provide a principled model of how ideas transform as they propagate across citation chains.

More broadly, existing frameworks rarely provide a unified explanation for several fundamental empirical phenomena observed in scientific evolution. These include:

1. **Credit concentration.** Scientific credit is extremely unevenly distributed, with a small number of papers becoming canonical references while most contributions receive limited recognition.
2. **Conceptual stabilization.** Mature scientific fields often converge on a small number of dominant conceptual frameworks that structure subsequent research.
3. **Paradigm shifts.** Periodically, scientific fields undergo abrupt reorganizations in which existing conceptual structures are replaced by new ones, a phenomenon famously described as "paradigm shifts" by Thomas Kuhn.

Existing models typically explain these phenomena separately—through network growth mechanisms (Wang et al., 2013), sociological processes (Merton, 1968), or historical narratives (Kuhn, 1962). What remains largely missing is a **mathematical framework that links the propagation of ideas to the accumulation of credit and the large-scale evolution of scientific paradigms**.

The key limitation underlying these approaches is that they treat citation networks primarily as **static or combinatorial structures**, rather than as **dynamical systems governing the flow of conceptual information**. In reality, when a paper cites prior work it does not merely point to it; it **interprets and transforms the ideas contained in that work**. These transformations propagate along citation chains, gradually reshaping the conceptual structure of a field. Understanding this process requires modeling citations as operators acting on ideas rather than as simple edges in a network.

The framework developed in this paper addresses this gap by modeling citations as **interpretation operators** acting on vectors in an idea space. Along citation cascades, these operators compose multiplicatively, producing a dynamical system whose long-run behavior can be analyzed using spectral theory. As we will show, this perspective provides a unified explanation for several central phenomena in the evolution of science, including conceptual compression, the emergence of canonical ideas, and the heavy-tailed distribution of scientific credit. Our theory derives conditional indicators for a field approaching revolution.

1.3 Overview of our framework

To address the limitations of existing approaches, we develop a mathematical framework that models the evolution of scientific knowledge as a **dynamical process in a conceptual space**. The central idea is to treat scientific papers not merely as nodes in a citation network but as **representations of ideas**, and citations as **transformations that reinterpret and propagate those ideas across generations of research**.

We begin by representing each scientific work as a vector in a conceptual space $\mathbb{R}^d$. The coordinates of this vector correspond to the paper's conceptual components—its key ideas, methods, or theoretical elements. In practice such representations can be approximated using semantic embeddings derived from textual content, but the theoretical framework does not depend on a particular embedding method. Instead, the embedding serves as an abstract representation of the **conceptual structure of scientific knowledge**.

Within this representation, a citation is interpreted as an **epistemic transformation**. When paper j cites paper i, it does not merely reproduce the ideas of i; it interprets them within a new intellectual context. We model this process using an **interpretation operator**

$$e_j = A_{ji} e_i,$$

where e_i and e_j denote the conceptual representations of the two papers and A_{ji} is a linear operator describing how the ideas of paper i are transformed when incorporated into paper j. More generally, when a paper integrates ideas from multiple predecessors, its conceptual representation can be written as a combination of such transformations applied to its cited sources.

This formulation naturally leads to the notion of **epistemic transport along citation chains**. Consider a sequence of papers

$$i_0 \to i_1 \to i_2 \to \cdots \to i_n$$

where each paper cites the previous one. The cumulative transformation transporting conceptual contributions from the ancestor i_0 to the descendant i_n is given by the product of interpretation operators

$$T_n = A_{i_n i_{n-1}} A_{i_{n-1} i_{n-2}} \cdots A_{i_1 i_0}.$$

Thus, the propagation of ideas across generations of research is governed by **products of interpretation operators**. Because these operators vary across papers and contexts, the resulting dynamics form a **random dynamical system on idea space**.

A key consequence of this formulation is that the long-run evolution of ideas is governed by the **spectral properties of these operator products**. Using tools from the theory of random matrix products, particularly the multiplicative ergodic theorem of Vladimir Oseledets, we show that conceptual space decomposes into subspaces characterized by distinct growth rates. Some conceptual directions are amplified through repeated reinterpretation, others remain approximately conserved, and many decay over time.

This spectral structure generates several striking implications. First, it explains why scientific fields tend to organize around a small number of **canonical conceptual modes**, as repeated reinterpretation suppresses most directions in idea space. Second, it implies that the propagation of intellectual contributions along citation cascades follows a **multiplicative process**, leading naturally to the heavy-tailed distribution of scientific credit. Finally, the theory predicts the existence of **critical transitions in the dynamics of knowledge evolution**, where shifts in the spectral structure of interpretation operators reorganize the conceptual foundations of entire fields.

Taken together, this framework provides a unified mathematical description of how ideas propagate, stabilize, and occasionally reorganize within the scientific enterprise. By linking the dynamics of conceptual evolution to the structure of citation networks, it allows us to analyze the emergence of scientific paradigms, the concentration of intellectual credit, and the large-scale evolution of knowledge within a single theoretical framework.

1.4 Roadmap of the paper

The remainder of the paper develops the theoretical framework outlined above and explores its implications for the evolution of scientific ideas and the distribution of intellectual credit.

Section 2 introduces the mathematical representation of scientific knowledge. We formalize the notion of an **idea space** in which scientific works are represented as vectors and define **interpretation operators** that model how citing papers transform the ideas they inherit. These definitions lead naturally to the concept of **epistemic transport**, describing how conceptual contributions propagate along citation chains. We also introduce a cooperative attribution framework that allows the epistemic contributions of ancestor papers to be quantified.

Section 3 examines the consequences of repeated interpretation for the geometry of idea space. We show that interpretation tends to act as a form of **information compression**, suppressing many conceptual directions while amplifying a smaller set of persistent ones. This process leads to **conceptual dimensionality collapse**, in which mature scientific fields become organized around a limited number of dominant conceptual modes. We also discuss how the dimensionality of conceptual space evolves during the life cycle of a scientific field.

Section 4 develops the spectral theory underlying epistemic transport. By modeling interpretation operators as a stochastic dynamical system and applying the multiplicative ergodic theorem of Vladimir Oseledets, we show that idea space decomposes into subspaces associated with distinct **Lyapunov exponents**. This decomposition provides the mathematical foundation for analyzing the long-run evolution of conceptual contributions along citation cascades.

Section 5 studies the structure of **conserved epistemic subspaces** arising from this spectral decomposition. We show that the dimension of the neutral Lyapunov subspace determines the number of conceptual quantities that remain approximately invariant under repeated reinterpretation. These conserved directions correspond to the stable conceptual frameworks that structure mature scientific fields.

Section 6 analyzes the dynamics of scientific credit within this framework. We demonstrate that conceptual contributions propagate multiplicatively along citation cascades, implying exponential decay of influence with citation distance and naturally producing **heavy-tailed distributions of intellectual credit**. The theory also predicts measurable quantities such as the **credit half-life of scientific ideas** and provides a quantitative relation linking credit inequality to both network structure and epistemic growth rates.

Section 7 explores the implications of the theory for **paradigm formation and scientific revolutions**. We show that shifts in the spectral structure of epistemic transport can induce phase transitions in the organization of scientific fields, corresponding to the emergence or collapse of dominant conceptual frameworks. This perspective provides a dynamical interpretation of paradigm shifts in the sense described by Thomas Kuhn.

Section 8 concludes by discussing the broader implications of the framework for the quantitative study of knowledge evolution, connecting the dynamics of epistemic transport to theories of cultural evolution and the sociology of science.

Technical details and mathematical proofs are provided in the appendices.

2. Mathematical Representation of Scientific Ideas

2.1 Idea Space

To study the evolution of scientific knowledge mathematically, we first introduce a geometric representation of scientific ideas. The central assumption of our framework is that the conceptual content of a scientific work can be represented as a vector in a high-dimensional **idea space**.

Let

$$\mathcal{I} \cong \mathbb{R}^d$$

denote a d-dimensional vector space whose coordinates correspond to latent conceptual directions. A scientific paper i is represented by a vector

$$\boldsymbol{e}_i \in \mathcal{I}.$$

The coordinates of $\boldsymbol{e}_i$ encode the conceptual content of the work along different dimensions of the idea space. In practice these coordinates may correspond to latent semantic directions derived from text embeddings, topic models, or other representations of scientific content. In the present theory, however, the specific construction of the coordinates is not essential; what matters is that the conceptual structure of ideas can be embedded in a vector space where linear transformations capture reinterpretations of ideas.

Citation Network. Scientific works are connected through a directed acyclic graph (DAG) representing the citation structure of the literature. If paper i cites paper j, we write

$$j \to i.$$

The set of parents (cited papers) of i is denoted

$$Pa(i).$$

Thus each paper inherits conceptual material from the vectors associated with its cited predecessors.

Local Idea Construction. We model the conceptual content of a paper as emerging from the reinterpretation of ideas inherited from its cited ancestors. Formally, the idea vector $\boldsymbol{e}_i$ is generated from the ideas of its parents through an **interpretation operator**

$$\boldsymbol{e}_i = f_i(\boldsymbol{e}_{Pa(i)}),$$

where f_i is a transformation that combines and modifies the ideas present in the cited works. Because each paper typically cites multiple predecessors, $\boldsymbol{e}_{Pa(i)}$ represents the collection of parent vectors

$$\boldsymbol{e}_{Pa(i)} = (\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \dots, \boldsymbol{e}_{j_k}), j_\ell \in Pa(i).$$

The function f_i captures the intellectual process by which authors reinterpret prior work: selecting certain conceptual directions, modifying them, and synthesizing them into new contributions.

Linearization and Local Geometry. To analyze the propagation of ideas through long citation chains, it is convenient to study the local linear approximation of the interpretation function. Around a given conceptual configuration, the interpretation map can be linearized using its Jacobian:

$$A_i = Df_i(\boldsymbol{e}_{Pa(i)}).$$

The matrix A_i describes how small conceptual perturbations in the parent papers propagate into the conceptual content of paper i . In other words, A_i captures how the citing work amplifies, suppresses, or mixes conceptual directions inherited from earlier work.

This local linearization plays a central role in our theory because repeated multiplication of these matrices along citation chains determines the long-run dynamics of ideas.

Conceptual Contributions as Tangent Vectors. We represent the marginal conceptual contribution of an ancestor paper as a vector in the tangent space of the idea space. Let

$$v_j \in T_{\boldsymbol{e}_j}\mathcal{I}$$

denote a conceptual perturbation introduced by paper j. Under interpretation by a descendant paper i, this perturbation transforms according to the derivative of the interpretation operator:

$$v_i = Df_i(\boldsymbol{e}_{Pa(i)})v_j.$$

This rule defines how conceptual contributions propagate through the citation network.

Conceptual Transport on Citation Networks. Taken together, the idea vectors $\boldsymbol{e}_i$ and the linear operators A_i define a **dynamical system on the citation graph**. Conceptual contributions move through this network via repeated application of interpretation operators along citation paths.

This process will be referred to as **epistemic transport**: the dynamical propagation of ideas through the scientific literature. The spectral properties of the resulting operator products will determine which conceptual directions persist, which decay, and how intellectual credit accumulates across generations of scientific work.

In the following sections we formalize the structure of these interpretation operators and derive the dynamical laws governing epistemic transport along citation chains.

2.2 Interpretation Operators

Scientific progress rarely consists of simple replication of prior ideas. Instead, each new work **interprets**, **reorganizes**, and **compresses** the conceptual material inherited from earlier research. In our framework, this intellectual transformation is represented by an **interpretation operator** acting on idea space.

Definition: Interpretation Operator. Let paper i cite papers $j_1, \dots, j_k$. The conceptual content of paper i is generated by applying an interpretation operator to the ideas of its cited predecessors. Formally, we define

$$\boldsymbol{e}_i = A_i\left(\sum_{j\in Pa(i)} w_{ij}\boldsymbol{e}_j\right)$$

where $A_i \in \mathbb{R}^{d\times d}$ is the **interpretation operator** associated with paper i, local linearization of f_i; w_{ij} are weights describing how strongly paper i relies on parent j; $\boldsymbol{e}_j$ are the idea vectors of cited works. The weights satisfy

$$\sum_{j\in Pa(i)} w_{ij} = 1,$$

so the parent ideas form a convex conceptual mixture before reinterpretation.

Interpretation as Conceptual Transformation. The matrix A_i captures how the citing paper transforms inherited ideas. Conceptually, this operator represents several common intellectual processes:

1. **Selection**
   Some conceptual directions from the cited literature are emphasized while others are ignored.
2. **Synthesis**
   Ideas from multiple ancestors may be combined into new conceptual directions.
3. **Compression**
   Complex sets of prior ideas may be summarized into simpler conceptual frameworks.

4. **Expansion**
Occasionally a paper introduces new conceptual directions not previously present in the literature.

Mathematically, these operations correspond to **linear transformations** of the idea space.

Spectral Structure of Interpretation. The singular value decomposition of A_i

$$A_i = U_i \Sigma_i V_i^{\top}$$

reveals how interpretation reshapes conceptual space. The components have natural epistemic interpretations:

- V_i: conceptual directions inherited from prior work,
- Σ_i: scaling factors indicating amplification or suppression,
- U_i: conceptual directions expressed in the new paper.

If a singular value is large, the corresponding conceptual direction is strongly amplified. If it is small, that conceptual direction is effectively compressed or forgotten.

Dimensional Compression. In practice, scientific papers often reinterpret complex bodies of literature through simpler conceptual narratives. As a result, interpretation operators frequently act as low-rank approximations:

$$rank(A_i) < d.$$

This reflects the fact that a citing paper typically highlights only a small number of conceptual directions from a large body of prior work. Many directions of the idea space are therefore suppressed during interpretation. Repeated interpretation along citation chains can therefore lead to progressive **dimensionality collapse** of idea space, a phenomenon we analyze formally in later sections.

Interpretation Along Citation Chains. Consider a citation chain

$$k \to i_1 \to i_2 \to \cdots \to i_m \to F.$$

The conceptual influence of ancestor k on the final paper F is obtained by composing interpretation operators along the chain:

$$A_{F:i_1} = A_F A_{i_m} \cdots A_{i_1}.$$

This product describes how conceptual contributions are transformed as they propagate through successive generations of scientific work. The long-run behavior of such operator products determines which conceptual directions persist through the literature and which vanish over time. The mathematical theory governing these dynamics is provided by the **multiplicative ergodic theorem**, which we introduce in Section 4.

Interpretation as a Stochastic Process. Interpretation operators vary across papers because different authors reinterpret prior ideas in different ways. We therefore model the sequence of operators A_i encountered along citation paths as realizations of a stochastic process drawn from some distribution over linear transformations. This stochastic perspective will allow us to apply tools from random matrix theory and ergodic dynamics to characterize the typical long-run evolution of ideas. In particular, the spectral properties of products of interpretation operators will determine the emergence of **canonical conceptual modes**, the **collapse of conceptual dimensionality**, and the **distribution of scientific credit across ancestors**. These results form the technical foundation of the theory developed in the following sections.

2.3 Epistemic Transport

Scientific knowledge evolves through a network of citations in which ideas are repeatedly interpreted, recombined, and transmitted. In the mathematical framework introduced above, interpretation operators transform conceptual vectors locally at each paper. We now formalize how ideas **propagate through the citation network as a transport process**.

Citation Networks as Directed Graphs. Let the scientific literature be represented as a directed acyclic graph (DAG)

$$G = (V, E),$$

where V is the set of papers; E is the set of citation edges. An edge $j \to i$ indicates that paper i cites paper j.

Because citations always point backward in time, the resulting graph is acyclic and naturally defines a **temporal flow of ideas** from earlier papers to later ones. Each node $i \in V$ carries an idea vector $\boldsymbol{e}_i \in \mathbb{R}^d$. The goal of epistemic transport theory is to describe how the conceptual influence of one paper propagates through this network.

Transport Along Citation Edges. From Section 2.2, a citing paper constructs its conceptual representation through

$$\boldsymbol{e}_i = A_i \left(\sum_{j \in Pa(i)} w_{ij} \boldsymbol{e}_j \right).$$

This equation describes a **local transport rule**: conceptual information from parents j is first aggregated and then transformed by the interpretation operator A_i. Consequently, the conceptual influence of an ancestor k on paper i propagates through chains of interpretation operators.

Transport Along Citation Paths. Consider a citation path $k \to i_1 \to i_2 \to \cdots \to i_m \to F$. The conceptual influence transmitted along this path is given by the operator product

$$T_{k \to F}^{(p)} = A_F A_{i_m} \cdots A_{i_1}.$$

Applying this operator to the original idea vector $\boldsymbol{e}_k$ yields the portion of the final paper's idea space attributable to ancestor $\boldsymbol{e}_k$ through path p.

Multiple Transport Paths. In realistic citation networks, an ancestor may influence a paper through many distinct citation paths. Let

$$\mathcal{P}(k, F)$$

denote the set of all citation paths connecting ancestor k to descendant F.

The total conceptual influence transmitted from k to F is therefore the sum over all transport paths:

$$T_{k \to F} = \sum_{p \in \mathcal{P}(k,F)} T_{k \to F}^{(p)}$$

This operator aggregates all ways in which the conceptual content of paper k can propagate through the literature to influence paper F.

Transport Distance. Citation paths may have different lengths. Let $\ell(p)$ denote the number of citation steps in path p. The epistemic distance between two papers can therefore be defined as

$$\ell(k, F) = \min_{p \in \mathcal{P}(k,F)} \ell(p),$$

the length of the shortest citation path connecting them. This notion of distance will later play an important role in quantifying **how influence decays with citation distance**.

Interpretation as Transport with Distortion. Unlike classical diffusion processes, epistemic transport does not simply move information through a network. Instead, interpretation operators distort the conceptual representation at every step. Consequently, some conceptual directions are amplified, others are attenuated or lost and new directions may occasionally be introduced. The cumulative effect of these transformations determines which ideas survive long chains of reinterpretation.

From Local Transport to Global Dynamics. Epistemic transport thus arises from the repeated multiplication of interpretation operators along citation paths. Over long citation chains, these operator products exhibit structured statistical behavior governed by the spectral properties of random matrix products.

In later sections we will show that this transport process leads naturally to several large-scale regularities in scientific knowledge:

- **dimensional compression of idea space**,
- **emergence of canonical conceptual modes**,
- **exponential decay of influence with citation distance**, and
- **heavy-tailed distributions of scientific credit**.

To analyze these phenomena rigorously, we next introduce mechanisms for attributing epistemic credit across citation networks.

2.4 Shapley Epistemic Attribution

The framework developed above describes how ideas propagate through citation networks via interpretation operators. However, understanding the **distribution of scientific credit** requires an additional step: we must determine how much of a descendant paper's conceptual content should be attributed to each of its intellectual ancestors.

This attribution problem is inherently **cooperative**. A scientific paper typically synthesizes ideas from multiple predecessors, and the resulting conceptual contribution cannot be uniquely assigned to any single source. Instead, credit must be allocated among multiple ancestors whose ideas jointly contribute to the final work. To formalize this process we adopt a principle from cooperative game theory originally introduced by Lloyd Shapley (1953): the **Shapley value**.

Conceptual Contribution as a Cooperative Game. Consider a paper F whose conceptual content is influenced by a set of ancestor papers

$$A(F) = \{k_1, k_2, \dots, k_n\}.$$

Each subset of ancestors contributes some portion of the final idea vector. Let

$$v(S)$$

denote the conceptual contribution generated when only the subset $S \subseteq A(F)$ is present.

In our setting, this contribution arises through epistemic transport along citation paths originating from papers in S. Formally, we can measure the magnitude of conceptual influence by the norm of the transported idea vectors:

$$v(S) = \left\| \sum_{k \in S} T_{k \to F} \mathbf{e}_k \right\|.$$

This function defines a cooperative game in which the "players" are the ancestor papers and the value of a coalition S is the conceptual contribution produced by those ancestors.

Shapley Value Attribution. The Shapley value assigns each ancestor a fair share of the total conceptual contribution. For ancestor k, the Shapley attribution is defined as

$$\phi_k(F) = \sum_{S \subseteq A(F) \setminus \{k\}} \frac{|S|!\,(n - |S| - 1)!}{n!} [v(S \cup \{k\}) - v(S)].$$

Intuitively, this formula averages the **marginal contribution** of ancestor k across all possible orders in which the ancestors could be combined to construct the final paper.

The Shapley value has several unique properties desirable for fair credit attribution: **Efficiency** (the total credit assigned to all ancestors equals the total conceptual contribution), **Symmetry** (Ancestors with identical contributions receive equal credit), **Null player property** (Ancestors whose ideas do not influence the final paper receive zero credit) and **Additivity**. These properties make the Shapley value an ideal candidate for quantifying intellectual credit within a cooperative system of idea production.

Connection to Epistemic Transport. Because conceptual contributions propagate multiplicatively through interpretation operators, the transport operators

$$T_{k\to F}$$

determine the magnitude of each ancestor's influence. Substituting the transport operators into the value function yields

$$v(S) = \| \sum_{k\in S} \left(\sum_{p\in\mathcal{P}(k,F)} T_{k\to F}^{(p)} \right) \boldsymbol{e}_k \|.$$

Thus the credit attributed to each ancestor depends on the cumulative transformation of its ideas across all citation paths leading to the focal paper.

Aggregated Credit Across the Literature. The total epistemic credit of paper k within the scientific literature can then be defined as the sum of its contributions to all descendant papers:

$$C_k = \sum_{F\in Desc(k)} \phi_k(F).$$

This quantity measures the cumulative conceptual influence of paper k across the entire citation network.

From Attribution to Credit Dynamics. The Shapley attribution framework connects the *local* propagation of ideas to the *global* distribution of scientific credit. Because conceptual influence propagates multiplicatively through interpretation operators, the credit assigned to ancestors inherits the statistical properties of the epistemic transport process. As we will show in later sections, these dynamics generate several striking consequences:

- ancestor influence decays exponentially with citation distance,
- credit accumulates multiplicatively along citation cascades,
- the resulting distribution of scientific credit becomes heavy-tailed.

These results provide a mathematical explanation for the extreme inequality observed in citation distributions and for the emergence of highly influential canonical papers in the scientific literature.

3. Compression of Idea Space

3.1 Interpretation as Information Compression

The framework introduced in Section 2 describes how scientific ideas propagate through citation networks via interpretation operators. A key empirical observation about scientific discourse, however, is that reinterpretation rarely preserves the full conceptual complexity of prior work. Instead, citing papers typically **select, summarize, and reorganize** the ideas they inherit. As a result, interpretation often acts as a form of **information compression**.

At a conceptual level, scientific papers rarely reproduce the entire intellectual content of their predecessors. Instead, they extract a smaller set of conceptual directions—core hypotheses, theoretical principles, or methodological innovations—and embed them within new contexts. This process can be viewed as a mapping from a higher-dimensional conceptual representation of prior work to a **lower-dimensional effective representation** in the new paper.

Within the idea-space framework, this phenomenon corresponds to the action of interpretation operators that reduce the effective dimensionality of conceptual representations. Recall from Section 2 that a citing paper i transforms inherited ideas through a linear operator $A_i \in \mathbb{R}^{d\times d}$. When the rank of this operator is smaller than the ambient dimension of the idea space,

$$rank(A_i) < d,$$

the transformation necessarily compresses the conceptual information inherited from the cited literature. This compression reflects the fact that a paper typically highlights only a limited number of conceptual directions from a potentially large body of prior work. Many dimensions of the inherited idea vectors are therefore suppressed during reinterpretation.

Compression Through Singular Value Structure. The degree of compression induced by an interpretation operator can be understood through its singular value decomposition $A_i = U_i\Sigma_iV_i^\top$,where the diagonal matrix Σ_i contains singular values

$$\sigma_{i1} \geq \sigma_{i2} \geq \cdots \geq \sigma_{id} \geq 0.$$

Conceptual directions associated with large singular values are preserved or amplified by the interpretation, while directions associated with small singular values are strongly attenuated. If many singular values are close to zero, the operator effectively projects the inherited ideas onto a **lower-dimensional conceptual subspace**. From this perspective, the process of scientific interpretation can be viewed as a sequence of **conceptual projections** that gradually concentrate intellectual discourse into a smaller number of dominant directions.

Iterated Compression Along Citation Chains. Because interpretation operators compose multiplicatively along citation chains, conceptual compression compounds across generations of scientific work. Consider a sequence of papers connected by citations,$i_0 \to i_1 \to i_2 \to \cdots \to i_n$.The cumulative transformation applied to the ideas of the original paper is

$$T_n = A_{i_n}A_{i_{n-1}}\cdots A_{i_1}.$$

Even if each individual operator reduces dimensionality only modestly, repeated multiplication can strongly concentrate conceptual influence into a small number of directions. Many components of the original idea vector are gradually attenuated, while a few directions remain persistent. This process produces a progressive **collapse of effective conceptual dimensionality** across the literature.

Compression and Scientific Narratives. The compression mechanism described above corresponds closely to the way scientific narratives emerge in practice. Over time, complex bodies of research are often summarized through simplified conceptual frameworks: a small set of central ideas, principles, or models that capture the essence of the field. These frameworks serve as organizing structures for subsequent research, guiding both interpretation and communication.

Within our model, such frameworks correspond to **dominant directions in idea space** that survive repeated compression through interpretation operators. Conceptual directions that fail to align with these dominant modes gradually disappear from the active discourse of the field.

This observation motivates the next section, where we analyze the consequences of repeated interpretation for the dimensional structure of idea space and show how conceptual compression leads to the emergence of a small number of **canonical conceptual modes** that organize mature scientific fields.

3.2 Conceptual Dimensionality Collapse

The compression mechanism described above has an important dynamical consequence for the evolution of scientific ideas. When interpretation operators repeatedly compress conceptual information, the effective dimensionality of idea space gradually decreases. Over long citation chains, only a small number of conceptual directions remain influential, while most others are suppressed. We refer to this phenomenon as **conceptual dimensionality collapse**.

Effective Conceptual Dimension. Let the idea space have ambient dimension d. For a collection of papers within a scientific field, we can measure the effective conceptual dimension as the number of principal directions needed to explain most of the variance in their idea vectors. Formally, if E denotes the matrix whose rows are idea vectors $\boldsymbol{e}_i$, the covariance matrix

$$\Sigma = \mathbb{E}\left[\boldsymbol{e}_i \boldsymbol{e}_i^\top\right]$$

captures the distribution of conceptual content in the field. Let

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

be the eigenvalues of Σ. The effective conceptual dimension

$$d_{eff}$$

can be defined as the number of eigenvalues exceeding a threshold or, more generally, as the number of dominant principal components capturing the majority of conceptual variance. Empirically, this effective dimension is often **much smaller than the ambient dimension** of the embedding space.

Dimensional Collapse Under Repeated Interpretation. The theoretical reason for this dimensional reduction lies in the multiplicative structure of interpretation operators. Consider again the cumulative transformation along a citation chain, $T_n = A_n A_{n-1} \cdots A_1$. If the operators A_i tend to suppress many conceptual directions (i.e., have small singular values in those directions), repeated multiplication amplifies the disparity between dominant and suppressed components. Over time, the operator product becomes effectively low-rank. More precisely, if the singular values of T_n are

$$\sigma_1(T_n) \geq \sigma_2(T_n) \geq \cdots \geq \sigma_d(T_n),$$

then the ratio

$$\frac{\sigma_k(T_n)}{\sigma_1(T_n)}$$

tends to decay exponentially for sufficiently large k. As a result, only the leading singular directions remain significant after many generations of reinterpretation.

Emergence of Low-Dimensional Idea Manifolds. Because most conceptual directions are attenuated, the idea vectors of papers in a mature field tend to concentrate near a **low-dimensional manifold** embedded within the larger idea space. That is, although the theoretical embedding space may have high dimension d, the set of realized scientific ideas occupies only a much smaller subspace. Intuitively, this means that a mature scientific field tends to revolve around a limited number of conceptual axes. New papers rarely introduce entirely new directions in idea space; instead, they typically refine or extend existing conceptual structures.

Dynamical Mechanism. The collapse of conceptual dimensionality can therefore be understood as a dynamical consequence of repeated interpretation. Each interpretation operator selects and reshapes conceptual directions inherited from prior work. Directions that align with widely used conceptual frameworks tend to be amplified, while others are gradually forgotten. Over time, this selective amplification produces a **hierarchy** of conceptual directions.

In later sections we will show that this hierarchy can be formally characterized using **Lyapunov exponents** associated with the products of interpretation operators. These exponents determine the long-term growth or decay rates of different conceptual directions, providing a mathematical description of the dimensional structure of scientific idea space.

Conceptual dimensionality collapse has several important implications for the organization of scientific knowledge:

1. **Stabilization of conceptual frameworks.**
   As dimensionality collapses, a field becomes organized around a small set of dominant conceptual axes.
2. **Predictability of research trajectories.**
   Because most conceptual directions are suppressed, future work tends to develop along established frameworks rather than exploring entirely new conceptual directions.
3. **Emergence of canonical ideas.**
   Certain conceptual directions become persistent across generations of research and serve as the intellectual backbone of the field.

In the next section we examine these persistent directions more closely and show how they give rise to **canonical conceptual modes** that structure the long-term evolution of scientific ideas.

3.3 Canonical Conceptual Modes

The dimensional collapse described in the previous section implies that only a limited number of conceptual directions remain persistent under repeated interpretation. These persistent directions form the **structural backbone of a scientific field**. We refer to them as **canonical conceptual modes**.

Definition: Canonical Conceptual Mode. Let $T_n = A_n A_{n-1} \cdots A_1$ denote the cumulative interpretation operator along a citation chain. Consider the singular value decomposition

$$T_n = U_n \Sigma_n V_n^\top,$$

where

$$\Sigma_n = diag(\sigma_1(T_n), \sigma_2(T_n), \dots, \sigma_d(T_n))$$

contains the singular values ordered by magnitude. As n increases, the ratios

$$\frac{\sigma_k(T_n)}{\sigma_1(T_n)}$$

typically decay exponentially for sufficiently large k. Consequently, only a small number of leading singular directions contribute significantly to the transported ideas. The corresponding vectors in the leading singular subspace define the **canonical conceptual modes** of the evolving scientific field.

A canonical conceptual mode represents a direction in idea space that remains stable under repeated reinterpretation across generations of research. When papers reinterpret the literature, they tend to preserve or amplify these directions while suppressing others. Conceptually, these modes correspond to the **core intellectual themes** around which scientific discourse organizes. Examples in various scientific domains include: **foundational theoretical principles, dominant methodological paradigms,** and **widely accepted explanatory frameworks.** Although individual papers introduce local variations, the underlying conceptual structure of the field tends to align with these persistent directions.

Stability Under Epistemic Transport. The stability of canonical conceptual modes arises from the multiplicative dynamics of interpretation operators. Suppose that a conceptual direction v is repeatedly amplified relative to other directions along citation chains. Then the transported contribution

$$T_n v$$

remains large even after many reinterpretations, whereas contributions along suppressed directions decay.

In the limit of long citation chains, the transported ideas become approximately aligned with the dominant singular vectors of T_n. Thus the conceptual content of many papers in a mature field becomes concentrated near a **small subspace spanned by the canonical modes**.

Canonical Modes and Scientific Narratives. From a sociological perspective, canonical conceptual modes correspond to the *simplified narratives through which scientific knowledge is communicated and taught*. Over time, complex bodies of research are distilled into a small number of core concepts or theoretical frameworks that structure both research and pedagogy. For example, Adam Smith and Karl Marx are often summarized through a few central ideas that become standard reference points in textbooks. Subsequent research typically elaborates on these ideas rather than exploring the full richness offered by their original multi-faceted books. Within the present framework, such narratives arise naturally from the compression dynamics of interpretation. Repeated reinterpretation filters out most conceptual variation while preserving a limited set of stable directions.

Mathematical Characterization. The formal identification of canonical conceptual modes requires analyzing the long-run spectral structure of the operator products T_n. As we will show in Section 4, the multiplicative ergodic theorem of Vladimir Oseledets guarantees that under broad conditions the space of ideas decomposes into invariant subspaces associated with distinct exponential growth rates. The canonical conceptual modes correspond to the subspace associated with the largest Lyapunov exponents of the epistemic transport process.

The emergence of canonical conceptual modes has several important implications for the structure of scientific knowledge:

1. **Conceptual convergence.**
   Mature scientific fields become organized around a small set of stable conceptual directions.
2. **Persistence of foundational ideas.**
   Certain conceptual contributions remain influential across long citation chains.
3. **Limited conceptual capacity.**
   Because only a finite number of directions remain stable, a field can sustain only a bounded number of dominant conceptual frameworks.

In the next section we extend this analysis by examining how the dimensional structure of idea space evolves during the life cycle of a scientific field, from periods of conceptual exploration to phases of mature conceptual compression.

3.4 Evolution of Conceptual Dimensionality

The previous sections showed that interpretation operators compress conceptual information and that repeated interpretation leads to the emergence of a small number of canonical conceptual modes. These results describe the **long-run structure of mature scientific fields**. However, the dimensional structure of idea space is not static. Instead, it evolves over the life cycle of a field. In this section we analyze the **dynamical trajectory of conceptual dimensionality**, showing that the number of active conceptual directions typically follows a characteristic evolutionary pattern.

Conceptual Exploration in Early Fields. In the early stages of a scientific field, conceptual space is typically **highly exploratory**. Researchers investigate many alternative theoretical directions, methods, and explanatory frameworks. As a result, the idea vectors of early papers tend to span a relatively large portion of the available conceptual space. Formally, if $\boldsymbol{e}_i$ denotes the idea vector of paper i, the covariance matrix

$$\Sigma_t = \mathbb{E}[\boldsymbol{e}_i \boldsymbol{e}_i^\top]$$

computed over papers published around time t tends to exhibit a relatively **broad eigenvalue spectrum** during the exploratory phase. This indicates that conceptual variance is distributed across many directions in idea space.

Emergence of Dominant Frameworks. As the field develops, certain conceptual directions prove more fruitful or explanatory than others. These directions begin to accumulate citations and influence subsequent work. Through repeated interpretation, ideas aligned with these directions are amplified, while others gradually lose prominence.

In terms of the covariance spectrum, this corresponds to a **progressive concentration of variance** in the leading eigenvalues of Σ_t. The effective conceptual dimension of the field therefore begins to decrease. This phase often coincides with the formation of recognizable **research programs or theoretical frameworks**, which organize the conceptual structure of the field.

Maturation and Dimensional Stabilization. In mature scientific fields, the process of dimensional collapse eventually stabilizes. Most papers align with a small number of canonical conceptual modes that structure the discourse of the field. At this stage, the effective conceptual dimension approaches a stable value

$$d^* \ll d.$$

This quantity can be interpreted as the **intrinsic conceptual dimension** of the field—the number of conceptual directions that remain stable under repeated interpretation. Importantly, this intrinsic dimension is not determined by the embedding method used to represent ideas. Instead, it arises from the **dynamical properties of epistemic transport** within the citation network.

Sigmoid Trajectory of Conceptual Diversity. Combining the phases described above suggests a characteristic trajectory for conceptual dimensionality during the life cycle of a scientific field:

- *Exploration phase* – conceptual dimensionality grows as new directions are explored.
- *Selection phase* – competing conceptual directions are filtered through interpretation and citation dynamics.
- *Compression phase* – the field converges toward a small set of canonical conceptual modes.

This trajectory produces a **sigmoid-shaped evolution** of conceptual diversity: an initial expansion followed by stabilization and compression (see derivations in Appendix E).

Dynamical Interpretation. The mechanism underlying this trajectory is the interaction between two opposing forces: *innovation,* which introduces new conceptual directions into idea space, and *interpretation compression,* which selectively amplifies certain directions while suppressing others. When innovation dominates, conceptual dimensionality expands. When compression dominates, dimensionality collapses toward the stable subspace determined by the spectral structure of interpretation operators.

Implications for Scientific Development. The evolution of conceptual dimensionality has important implications for understanding the dynamics of scientific fields:

- **Early-stage fields exhibit conceptual diversity**, reflecting exploration of multiple theoretical possibilities.
- **Mature fields exhibit conceptual concentration**, organized around stable frameworks.
- **Scientific revolutions** may occur when new conceptual directions suddenly become amplified, reorganizing the dominant subspace of ideas.

In the following sections we develop the mathematical tools needed to analyze these dynamics more formally. In particular, we show that the long-run behavior of epistemic transport can be characterized using spectral theory and Lyapunov exponents associated with products of interpretation operators.

4. Spectral Dynamics of Epistemic Transport

The previous sections introduced a conceptual framework in which scientific ideas propagate through citation networks via interpretation operators. We now develop the **spectral theory of epistemic transport**, which provides the mathematical foundation for analyzing the long-run evolution of conceptual influence.

The key observation is that conceptual contributions propagate through **products of interpretation operators** along citation chains. The asymptotic behavior of these operator products determines which conceptual directions persist across generations of scientific work.

4.1 Transport Along Citation Chains

We begin by formalizing how conceptual contributions propagate along a single citation chain.

Citation Chains. Consider a sequence of papers connected through citations

$$k \to i_1 \to i_2 \to \cdots \to i_m \to F.$$

Here paper k is an ancestor and F is a descendant paper reached after $m + 1$ interpretation steps. Each citing paper i applies an interpretation operator A_i that transforms inherited conceptual directions. As a result, conceptual contributions from ancestor k are repeatedly transformed as they move through the chain.

Transport Operator Along a Chain. Let $T_{k \to F}$ denote the cumulative transport operator that maps conceptual contributions from ancestor k to descendant F. Along the citation chain above, this operator is given by the product

$$T_{k \to F} = A_F A_{i_m} \cdots A_{i_1}.$$

This operator describes how a conceptual perturbation introduced by paper k is transformed through successive reinterpretations before appearing in paper F.

Propagation of Conceptual Contributions. Let $v_k \in \mathbb{R}^d$ represent a conceptual contribution introduced by paper k. After propagation through the citation chain, the contribution appearing in paper F becomes

$$v_F = T_{k \to F} v_k.$$

Thus the cumulative operator $T_{k \to F}$ determines both the **magnitude** and **direction** of conceptual influence transmitted from ancestor to descendant.

Norm Growth and Influence. To measure the strength of conceptual influence, we consider the norm of the transported contribution:

$$\| v_F \| = \| T_{k \to F} v_k \|.$$

The growth or decay of this norm as the chain length increases determines whether a conceptual direction remains influential across generations of scientific work.

If the operator product expands a direction v_k, the corresponding idea becomes increasingly prominent. If it contracts the direction, the idea gradually disappears from the literature.

Citation Distance. Let $\ell(k, F)$ denote the number of interpretation steps separating ancestor k from descendant F. The cumulative transport operator therefore depends on this citation distance:

$$T_{k \to F} = A_\ell A_{\ell-1} \cdots A_1.$$

As the citation distance grows, the statistical properties of these operator products begin to dominate the behavior of conceptual transport.

Asymptotic Regime. For long citation chains, the behavior of the operator product $T_{k \to F}$ becomes increasingly regular. Rather than depending on the details of individual interpretation steps, the growth rates of conceptual contributions converge toward characteristic exponential rates.

These rates are described by **Lyapunov exponents** associated with the random sequence of interpretation operators. The mathematical theorem that guarantees the existence of these exponents is the **multiplicative ergodic theorem** of Vladimir Oseledets (1968). This result implies that the idea space decomposes into subspaces associated with different exponential growth rates of conceptual influence. In the next section we introduce the stochastic interpretation dynamics that allow this theorem to be applied to citation networks.

4.2 Random Interpretation Dynamics

The transport of ideas along citation chains depends on the sequence of interpretation operators applied by successive papers. In practice, these operators vary widely because different authors reinterpret prior work in different ways. To analyze the long-run behavior of epistemic transport, it is therefore natural to model interpretation as a **stochastic process**. This stochastic formulation allows us to apply powerful results from the theory of random matrix products to characterize the asymptotic dynamics of conceptual influence.

Random Sequence of Interpretation Operators. Let

$$\{A_t\}_{t=1}^{\infty}$$

be a sequence of interpretation operators corresponding to successive citation steps along a chain. Each operator

$$A_t \in \mathbb{R}^{d \times d}$$

transforms conceptual directions inherited from the previous paper. Because interpretation varies across papers, we treat these operators as random matrices drawn from some distribution over linear transformations of idea space.

Stationary Interpretation Process. For analytical tractability, we assume that the sequence of interpretation operators forms a *stationary stochastic process*. That is, the statistical properties of interpretation do not change systematically along long citation chains. Formally, we assume the matrices A_t are drawn from a probability distribution μ on the space of $d \times d$ matrices satisfying

$$A_t \sim \mu.$$

This assumption does not imply that every paper interprets ideas identically. Rather, it states that the **statistical structure** of interpretation remains stable across the literature.

Products of Random Interpretation Operators. Under this stochastic model, epistemic transport along a citation chain of length n is governed by the random matrix product

$$T_n = A_n A_{n-1} \cdots A_1.$$

The properties of this product determine the long-run behavior of conceptual contributions transmitted through the citation network. A central quantity of interest is the growth rate of transported vectors: $\| T_n v \|$. For large n, this quantity typically grows or decays exponentially depending on the direction of the initial vector v.

Lyapunov Growth Rates. The exponential growth rate of a conceptual direction v is defined by

$$\lambda(v) = \lim_{n \to \infty} \frac{1}{n} \log \| T_n v \|.$$

These growth rates characterize how conceptual contributions evolve under repeated reinterpretation. Some directions in idea space may exhibit positive growth rates, meaning their influence tends to amplify along citation chains. Other directions exhibit negative growth rates and therefore decay over time.

Spectral Hierarchy of Conceptual Directions. The set of possible growth rates forms a finite sequence

$$\lambda_1 > \lambda_2 > \cdots > \lambda_r.$$

Each exponent corresponds to a subspace of conceptual directions that share the same asymptotic growth rate. This hierarchy describes the *spectral structure of epistemic transport*: directions associated with larger exponents dominate long-run conceptual influence; directions associated with smaller exponents gradually disappear from the literature.

In the next section we formalize this structure using the multiplicative ergodic theorem, which provides a rigorous decomposition of idea space into invariant subspaces associated with these Lyapunov growth rates.

4.3 Application of the Multiplicative Ergodic Theorem

Operator Products in Epistemic Transport. Recall that conceptual contributions propagate along citation chains through products of interpretation operators:

$$T_n = A_n A_{n-1} \cdots A_1.$$

Here the matrices A_t represent interpretation steps applied by successive papers. Under the stochastic model described in Section 4.2, the sequence $\{A_t\}$ forms a stationary random process. The key question is how the norm of transported vectors$|T_n v|$behaves as the length of the citation chain n becomes large.

Lyapunov Exponents. The multiplicative ergodic theorem guarantees that, under broad regularity conditions, the exponential growth rates of such products converge to well-defined constants called Lyapunov exponents. Formally, there exists a sequence of numbers

$$\lambda_1 > \lambda_2 > \cdots > \lambda_r$$

such that for almost every realization of the random operator sequence,

$$\lim_{n\to\infty} \frac{1}{n} \log|T_n v| = \lambda_i$$

whenever v lies in a certain associated subspace. Each exponent therefore represents the asymptotic growth or decay rate of conceptual directions belonging to the corresponding subspace.

Decomposition of Idea Space. A crucial implication of the multiplicative ergodic theorem is that the idea space $\mathbb{R}^d$ decomposes into invariant subspaces associated with the different Lyapunov exponents. That is, there exists a filtration of subspaces

$$\mathbb{R}^d = V_1 \supset V_2 \supset \cdots \supset V_r$$

such that vectors in $(V_i \setminus V_{i+1})$ exhibit exponential growth rate λ_i under repeated interpretation. This decomposition partitions idea space into directions that are amplified at different exponential rates by epistemic transport.

Interpretation for Scientific Ideas. Within the context of scientific knowledge evolution, the Lyapunov spectrum provides a natural interpretation:

- **Positive or large exponents** correspond to conceptual directions that tend to be amplified through reinterpretation and citation.
- **Negative exponents** correspond to directions that decay and gradually disappear from scientific discourse.

- **Neutral exponents** correspond to directions that remain approximately stable under repeated interpretation.

Thus the multiplicative ergodic theorem provides a mathematical mechanism for understanding why certain ideas persist across generations of research while others fade away.

Dominant Conceptual Subspace. Over long citation chains, contributions from directions associated with smaller exponents become negligible relative to those associated with the largest exponent. As a result, transported conceptual vectors tend to align with the subspace corresponding to the leading Lyapunov exponent.

This phenomenon provides the mathematical foundation for the **canonical conceptual modes** discussed in Section 3.3: these modes correspond to the dominant directions selected by the long-run dynamics of epistemic transport.

Toward a Spectral Theory of Scientific Ideas. The multiplicative ergodic theorem therefore provides the key theoretical bridge between the **local interpretation of ideas** by individual papers, and the **global structure of scientific knowledge** observed across the citation network. By analyzing the Lyapunov spectrum of interpretation operators, we can characterize the long-run dimensional structure of idea space and the persistence of conceptual directions.

In the next section we examine this structure more closely by introducing the **Lyapunov subspace decomposition**, which will play a central role in understanding conserved epistemic quantities and the dynamics of scientific credit.

4.4 Lyapunov Subspace Decomposition

The multiplicative ergodic theorem introduced in the previous section guarantees that long products of interpretation operators exhibit well-defined exponential growth rates. Beyond the existence of these growth rates, the theorem also provides a **geometric decomposition of idea space** into invariant subspaces associated with different Lyapunov exponents. This structure plays a central role in understanding the long-run dynamics of epistemic transport.

Oseledec Splitting of Idea Space. Under the conditions of the multiplicative ergodic theorem, the idea space $\mathbb{R}^d$ admits a direct-sum decomposition

$$\mathbb{R}^d = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_r},$$

where each subspace E_{λ_i} corresponds to a Lyapunov exponent λ_i. Vectors in the subspace E_{λ_i} grow or decay at the exponential rate λ_i under repeated interpretation:

$$|T_n v| \approx e^{\lambda_i n} |v| for v \in E_{\lambda_i}$$

This decomposition is often referred to as the **Oseledec splitting (**Oseledets, 1968).

Interpretation for Epistemic Transport. Within the context of scientific knowledge evolution, the Lyapunov subspaces correspond to distinct categories of conceptual directions:

1. *Expanding directions* ($\lambda_i > 0$)
   These conceptual directions become increasingly influential along citation chains.
2. *Contracting directions* ($\lambda_i < 0$)
   Contributions along these directions decay exponentially and gradually disappear from scientific discourse.
3. *Neutral directions* ($\lambda_i \approx 0$)
   These directions remain approximately conserved under repeated interpretation.

The existence of these subspaces implies that the dynamics of scientific ideas are highly structured rather than arbitrary.

Alignment of Conceptual Contributions. Consider an arbitrary conceptual vector v introduced by an early paper. Using the Lyapunov decomposition, we can write

$$v = v_1 + v_2 + \cdots + v_r$$

where $v_i \in E_{\lambda_i}$. After many interpretation steps,

$$T_n v = T_n v_1 + T_n v_2 + \cdots + T_n v_r.$$

Because the components grow at different exponential rates, the contribution from the largest exponent eventually dominates. Thus long citation chains tend to align conceptual influence with the **leading Lyapunov subspace**. This mechanism explains why the conceptual structure of mature scientific fields often becomes concentrated around a small number of dominant directions.

Neutral Subspace and Conceptual Conservation. Of particular importance is the *neutral Lyapunov subspace* corresponding to exponents near zero. Conceptual directions in this subspace neither grow nor decay significantly under repeated interpretation.These directions therefore represent structurally conserved components of scientific ideas. Contributions aligned with this subspace can persist across long citation chains without amplification or attenuation. In later sections we will show that the dimension of this neutral subspace determines the number of **conserved epistemic quantities** in the evolution of scientific knowledge.

Spectral Structure of Scientific Fields. The Lyapunov subspace decomposition provides a compact spectral description of epistemic transport. Rather than tracking the detailed history of individual citations, the long-run behavior of ideas can be summarized by the Lyapunov spectrum $\lambda_1, \lambda_2, \ldots, \lambda_r$, and the associated invariant subspaces E_{λ_i}. Together, these objects characterize the fundamental dynamical structure governing the propagation of scientific ideas.

In the next section we explore how the neutral subspace gives rise to **conserved epistemic quantities** that structure the long-run evolution of scientific ideas and influence the distribution of intellectual credit.

5. Conserved Epistemic Structures

The spectral decomposition developed in the previous section reveals that the evolution of scientific ideas is governed by a structured hierarchy of conceptual directions with different exponential growth rates. Among these directions, the **neutral Lyapunov components** play a special role: they neither grow nor decay under repeated interpretation. This section shows that these neutral directions correspond to **conserved epistemic structures**—quantities that remain approximately invariant as ideas propagate through citation networks. Such conservation laws provide a new perspective on the stability of scientific knowledge.

5.1 Conserved Epistemic Quantities

The Lyapunov subspace decomposition introduced in Section 4 implies that idea space contains directions whose contributions remain approximately unchanged under repeated interpretation. These directions give rise to quantities that are conserved along citation chains.

Definition. Let $T_n = A_n A_{n-1} \cdots A_1$ denote the cumulative interpretation operator along a citation chain of length n. A linear functional

$$Q(v) = w^\top v$$

is called a **conserved epistemic quantity** if

$$Q(T_n v) \approx Q(v) \textit{for large} n$$

In other words, the quantity measured by Q remains approximately invariant as the idea vector v propagates through successive reinterpretations.

Relation to the Neutral Lyapunov Subspace. Conserved epistemic quantities arise from the neutral Lyapunov subspace identified in the spectral decomposition of epistemic transport. Let

$$E_0$$

denote the subspace associated with Lyapunov exponent $\lambda = 0$. Vectors lying in this subspace satisfy

$$|T_n v| \approx |v| \textit{for large} n.$$

Thus conceptual components aligned with E_0 remain stable under repeated interpretation. Linear functionals that measure these components therefore define conserved epistemic quantities.

In the context of scientific discourse, conserved epistemic quantities correspond to **core conceptual invariants** that persist across generations of research. These invariants represent intellectual structures that remain recognizable even as ideas are repeatedly reinterpreted. Examples of such structures include fundamental theoretical principles, stable conceptual distinctions within a field, or methodological frameworks that persist across research programs. Although the language used to describe these ideas may change, their underlying conceptual structure remains approximately preserved.

Dimensionality of Conserved Quantities

An important consequence of the spectral theory developed earlier is that the number of independent conserved epistemic quantities is determined by the dimension of the neutral Lyapunov subspace. If

$$dim(E_0) = d^*,$$

then the epistemic transport process admits exactly d^* independent conserved quantities. This result provides a direct connection between the **spectral dynamics of interpretation operators** and the **structural stability of scientific ideas**.

Conceptual Significance. The existence of conserved epistemic quantities suggests that scientific knowledge evolves under constraints analogous to conservation laws in physical systems. While individual ideas may be amplified, attenuated, or recombined, certain conceptual structures remain invariant across the literature. These conserved structures provide the **scaffolding that organizes scientific knowledge**, ensuring continuity across generations of research. In the next section we examine the neutral Lyapunov subspace itself in greater detail and analyze how it governs the persistence of conceptual frameworks within mature scientific fields.

5.2 Neutral Lyapunov Subspace

The conserved epistemic quantities introduced in the previous section arise from a specific component of the spectral decomposition of epistemic transport: the **neutral Lyapunov subspace**. This subspace contains conceptual directions whose magnitude remains approximately stable under repeated interpretation. As we show below, the structure of this subspace plays a central role in determining the long-run organization of scientific knowledge.

Definition. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ denote the Lyapunov exponents associated with the sequence of interpretation operators, and let E_{λ_i} be the corresponding invariant subspaces introduced in Section 4.4. The **neutral Lyapunov subspace** is defined as

$$E_0 = v \in \mathbb{R}^d : \lambda(v) = 0.$$

Vectors in this subspace exhibit neither exponential amplification nor exponential decay under repeated interpretation.

Dynamical Behavior. For vectors $v \in E_0$, the cumulative transport operator satisfies $|T_n v| \approx |v| for\, large n.$Thus conceptual contributions aligned with the neutral subspace persist across arbitrarily long citation chains without systematic amplification or attenuation. In contrast, vectors in expanding subspaces grow exponentially; vectors in contracting subspaces decay exponentially. Over long citation distances, most components of an idea vector therefore vanish or explode relative to the neutral components. As a result, the neutral subspace often provides the stable conceptual backbone of the field.

Dimension of the Neutral Subspace. Let $d^* = dim(E_0)$ denote the dimension of the neutral Lyapunov subspace. This quantity has a direct epistemic interpretation: it represents the number of independent conceptual directions that remain stable under the dynamics of epistemic transport. Equivalently, d^* determines the number of conserved epistemic quantities discussed in Section 5.1.

The intrinsic conceptual dimensionality of a mature scientific field is therefore determined not by the ambient dimension of the embedding space but by the dimension of the neutral Lyapunov subspace. In other words, even if idea vectors are embedded in a high-dimensional representation, the effective conceptual structure of the field is governed by the much smaller quantity d^*. This observation provides a theoretical explanation for the empirical phenomenon discussed earlier in Section 3: mature scientific fields often appear to be organized around a **small number of dominant conceptual frameworks**.

Stability of Scientific Frameworks. Because conceptual directions in the neutral subspace are neither amplified nor suppressed, they can persist across long citation chains with relatively little distortion. As a result, these directions tend to correspond to the stable conceptual frameworks that organize scientific discourse. Such frameworks provide the reference structure within which new ideas are interpreted and evaluated.

In the next section we examine how these neutral conceptual directions give rise to **canonical conceptual structures**—the stable intellectual frameworks that characterize mature scientific paradigms.

5.3 Canonical Conceptual Structures

Definition. Let E_0 denote the neutral Lyapunov subspace introduced in Section 5.2. A **canonical conceptual structure** is defined as a dominant conceptual direction within this subspace that remains stable under the dynamics of epistemic transport. Formally, a vector $u \in E_0$ represents a canonical conceptual structure if it satisfies

$$T_n u \approx u \, for \, large n.$$

Thus repeated interpretation preserves the conceptual meaning encoded in this direction.

Persistence Across Citation Cascades. Because the neutral Lyapunov subspace is stable under the transport operators, conceptual contributions aligned with these directions propagate through citation chains without systematic distortion. If a conceptual contribution v contains a component along a canonical direction u, we may write

$$v = \alpha u + v_{\perp},$$

where $v_{\perp}$ lies outside the neutral subspace. Under repeated interpretation, the non-neutral components typically decay or expand relative to the neutral component. Consequently, the long-run conceptual content of the idea becomes increasingly aligned with the canonical structure u. This process gradually filters the conceptual diversity of the literature, leaving behind a stable set of dominant conceptual frameworks.

Emergence of Canonical Frameworks. The existence of canonical conceptual structures explains a widely observed feature of mature scientific fields: despite large numbers of publications and theoretical variations, the field typically converges toward a small number of stable conceptual narratives, organizing empirical findings, providing shared terminology, and defining the central problems and methods of the field. Within our framework, such narratives correspond to **low-dimensional structures embedded within the neutral Lyapunov subspace**.

Bounded Number of Canonical Structures. An immediate implication of the spectral theory developed earlier is that the number of canonical conceptual structures is bounded by the dimension of the neutral subspace:

$$N_{canonical.Str} \leq d^*.$$

Because $d^* = dim(E_0)$, this bound implies that a scientific field can only sustain a limited number of stable conceptual frameworks.

This result provides a theoretical explanation for an empirical regularity of scientific discourse: although the number of published papers may grow indefinitely, the number of **foundational conceptual narratives** typically remains small. In a highly mature or strongly consolidated scientific paradigm, where $d^* = 1$, a single canonical narrative exists, dominates and persists.

Relation to Scientific Paradigms. Canonical conceptual structures provide a natural mathematical interpretation of the idea of *scientific paradigms*, as described in the philosophy of science literature. In particular, paradigms correspond to stable conceptual directions that organize the interpretation of new research. Within the present framework, such structures arise naturally from the spectral dynamics of epistemic transport rather than being imposed externally.

Transition to Credit Dynamics. The canonical conceptual structures identified in this section not only organize scientific ideas but also shape the distribution of intellectual credit across the citation network. Because conceptual contributions align with a small number of dominant directions, influence tends to concentrate around a limited set of ancestral works. In the next section we examine how these spectral dynamics govern the **propagation of epistemic credit** along citation cascades and lead to the emergence of highly unequal credit distributions in scientific fields.

5.4 Bounded Canonical Ancestors

The existence of canonical conceptual structures has an important implication for the structure of citation networks. If scientific ideas ultimately align with a small number of stable conceptual directions, then only a limited number of ancestral works can meaningfully contribute to the conceptual foundation of a new paper. This leads to the principle of **bounded canonical ancestors**.

Conceptual Contribution from Ancestors. Consider a paper whose idea vector is v. Through the citation network, this idea inherits conceptual contributions from a set of ancestral papers. Let

$$v = \sum_{i=1}^{N} \alpha_i a_i,$$

where a_i represents the conceptual contribution inherited from ancestor i, and α_i measures its influence. While a paper may cite many predecessors, only a subset of these ancestors contributes meaningfully to the conceptual structure of the resulting idea.

Spectral Filtering of Ancestral Contributions. Under repeated epistemic transport, contributions along contracting Lyapunov directions decay exponentially. If an ancestral idea contains components outside the neutral subspace, those components vanish as the idea propagates through successive reinterpretations. Thus, for large citation distance n,

$$T_n a_i \approx P_{E_0}(a_i),$$

where P_{E_0} denotes projection onto the neutral Lyapunov subspace. This process acts as a **spectral filter**, eliminating conceptual components that do not align with the canonical structures of the field.

Canonical Ancestors. We define a canonical ancestor as an ancestral paper whose conceptual contribution has a non-negligible projection onto the neutral Lyapunov subspace. Formally, ancestor i is canonical if

$$|P_{E_0}(a_i)| > 0.$$

Such ancestors contribute to the stable conceptual structure that survives across long citation chains.

Bounded Number of Canonical Ancestors. Because the neutral subspace has dimension d^*, the number of independent conceptual directions that can persist is limited. Consequently, the number of canonical ancestors that meaningfully contribute to a paper is also bounded.
This yields the following structural result. Bounded Canonical Ancestor Principle:

$$N_{canonica.Anc} \leq d^*.$$

Even if a paper cites dozens or hundreds of predecessors, the number of ancestors that contribute independent conceptual directions cannot exceed the intrinsic epistemic dimension d^*.

This result explains an empirical pattern widely observed in scientific writing. Although reference lists may be long, authors typically rely on only a small number of **foundational works** when framing their conceptual contributions. These works function as canonical ancestors that define the intellectual lineage of the research. Within our framework, this phenomenon arises naturally from the spectral filtering properties of epistemic transport.

Consequences for Citation Networks. The bounded canonical ancestor principle has several implications: Citation networks may be large and dense, but conceptual influence concentrates around a small number of foundational works; The effective intellectual ancestry of a paper is much smaller than its formal citation list; Mature fields tend to develop stable genealogies of canonical works. These patterns emerge directly from the low-dimensional structure of the neutral Lyapunov subspace.

Bridge to Credit Dynamics. Because canonical ancestors occupy the conceptual directions that persist across citation cascades, they tend to accumulate a disproportionate share of epistemic credit. In other words, the spectral structure of epistemic transport naturally leads to **credit concentration** around a small number of foundational papers.

The next section formalizes this idea by introducing a dynamical theory of **credit propagation along citation cascades**, showing how spectral growth rates determine the distribution of intellectual credit across scientific lineages.

6. Credit Dynamics Along Citation Cascades

The spectral structure of epistemic transport developed in the previous sections has direct implications for how intellectual credit propagates through citation networks. As ideas are repeatedly interpreted and transmitted across generations of papers, conceptual influence is progressively filtered by the Lyapunov structure of the transport process.

This section develops a dynamical theory of **epistemic credit propagation** along citation cascades. We begin by establishing a fundamental property of citation influence: **credit decays exponentially with citation distance**.

6.1 Influence Decay

When a paper contributes an idea to the literature, its conceptual influence propagates through a chain of subsequent works that reinterpret, extend, or apply that idea. However, the spectral dynamics of epistemic transport imply that this influence cannot persist indefinitely. Instead, it decays systematically with the length of the citation chain.

Citation Chains as Transport Processes. Consider a citation chain

$$P_0 \to P_1 \to P_2 \to \cdots \to P_n$$

in which each paper interprets the ideas of its predecessors. If the conceptual state of paper P_k is represented by a vector v_k, then successive interpretations generate the relation

$$v_{k+1} = A_{k+1} v_k,$$

where A_{k+1} is the interpretation operator introduced in Section 2.

The conceptual state of the descendant paper P_n relative to the original idea v_0 is therefore

$$v_n = T_n v_0, T_n = A_n A_{n-1} \cdots A_1.$$

Thus the propagation of conceptual influence along citation chains is governed by the same multiplicative dynamics studied in Section 4.

Spectral Attenuation of Influence. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ denote the Lyapunov exponents associated with the interpretation process. Decomposing the initial idea v_0 into Lyapunov components yields

$$v_0 = \sum_i v_i, v_i \in E_{\lambda_i}.$$

Under repeated transport, the contribution of each component evolves as

$$|T_n v_i| \approx e^{\lambda_i n} |v_i|.$$

If $\lambda_i < 0$, the corresponding conceptual contribution decays exponentially with citation distance.

Influence Decay Law. Because most conceptual components lie in contracting Lyapunov directions, the overall influence of an ancestor paper decreases exponentially as the citation chain length increases. This leads to the following fundamental result:

The expected epistemic influence of an ancestor at citation distance n satisfies

$$I(n) \propto e^{-\gamma n},$$

where $\gamma = |\lambda| > 0$ is the absolute value of the negative Lyapunov exponents of the epistemic transport process. Thus conceptual influence decays exponentially with citation distance.

Credit Half-Life. The exponential decay law implies the existence of a characteristic credit half-life for scientific ideas. Let $n_{1/2}$ denote the citation distance at which influence falls to half its original magnitude:

$$I(n_{1/2}) = \frac{1}{2} I(0).$$

Solving for $n_{1/2}$ yields

$$n_{1/2} = \frac{log2}{\gamma}.$$

The credit half-life therefore provides a natural measure of how rapidly conceptual influence dissipates within a scientific field.

The influence decay law leads to a clear empirical prediction: the epistemic credit attributed to an ancestor should decline approximately exponentially with its distance in the citation network. This prediction is consistent with observed patterns in citation analysis, where distant ancestors tend to receive progressively smaller attribution in intellectual lineages. Moreover, variation in the decay parameter γ across fields may help explain differences in citation inequality and the persistence of foundational works.

While influence decays with citation distance, the **multiplicative structure** of epistemic transport also allows credit to accumulate along branching citation cascades. These two forces—**exponential decay along chains** and **multiplicative amplification across branches**—jointly determine the distribution of scientific credit. In the next section we formalize this process by developing a model of **multiplicative propagation of epistemic credit** through citation networks.

6.2 Multiplicative Propagation of Credit

The previous section established that conceptual influence decays exponentially with citation distance. However, citation networks do not consist of single linear chains. Scientific ideas propagate through **branching cascades**, where each paper may generate multiple descendants. As a result, epistemic credit evolves through a combination of **exponential attenuation along chains** and **branching multiplication across the network**.

In this section we show that these dynamics lead naturally to multiplicative credit propagation and concentration of influence among a small set of foundational works.

Credit Flow Through Citation Networks. Consider a paper P_0 whose conceptual contribution propagates through the citation network. Let b denote the average number of descendant papers generated by each paper in the cascade. After n generations of citations, the expected number of descendants is approximately

$$N(n) \approx b^n.$$

Each descendant inherits a fraction of the conceptual influence of the original work, determined by the epistemic transport process.

Combined Growth–Decay Dynamics. From Section 6.1, the influence transmitted along a citation chain of length n decays as

$$I(n) \propto e^{-\gamma n}.$$

Meanwhile, the number of descendant papers grows approximately as b^n. The total influence of the original paper across the entire citation cascade is therefore proportional to

$$C(n) \propto b^n e^{-\gamma n}.$$

This expression captures the central tension in credit dynamics: branching growth tends to amplify influence, while spectral attenuation suppresses it.

Spectral Interpretation. The multiplicative structure of epistemic transport implies that conceptual contributions propagate through repeated application of random interpretation operators. As shown in Section 4, the dominant growth rate of this process is governed by the largest Lyapunov exponent λ.

Consequently, the effective amplification of conceptual influence along the citation cascade follows

$$|T_n v| \approx e^{\lambda n}|v|.$$

The largest Lyapunov exponent therefore determines the intrinsic **growth rate of epistemic influence** under repeated reinterpretation.

Combining the branching growth rate b with the spectral growth rate λ yields a fundamental result governing the distribution of scientific credit.

Credit Concentration Theorem. If the effective growth rate of epistemic transport exceeds the branching rate of the citation network, influence becomes concentrated among a small number of ancestral works. More precisely, the total credit propagated from an ancestor grows approximately as

$$C(n) \propto e^{(\lambda+logb)n}.$$

When this exponent is positive, small differences in early conceptual contributions are exponentially amplified across the citation cascade.

Emergence of Superstar Papers. The multiplicative propagation of credit provides a natural explanation for the emergence of superstar papers in scientific fields. Even modest differences in the conceptual alignment of early papers with dominant Lyapunov directions can produce exponentially different influence trajectories. As the citation cascade unfolds, this multiplicative amplification leads to large disparities in epistemic credit between different ancestors.

Interaction with Spectral Filtering. The canonical conceptual structures identified in Section 5 further intensify this effect. Because conceptual influence is filtered through a low-dimensional neutral subspace, only ancestors aligned with these stable directions can accumulate substantial long-run credit. As a result, epistemic credit tends to concentrate among the *canonical ancestors* that define the dominant conceptual frameworks of the field.

Bridge to Heavy-Tailed Credit Distributions. The combination of branching growth and multiplicative amplification has an important statistical consequence. Processes governed by repeated multiplicative dynamics typically produce heavy-tailed distributions, where a small number of elements account for a large fraction of the total quantity. In the context of scientific credit, this implies that influence across ancestors should follow a heavy-tailed distribution. In the next section we formalize this intuition and derive the **heavy-tailed distribution** of epistemic credit that emerges from multiplicative citation dynamics.

6.3 Emergence of Heavy-Tailed Credit Distributions

Sections 6.1 and 6.2 established two key mechanisms governing the propagation of scientific credit. First, conceptual influence decays exponentially along citation chains. Second, citation cascades grow multiplicatively through branching descendants. When these two processes interact across many generations, they generate a characteristic statistical pattern: **heavy-tailed distributions of epistemic credit**.

This section shows that the multiplicative dynamics of epistemic transport naturally produce power-law–like distributions of ancestor credit.

Credit as a Multiplicative Process. Consider an ancestral paper P_0. Each citation step transmits a fraction of the conceptual influence of the original idea. Let w_k denote the random fraction of credit passed from generation $k-1$ to generation k. Along a citation chain of length n, the credit transmitted to a descendant is therefore

$$C_n = C_0 \prod_{k=1}^{n} w_k.$$

This multiplicative structure is a direct consequence of repeated epistemic transport: each interpretation rescales the conceptual contribution inherited from previous work.

Logarithmic Growth Dynamics. Taking the logarithm of the multiplicative process yields

$$logC_n = logC_0 + \sum_{k=1}^{n} logw_k.$$

If the weights w_k are random variables generated by the interpretation process, then the logarithm of credit evolves as a sum of random increments. By standard limit results, this implies that credit across citation chains tends toward **log-normal variability**. However, when combined with branching citation growth, the resulting distribution becomes substantially heavier-tailed.

Branching Citation Cascades. Let b denote the average number of descendants produced by each paper. After n citation generations, the number of descendant paths grows approximately as

$$N(n) \approx b^n.$$

Each path contributes a multiplicative credit factor determined by the product of interpretation weights along that path. The total credit attributed to the ancestor is therefore the sum of contributions from a large number of multiplicative paths.

The interaction between multiplicative transport and branching growth leads to a characteristic statistical law.

Paradigm Credit Theorem. Under multiplicative credit propagation on a branching citation network, the distribution of ancestor credit approaches a heavy-tailed form with approximate power-law scaling

$$P(C > x) \propto x^{-\alpha}.$$

The exponent α is determined by the balance between the spectral growth rate of epistemic transport and the branching structure of the citation network.

Spectral Determination of the Power-Law Exponent. Within the spectral framework developed earlier, the growth of conceptual influence is governed by the largest Lyapunov exponent λ. Combining this with the branching rate b yields a theoretical prediction for the credit distribution exponent (see Appendix D):

$$\alpha = 1 + \frac{logb}{\lambda}.$$

This relation provides a direct bridge between **bibliometric patterns** and the spectral dynamics of **epistemic transport**.

The heavy-tailed distribution of scientific credit arises naturally from the multiplicative structure of idea propagation. Most papers receive relatively small credit contributions, while a small number of foundational works accumulate disproportionately large influence. Importantly, this inequality is not merely a sociological phenomenon but a mathematical consequence of the dynamics governing how ideas propagate and transform across the scientific literature.

The **bounded canonical ancestor** principle established in Section 5 further sharpens this result. Because only conceptual contributions aligned with the neutral Lyapunov subspace persist over long citation distances, heavy-tailed credit distributions tend to concentrate around a small set of canonical works. These works form the intellectual backbone of the field, serving as the primary ancestors through which conceptual influence flows.

The emergence of heavy-tailed credit distributions is closely related to the formation of dominant **scientific paradigms**. As certain conceptual directions accumulate increasing influence, they begin to structure the intellectual landscape of the field.

In the next section we examine how these spectral dynamics can generate **paradigm formation and phase transitions in scientific fields**, linking the mathematical framework developed here to broader theories of scientific change.

7. Paradigm Formation and Phase Transitions in Science

7.1 Compression and Paradigm Formation

The preceding sections developed a mathematical description of how ideas propagate, compress, and accumulate credit across citation networks. We now turn to the broader structural implications of these dynamics for the organization of scientific knowledge. In particular, the spectral compression of idea space provides a natural mechanism for the emergence of **scientific paradigms**.

Paradigms as Low-Dimensional Conceptual Structures. Recall from Section 3 that repeated interpretation of ideas induces *conceptual compression*: the effective dimensionality of idea space contracts as epistemic transport amplifies some conceptual directions while suppressing others. Over long citation chains, this process causes most conceptual variation to collapse into a small number of dominant directions corresponding to the leading Lyapunov subspaces.

Let the vector representation of a scientific idea be $v \in \mathbb{R}^d$. Under repeated interpretation,

$$v_n = T_n v_0,$$

where T_n denotes the product of interpretation operators along a citation chain. The Multiplicative Ergodic Theorem implies that the asymptotic behavior of v_n is governed by a small number of dominant spectral directions. As a result, the effective idea space of a mature scientific field becomes concentrated within a low-dimensional conceptual manifold spanned by the dominant Lyapunov vectors.

Conceptual Modes and Paradigm Structure. Let $e_1, \ldots, e_k$ denote the basis vectors associated with the dominant Lyapunov subspace. Then most ideas in the field can be approximated as

$$v \approx \sum_{i=1}^{k} a_i e_i.$$

The coefficients a_i describe how a given scientific contribution combines these fundamental conceptual directions. These dominant modes form what may be interpreted as the **conceptual backbone of the field**: a small set of stable directions that organize how new ideas are interpreted and developed.

Mathematical Interpretation of Paradigms. Within this framework, a scientific paradigm corresponds to the emergence of a **stable low-dimensional conceptual structure** in idea space. Once compression has reduced the effective dimensionality of the field, most subsequent research can be expressed as variations within the span of these dominant modes. In other words, a paradigm is not merely a sociological phenomenon but a structural property of the **spectral geometry of epistemic transport**.

This interpretation provides a quantitative realization of the notion of paradigms introduced by Thomas Kuhn. In *The Structure of Scientific Revolutions*, Kuhn argued that scientific progress occurs within periods of **normal science**, during which researchers work within a shared conceptual framework that defines legitimate questions, methods, and interpretations.

In the present theory, such periods correspond to regimes in which idea space has already undergone substantial compression. The dominant Lyapunov directions define the conceptual coordinates of the paradigm, and most new contributions are interpreted within this established structure. Researchers working within the paradigm therefore produce ideas that lie close to the same low-dimensional conceptual manifold.

Stability of Paradigms. The spectral structure of epistemic transport also explains why paradigms can remain stable for long periods. Once conceptual compression has aligned most ideas with a small set of dominant directions, the interpretation process itself reinforces this structure. New ideas that deviate strongly from the dominant conceptual modes are progressively filtered out by the dynamics of interpretation. Consequently, paradigms act as **self-reinforcing attractors** in idea space.

Emergence of Canonical Narratives. As paradigms stabilize, the field's intellectual history becomes increasingly organized around a small set of canonical works. These papers define the conceptual directions along which subsequent ideas evolve, and they accumulate disproportionate epistemic credit through the mechanisms described in Section 6.

Thus paradigm formation and credit concentration are two manifestations of the same underlying process: the **spectral compression of idea space**.

While conceptual compression stabilizes paradigms, the spectral structure of epistemic transport also permits structural shifts in the dominant conceptual directions. When the balance of growth rates between different conceptual modes changes, the geometry of idea space can reorganize. Such reorganizations correspond to the **scientific revolutions** described by Kuhn. In the next section we formalize this phenomenon by introducing a mathematical theory of **conceptual phase transitions in scientific fields**.

7.2 Influence Phase Transition Theorem

The previous section interpreted scientific paradigms as low-dimensional conceptual structures emerging from compression in idea space. We now extend this perspective by showing that the dynamics of epistemic transport can undergo **qualitative regime shifts**. These shifts correspond to structural reorganizations in the way ideas propagate and accumulate influence across scientific communities.

In this section we formalize these changes as **influence phase transitions** in the dynamics of scientific knowledge.

Competing Forces in Idea Propagation

Two fundamental processes govern the propagation of ideas through citation networks:

1. **Branching expansion of the literature.**
   Each paper generates a number of descendants through citations. If the average branching factor is b, the number of citation paths grows approximately as
   $$N(n) \approx b^n.$$
2. **Spectral amplification of conceptual influence.**
   Through repeated interpretation, conceptual contributions evolve according to the spectral growth rate determined by the largest Lyapunov exponent λ:
   $$|T_n v| \approx e^{\lambda n}|v|.$$

The long-run structure of scientific credit and idea evolution emerges from the interaction between these two exponential processes.

Critical Balance Between Growth Rates. A critical structural threshold arises when the spectral growth rate of conceptual amplification equals the logarithmic growth rate of the citation network.

$$\lambda = \log b.$$

At this point the amplification of conceptual influence exactly balances the expansion of citation paths. The system lies at a **critical boundary** separating two qualitatively different regimes of scientific organization.

Influence Phase Transition Theorem. Let b denote the branching factor of the citation network and λ the largest Lyapunov exponent governing epistemic transport. Then the structure of scientific credit and influence undergoes a phase transition at the critical point

$$\lambda = logb.$$

Specifically:

- **Subcritical regime ($\lambda < logb$).**
  Network branching dominates conceptual amplification. Credit contributions diffuse across many ancestors, producing relatively egalitarian distributions of influence.
- **Supercritical regime ($\lambda > logb$).**
  Conceptual amplification dominates network expansion. Small differences in early ideas are exponentially amplified, concentrating influence among a small set of foundational works.

Egalitarian vs. Superstar Science. These two regimes correspond to distinct organizational structures of scientific fields. In the **egalitarian regime**, credit is distributed across a broad intellectual lineage. Many ancestor papers contribute meaningfully to the conceptual development of the field. In the **superstar regime**, the dynamics of epistemic amplification concentrate influence among a few canonical works. These papers serve as dominant conceptual anchors, shaping the trajectory of subsequent research.

The transition between these regimes resembles a phase transition in statistical physics: a small change in the balance between conceptual amplification and network branching can produce a large-scale reorganization of the credit structure of science.

This phase transition provides a mathematical interpretation of the scientific revolutions described by Thomas Kuhn. In Kuhn's account, revolutions occur when an existing paradigm loses its explanatory power and a new conceptual framework reorganizes the field. Within the present framework, such revolutions correspond to **structural changes in the spectral dynamics of epistemic transport**. When the growth rate of a new conceptual direction surpasses the effective branching rate of the existing citation network, the system crosses the critical threshold. A new set of dominant conceptual modes emerges, reorganizing idea space and redefining the field's intellectual structure.

The critical condition $\lambda \approx logb$ also suggests measurable precursors to paradigm shifts. As the system approaches this boundary, the **spectral radius** of epistemic transport approaches unity and conceptual modes become increasingly unstable. Empirically, this may manifest as: rising conceptual diversity in scientific discourse; increasing volatility in citation structures; rapid shifts in the dominant directions of idea embeddings. These patterns may therefore serve as early signals of impending paradigm transitions.

While the phase transition theorem explains structural reorganizations of scientific paradigms, it does not yet explain **why certain ideas become dominant while others disappear**. Addressing this question requires incorporating mechanisms of cultural transmission and selection.

In the next section we integrate the spectral framework developed here with theories of **cultural evolution**, providing a unified perspective on how scientific ideas propagate, compete, and survive within the scientific community.

7.3 Scientific Revolutions as Subspace Rotation

Sections 7.1 and 7.2 described how paradigms emerge from the compression of idea space. We now introduce a geometric interpretation of **scientific revolutions**. Within the spectral framework developed in this paper, paradigm shifts correspond to **rotations of the dominant conceptual subspace** governing epistemic transport.

Conceptual Geometry of Scientific Fields. Recall that repeated interpretation compresses idea space into a low-dimensional subspace spanned by the dominant Lyapunov vectors. Let

$$E^{+} = span(e_1, e_2, \dots, e_k)$$

denote the expanding conceptual subspace associated with the largest Lyapunov exponents. Most ideas produced within the field lie close to this subspace. Consequently, the vectors e_i define the **conceptual axes** that organize scientific thinking within a paradigm.

Under normal scientific development, the orientation of this subspace remains relatively stable. New ideas primarily represent combinations or refinements of the same dominant conceptual directions.

Instability of Conceptual Modes. However, the spectral structure of epistemic transport is not necessarily fixed. As new conceptual directions appear in the literature, the growth rates associated with different modes can change. When an emerging conceptual direction acquires a larger Lyapunov exponent than existing ones, the dominant expanding subspace may shift. Mathematically, this corresponds to a **rotation of the leading Lyapunov subspace** within idea space. Let E_t^{+} denote the dominant conceptual subspace at time t. A paradigm shift occurs when

$$E_{t+\Delta t}^{+} \neq E_t^{+}.$$

In other words, the principal directions along which ideas expand and propagate are replaced by new conceptual axes. We can formalize this phenomenon as follows.

Subspace Rotation Theorem. Let E_t^{+} denote the dominant Lyapunov subspace governing epistemic transport at time t. If the ordering of Lyapunov exponents changes due to shifts in interpretation dynamics, the expanding subspace rotates within idea space. This rotation produces a structural reorganization of conceptual coordinates within the field.

The consequence of this reorganization is a **paradigm shift**: ideas previously considered central may become peripheral, while previously marginal conceptual directions become dominant (see Appendix F). In Kuhn's account, new conceptual framework replaces existing one, reorganizing the questions and methods of a science discipline.

Subspace rotation produces distinctive empirical signatures in the structure of scientific literature. For example: embedding directions representing dominant concepts should change over time; canonical ancestor papers may shift across generations; clusters of scientific work may reorganize around new conceptual axes. These changes reflect a reorientation of the conceptual geometry of the field.

This geometric model complements the phase transition framework introduced in Section 7.2. While the phase transition describes changes in the ***intensity*** of conceptual amplification, subspace rotation captures changes in the ***direction*** of conceptual growth. Together, these mechanisms provide a mathematical account of how scientific fields reorganize over time.

7.4 Cultural Selection of Scientific Ideas

The spectral framework developed in this paper describes how ideas propagate and transform through chains of scientific interpretation. However, the dynamics of scientific knowledge are not purely mechanical processes of transmission. Ideas compete for attention, adoption, and continuation within the scientific community. To understand this evolutionary dimension, it is useful to connect the present model with theories of **cultural evolution**.

In particular, the dynamics of epistemic transport closely parallel the mechanisms described in Dual Inheritance Theory, a framework in evolutionary anthropology and cultural evolution developed by scholars such as Robert Boyd and Peter J. Richerson (1988). Dual Inheritance Theory proposes that human cultural systems evolve through processes analogous to biological evolution, operating through three fundamental mechanisms: **variation, selection, and inheritance**.

Within the present mathematical framework, these mechanisms emerge naturally from the dynamics of idea propagation.

Variation Through Interpretation. Scientific ideas rarely propagate unchanged. When a researcher cites and interprets a previous work, the conceptual content of the idea is typically modified, extended, or reframed. In the present model, this process is represented by **interpretation operators** acting on vectors in idea space. If v represents the conceptual structure of an idea, then the interpretation performed by a subsequent work applies a transformation

$$v' = Av.$$

Different interpretation operators correspond to different ways of extending or reframing earlier work. Over time, these transformations generate a population of conceptually diverse variants derived from the same intellectual ancestor. Thus the interpretation process provides a natural mechanism for **conceptual variation** within the scientific literature.

Selection Through Epistemic Amplification. Not all conceptual variants propagate equally. Some interpretations prove influential and generate large citation cascades, while others disappear rapidly from the intellectual landscape. Within the spectral framework, this differential survival arises from the **growth rates** of conceptual directions. Directions associated with larger Lyapunov exponents are amplified under repeated interpretation, while directions associated with negative exponents decay. Consequently, conceptual variants aligned with expanding spectral directions are more likely to propagate through future work. These directions function as **selection gradients** in idea space, determining which conceptual structures survive and proliferate.

Inheritance Through Citation Networks. Finally, ideas persist through **citation-based inheritance**. When a new paper builds upon earlier work, it transmits conceptual content across generations of research. The citation network therefore serves as the genealogical structure through which ideas reproduce and evolve. If a citation chain involves interpretation operators $A_1, A_2, \ldots, A_n$, the conceptual structure inherited by the descendant paper is

$$v_n = A_n A_{n-1} \cdots A_1 v_0.$$

This process constitutes the mechanism of **cultural inheritance** within the scientific system.

Evolutionary Dynamics of Scientific Knowledge. Taken together, these mechanisms show that scientific knowledge evolves through a process closely analogous to cultural evolution:

Evolutionary Mechanism	Scientific Process
Variation	reinterpretation and recombination of ideas
Selection	spectral amplification of successful conceptual directions
Inheritance	citation-based transmission across generations

In this sense, the scientific literature can be viewed as a large-scale **evolving population of ideas**.

Spectral Structure of Cultural Selection. The mathematical framework developed here adds a new dimension to cultural evolution theory by identifying the **spectral geometry of idea space** as the underlying mechanism that governs selection. Instead of treating cultural selection purely as a sociological process, the present model shows how selection pressures emerge from the structure of epistemic transport itself.

The dominant Lyapunov directions define the conceptual axes along which cultural selection operates. Ideas aligned with these directions are more likely to propagate, while ideas orthogonal to them gradually disappear from the intellectual landscape.

Toward a Unified Theory of Idea Evolution. By linking the spectral dynamics of epistemic transport with cultural evolutionary theory, this framework provides a unified perspective on the evolution of scientific knowledge. Scientific fields evolve not only through the accumulation of discoveries but also through the **selective amplification of certain conceptual structures** within idea space.

This evolutionary perspective complements the paradigm-based view introduced by Thomas Kuhn. While Kuhn emphasized the sociological dynamics of paradigm shifts, the present framework identifies the underlying mathematical processes through which ideas vary, compete, and stabilize within scientific communities.

8. Discussion

This paper develops a mathematical framework for understanding the evolution of scientific ideas and the distribution of scientific credit. By representing ideas as vectors in a conceptual space and modeling interpretation as a sequence of linear transformations along citation networks, the theory connects the dynamics of knowledge transmission with tools from dynamical systems, spectral theory, and information geometry. The resulting framework links three domains that are often studied separately: the **structure of scientific ideas, the dynamics of knowledge transmission, and the distribution of scientific credit**. The discussion below highlights four broader implications of the theory.

8.1 A Unifying Theory of Scientific Evolution

A central contribution of the framework is to provide a **unified mathematical model of scientific evolution**. Scientific ideas propagate through the literature via citation chains, but each transmission involves reinterpretation and conceptual transformation. Modeling these transformations as multiplicative operators allows the evolution of ideas to be analyzed using the spectral theory of dynamical systems.

The multiplicative structure of epistemic transport produces several fundamental regularities. Repeated interpretation compresses idea space into a small number of dominant conceptual directions, generating the canonical conceptual modes described in Section 3. The multiplicative ergodic theorem implies a decomposition of idea space into Lyapunov subspaces, providing a natural explanation for the emergence of conserved epistemic structures discussed in Section 5.

Within this framework, the long-run evolution of scientific fields is governed by a small number of dynamical quantities—Lyapunov exponents, spectral radii, and branching factors—that determine both conceptual stability and credit dynamics. This perspective suggests that the development of scientific knowledge may obey statistical regularities analogous to those found in other complex dynamical systems.

8.2 Bridging Mathematics and the Sociology of Science

The theory also builds a bridge between **formal mathematical models and classical theories in the sociology and philosophy of science**. In particular, the framework provides a quantitative interpretation of several well-known qualitative ideas:

- The notion of **scientific paradigms** described by Thomas Kuhn corresponds to strong compression of idea space into a small number of canonical conceptual modes.
- **Scientific revolutions** can be interpreted as structural transitions in epistemic dynamics, where the dominant Lyapunov subspaces rotate and new conceptual directions become stable.
- The evolutionary perspective on cultural knowledge emphasized in Dual Inheritance Theory is reflected in the processes of variation, selection, and inheritance that govern the propagation of ideas in citation networks.

By translating these sociological concepts into explicit dynamical quantities, the framework provides a formal language for analyzing the evolution of scientific knowledge.

8.3 Implications for Measuring Scientific Impact

The model also has direct implications for the measurement of scientific impact. Traditional bibliometric measures such as citation counts treat citations as additive signals of influence. In contrast, the present framework suggests that influence propagates **multiplicatively** through citation cascades. This perspective leads to several quantitative predictions:

- Credit contributions decay exponentially with citation distance.
- Scientific credit accumulates as products of singular values along epistemic transport chains.
- Credit distributions across intellectual ancestors follow heavy-tailed power laws.
- The exponent of these distributions depends on the balance between network branching and conceptual amplification.

These results imply that citation inequality may arise not primarily from social factors but from the **intrinsic multiplicative dynamics of knowledge transmission**. The framework therefore offers a new interpretation of highly skewed credit distributions commonly observed in science.

8.4 Directions for Empirical Research

The theoretical results presented here generate a number of empirical predictions that can be tested using modern bibliometric datasets and scientific embeddings. Several promising directions for future work include:

1. **Estimating epistemic transport operators.**
   Using text embeddings or citation-based representations to estimate interpretation operators between successive generations of papers.
2. **Measuring conceptual dimensionality.**
   Tracking the intrinsic dimensionality of scientific fields over time to test predictions about dimensionality collapse and sigmoid growth of conceptual diversity.
3. **Analyzing credit propagation.**
   Reconstructing citation genealogies to evaluate predictions about influence decay, credit half-lives, and heavy-tailed ancestor credit distributions.
4. **Detecting paradigm transitions.**
   Monitoring spectral properties of epistemic transport to identify early-warning signals of conceptual phase transitions in scientific fields.

The rapid growth of large-scale scholarly datasets and machine-learning representations of scientific texts makes such empirical investigations increasingly feasible.

8.5 Toward a Quantitative Science of Science

More broadly, the framework developed in this paper contributes to the emerging field of a **quantitative science of science**. By combining ideas from dynamical systems, network theory, and cultural evolution, the model provides a principled approach for studying how scientific knowledge accumulates and reorganizes over time.

If the predictions of the theory are empirically supported, they would suggest that the structure of scientific knowledge is shaped not only by social institutions and individual creativity but also by deeper dynamical principles governing the transmission and transformation of ideas.

Understanding these principles may ultimately allow us to better characterize the conditions under which scientific fields grow, stabilize, or undergo conceptual revolutions.

Appendices

Appendix A. Multiplicative Ergodic Theorem Background

The analysis of epistemic transport in Sections 4–6 relies on the spectral properties of products of interpretation operators along citation chains. These properties are formally described by the **Multiplicative Ergodic Theorem (MET)**, originally established by Vladimir Oseledets. This appendix briefly summarizes the aspects of the theorem used in the main text.

Let $\{J_i\}$ denote a sequence of interpretation operators associated with papers in a citation chain. For a chain of length n, the cumulative interpretation operator transporting conceptual perturbations from paper k to paper F is $\Phi_{F,k} = J_{i_n} J_{i_{n-1}} \cdots J_{i_1}$. This operator describes how conceptual perturbations propagate across successive reinterpretations of earlier work.

The multiplicative ergodic theorem states that under mild regularity conditions there exist real numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ called **Lyapunov exponents** such that for almost every realization of the operator sequence, $\lim_{n\to\infty} \frac{1}{n} log \parallel \Phi_n v \parallel = \lambda_i$ whenever the vector v lies in the corresponding invariant subspace. These exponents describe the asymptotic exponential growth or decay rates of perturbations under repeated application of interpretation operators.

The theorem further implies that the conceptual space $\mathbb{R}^d$ decomposes into invariant subspaces $\mathbb{R}^d = E^{(+)} \oplus E^{(0)} \oplus E^{(-)}$
corresponding respectively to: expanding directions $(\lambda > 0)$, neutral directions $(\lambda = 0)$, contracting directions $(\lambda < 0)$. In the main text these subspaces were interpreted as representing conceptual components that are respectively amplified, preserved, or suppressed during repeated reinterpretation of ideas.

Appendix B. Proofs of Spectral Results

This appendix provides formal justification for the spectral results used in the main text.

Proposition B.1. Let $\{J_i\}$ be a sequence of interpretation operators satisfying the conditions of the multiplicative ergodic theorem. Then the conceptual space admits a measurable decomposition

$$\mathbb{R}^d = \bigoplus_{j=1}^{r} E^{(j)}$$

such that for any nonzero $v \in E^{(j)}$

$$\lim_{n\to\infty} \frac{1}{n} log \parallel \Phi_n v \parallel = \lambda_j.$$

Sketch of proof. The result follows directly from the multiplicative ergodic theorem applied to the sequence of interpretation operators. Under standard integrability conditions on the matrix norms, the theorem guarantees the existence of invariant subspaces associated with distinct Lyapunov exponents. The decomposition arises by grouping together subspaces with positive, zero, and negative exponents. Section 6.1 used Lyapunov exponents to derive bounds on the influence of ancestor papers across citation chains.

Proposition B.2 (Influence decay). Let k and F be papers separated by citation distance $\ell(k, F)$. If the conceptual perturbation introduced by paper k lies primarily within a contracting Lyapunov subspace with exponent $-\gamma < 0$, then $\parallel \phi_k(F) \parallel \leq Ce^{-\gamma\ell(k,F)}$.

Proof. Let v denote the conceptual perturbation introduced at paper k. If v lies in a contracting Lyapunov subspace with exponent $-\gamma$, then by definition $\| \Phi_{\ell} v \| \approx e^{-\gamma\ell} \| v \|$. Setting $\ell = \ell(k, F)$ yields the stated inequality up to a constant factor C determined by the norm equivalence between the embedding space and the conceptual representation used in the model.

Appendix C. Credit Distribution Derivation

This appendix derives the heavy-tailed credit distributions discussed in Section 6.

Letc $c_k(F)$ denote the conceptual contribution of ancestor k to paper F. Along a citation chain of length n,

$$c_k(F) = c_0 \prod_{t=1}^{n} \sigma_t$$

where σ_t are singular values associated with interpretation operators. Taking logarithms, $log c_k(F) = log c_0 + \sum_{t=1}^{n} log \sigma_t$. Thus the logarithm of epistemic credit is a sum of random variables.

If the variables $log \sigma_t$ have finite variance, the central limit theorem implies that $log c_k(F)$ converges in distribution to a Gaussian random variable for moderate chain lengths. Consequently, credit contributions follow an approximate log-normal distribution in intermediate regimes.

Citation networks exhibit branching structures in which many citation paths descend from early papers. Aggregating multiplicative contributions across these paths yields a stochastic branching process. Under standard conditions for multiplicative branching processes, the resulting distribution of accumulated credit develops a **power-law tail** $P(C > x) \sim x^{-\alpha}$. The exponent α depends on both the branching rate of the citation network and the statistical properties of the multiplicative factors σ_t.

This result provides a theoretical explanation for the heavy-tailed concentration of scientific influence. Early contributions that align with conceptual directions preserved under interpretation dynamics propagate through large citation cascades and accumulate substantial epistemic credit.

Appendix D. Derivation of the Universal Scaling Law for Scientific Credit

This appendix derives the universal scaling relation presented in Section 6.5 linking the distribution of scientific credit to the interaction between **network branching structure** and **epistemic amplification dynamics**. The derivation connects the multiplicative transport model introduced in the main text with the heavy-tailed distributions observed in citation-based credit allocation.

The key result is that the tail exponent of the ancestor credit distribution is

$$\alpha = 1 + \frac{log b}{\lambda},$$

where b is the effective **branching factor** of the citation network, and λ is the dominant **Lyapunov exponent** governing epistemic amplification along citation chains. This relation provides a bridge between network topology and conceptual dynamics.

Consider a citation cascade originating from a focal paper P_0. The cascade traces backward through the citation network to its intellectual ancestors. Let the average number of references per paper be b. In a tree approximation of the citation network, the number of ancestors at citation distance k grows approximately as

$$N_k \approx b^k.$$

This captures the combinatorial expansion of the citation genealogy. Each step in the citation chain also involves an **interpretation operator**

$$v_{k+1} = A_k v_k,$$

as introduced in Section 2.2. The cumulative epistemic transformation along a chain of length k is

$$T_k = A_{k-1} A_{k-2} \cdots A_0.$$

The Multiplicative Ergodic Theorem (Appendix A) implies that for generic directions in idea space,

$$|T_k v| \sim e^{\lambda k} |v|,$$

where λ is the dominant Lyapunov exponent. This describes the exponential amplification or attenuation of conceptual influence along citation chains.

Scientific credit is propagated backward along citation chains through epistemic transport. Let the conceptual influence weight of an ancestor at distance k be proportional to the inverse amplification required to reconstruct the idea:

$$w_k \propto e^{-\lambda k}.$$

This exponential decay arises naturally from the Lyapunov growth of epistemic transport. Thus: deeper ancestors contribute less influence, but the number of ancestors grows exponentially with depth. This competition between **exponential growth in ancestor count** and **exponential decay in influence** determines the overall credit distribution.

To derive the distribution of ancestor credit, we relate citation depth to rank in the ancestor list. From the branching relation

$$N_k \approx b^k,$$

the cumulative number of ancestors up to distance k is approximately

$$r \sim b^k.$$

Solving for depth gives

$$k \approx \frac{log r}{log b}.$$

Substituting this into the influence decay relation

$$w_k \propto e^{-\lambda k}$$

yields the rank-credit relationship

$$w(r) \propto r^{-\lambda / log b}.$$

This implies that credit assigned to ancestors follows a **power-law decay with rank**.

Let C denote the credit received by an ancestor. The rank relation implies

$$C \propto r^{-\beta},$$

with

$$\beta = \frac{\lambda}{logb}.$$

The complementary cumulative distribution satisfies

$$P(C > x) \propto x^{-1/\beta}.$$

Differentiating yields the probability density

$$P(C) \propto C^{-\alpha},$$

with exponent

$$\alpha = 1 + \frac{1}{\beta}.$$

Substituting the expression for β gives

$$\alpha = 1 + \frac{logb}{\lambda}.$$

The resulting exponent reflects the balance between two fundamental forces: **Network expansion.** The citation network grows combinatorially with branching factor b. Larger b increases the number of potential ancestors contributing credit. **Conceptual amplification.** Interpretation dynamics amplify or attenuate conceptual signals according to the Lyapunov exponent λ. The ratio $logb/\lambda$ therefore determines the **degree of credit inequality**.

Weak conceptual amplification ($\lambda \ll logb$). Many ancestors contribute comparable influence. Credit distributions are relatively egalitarian. **Strong conceptual amplification** ($\lambda \gg logb$)**.** Only a few ancestors dominate intellectual lineage. Credit distributions become highly unequal.

The scaling law depends only on two coarse-grained quantities: branching factor b and Lyapunov exponent λ. Because these quantities summarize generic structural and dynamical properties of scientific fields, the resulting exponent is expected to be **approximately universal across disciplines**. This universality explains why heavy-tailed credit distributions appear consistently across many scientific citation networks.

The scaling law suggests a direct empirical test: estimate the **branching factor** b using average citation counts; estimate the **Lyapunov exponent** λ from singular-value growth of interpretation operators; measure the **ancestor credit distribution** and estimate the empirical power-law exponent. Agreement with $\alpha = 1 + \frac{logb}{\lambda}$ would provide strong evidence for the multiplicative epistemic transport model proposed in this paper.

Appendix E. Derivation of the Sigmoid Trajectory of Conceptual Dimensionality

Section 3.4 proposed that the intrinsic conceptual dimensionality of a scientific field follows a **sigmoid life-cycle trajectory**. This appendix derives that result from the underlying dynamics of **conceptual innovation** and **epistemic compression** introduced earlier in the paper.

The key idea is that the dimensionality of idea space evolves under two competing processes: **Conceptual innovation**, which introduces new directions in idea space; **Interpretation compression**, which collapses ideas toward dominant conceptual modes. When these forces are modeled together, the resulting dynamics naturally generate logistic (sigmoid) growth. Let

$$d(t)$$

denote the **effective conceptual dimensionality** of a scientific field at time t. This dimensionality measures the number of independent conceptual directions required to represent ideas in the field.

Empirically, $d(t)$ can be estimated as: intrinsic manifold dimension of paper embeddings, effective rank of the idea covariance matrix, number of dominant singular values in the interpretation operator. Conceptually: larger $d(t)$ indicates **diverse conceptual exploration**; smaller $d(t)$ indicates **compression toward canonical frameworks**.

New research directions introduce additional conceptual axes into idea space.
Let the **innovation rate** be proportional to the existing diversity:

$$innovation\ rate = \gamma d$$

where γ measures the rate at which new conceptual directions emerge. This captures the idea that fields with more active research directions generate more conceptual variation.

Interpretation operators compress ideas toward dominant conceptual modes. As dimensionality increases, many directions become redundant or unstable and collapse under repeated interpretation. Let the **compression rate** grow with dimensionality:

$$compression\ rate = \kappa d^2/d_{max}$$

where: d_{max} is the maximum conceptual diversity achievable in the field, κ measures the strength of epistemic compression. The quadratic term reflects the fact that compression becomes stronger when many directions compete for explanatory dominance.

Combining innovation and compression yields the differential equation

$$\frac{dd}{dt} = \gamma d - \kappa \frac{d^2}{d_{max}}.$$

Factoring the right-hand side gives

$$\frac{dd}{dt} = rd\left(1 - \frac{d}{d_{max}}\right)$$

where $r = \gamma$**.** This is the **logistic growth equation**, widely used to model systems with initial expansion and eventual saturation. Solving the logistic equation yields

$$d(t) = \frac{d_{max}}{1 + Ce^{-rt}}.$$

Applying the initial condition $d(0) = d_0$ determines the constant

$$C = \frac{d_{max} - d_0}{d_0}.$$

Thus

$$d(t) = \frac{d_{max}}{1 + \frac{d_{max} - d_0}{d_0} e^{-rt}}.$$

This function has the characteristic **sigmoid shape** predicted in Section 3.4.

The logistic dynamics imply three stages in the evolution of scientific fields.

Early exploration phase. When $d \ll d_{max}$,

$$\frac{dd}{dt} \approx rd$$

and dimensionality grows exponentially. This corresponds to early stages of a field where many competing ideas emerge.

Expansion phase. As dimensionality increases, conceptual diversity continues to grow but at a slower rate. Multiple theoretical frameworks coexist and compete.

Compression phase. When d approaches d_{max}, compression dominates:

$$\frac{dd}{dt} \to 0$$

and dimensionality stabilizes. The field matures and converges toward a small number of **canonical conceptual modes** (Section 3.3).

The logistic dynamics can also be interpreted in terms of the spectral properties of interpretation operators. Recall that epistemic transport compresses idea space toward dominant Lyapunov subspaces. Let $r_i(t)$ denote the singular values associated with conceptual directions. Directions with smaller Lyapunov exponents decay over time, reducing effective dimensionality. The logistic trajectory therefore describes the **balance between the creation of new conceptual directions and their spectral suppression by epistemic transport**.

The resulting trajectory suggests a natural life cycle for scientific disciplines:

Stage	Conceptual structure
Birth	rapid expansion of conceptual directions
Growth	coexistence of competing frameworks
Maturity	compression toward canonical modes

This provides a quantitative model for the evolution of conceptual diversity in science and connects directly to the paradigm dynamics discussed in Section 7.

Appendix F. Subspace Rotation Theorem and the Geometry of Scientific Revolutions

Section 7.4 proposed that **scientific revolutions correspond to rotations of dominant conceptual subspaces** in idea space. This appendix provides the mathematical derivation underlying that claim. The analysis shows that structural changes in epistemic transport operators cause rotations of the leading Lyapunov subspaces, thereby replacing the dominant conceptual directions that organize a scientific field.

Let the scientific idea space be a vector space

$$V = \mathbb{R}^n$$

in which each paper corresponds to a vector representation. Interpretation along a citation chain is described by the sequence of operators

$$A_0, A_1, A_2, \ldots$$

with epistemic transport

$$T_k = A_{k-1} A_{k-2} \cdots A_0.$$

Under the conditions of the multiplicative ergodic theorem, there exists a decomposition

$$V = E_1 \oplus E_2 \oplus \cdots \oplus E_m$$

into **Lyapunov subspaces**. Each subspace E_i corresponds to directions in idea space that grow or decay at rate

$$\lambda_i.$$

The leading subspace E_1 defines the **dominant conceptual framework** organizing a scientific field.

Scientific revolutions occur when the interpretation operators change due to the introduction of new conceptual frameworks. Suppose the epistemic transport operators shift from

$$A_k$$
to
$$A_k^{'} = A_k + \Delta_k$$
where Δ_k represents conceptual perturbations produced by new theoretical developments. The resulting transport operator becomes
$$T_k^{'} = A_{k-1}^{'} A_{k-2}^{'} \cdots A_0^{'}.$$
Because the Lyapunov subspaces depend on the asymptotic spectral structure of these products, perturbations in the operators can change the orientation of the dominant subspaces. Let
$$U$$
denote an orthonormal basis for the dominant conceptual subspace before the perturbation. Similarly, let
$$U^{'}$$
be the basis for the dominant subspace after the perturbation. The difference between conceptual paradigms can be quantified using the **principal angles** between these subspaces. If
$$\sigma_1, \ldots, \sigma_d$$
are the singular values of
$$U^{\top} U^{'},$$
then the principal angles are
$$\theta_i = arccos(\sigma_i).$$
Large principal angles correspond to strong conceptual reorientation.

Matrix perturbation theory implies that the invariant subspaces of linear operators vary continuously with the operator itself. Let
$$A$$
denote the expected interpretation operator in the stationary regime:
$$A = \mathbb{E}[A_k].$$
Let the perturbed operator be
$$A^{'} = A + \Delta.$$
The Davis–Kahan sin–θ theorem provides a bound on the rotation of invariant subspaces:
$$sin\Theta(E_1, E_1') \leq \frac{|\Delta|}{\delta}$$
where E_1 is the original dominant subspace, $E_1^{'}$ is the perturbed subspace, δ is the spectral gap between the leading eigenvalues. Thus, conceptual subspaces rotate when perturbations exceed the stabilizing spectral gap. We can now state the central result.

Subspace Rotation Theorem. Let A_k be a stationary sequence of interpretation operators satisfying the conditions of the multiplicative ergodic theorem. Let E_1 denote the dominant Lyapunov subspace with spectral gap δ. If the interpretation dynamics are perturbed by operators Δ_k such that
$$|\Delta_k| \gtrsim \delta,$$
then the dominant conceptual subspace rotates by a non-negligible angle. Equivalently,
$$\Theta(E_1, E_1') > 0.$$
This rotation corresponds to the replacement of one conceptual framework by another.

The theorem implies that scientific revolutions are **geometric transformations of idea space**. Before the revolution: conceptual activity is concentrated along a dominant subspace E_1.

After the revolution: a new subspace $E_1^{'}$ becomes dominant. The coordinate system organizes the field changes. Old explanatory axes lose explanatory power, and new conceptual directions become central.

The theory also suggests measurable signals preceding revolutions. If the spectral gap

$$\delta = \lambda_1 - \lambda_2$$

shrinks, the dominant conceptual subspace becomes unstable. Small perturbations can then produce large rotations. Impending paradigm shifts may be detectable by observing: decreasing spectral gaps, increasing variance in conceptual directions, rising conceptual dimensionality.

Appendix G. Neutral Subspace Dimension Theorem

Section 5 introduced the concept of conserved epistemic quantities and argued that these correspond to the neutral Lyapunov subspace of the epistemic transport dynamics. This appendix provides a formal derivation of the result:

$$\textit{Number of conserved epistemic quantities} = dim(E_0),$$

where E_0 denotes the neutral Lyapunov subspace associated with Lyapunov exponent 0. This theorem establishes a precise relationship between the spectral structure of interpretation dynamics and the number of stable conceptual invariants in a scientific field.

Recall that interpretation along citation chains is modeled by a sequence of operators

$$A_0, A_1, A_2, \ldots$$

acting on the idea space

$$V = \mathbb{R}^n.$$

The cumulative epistemic transport after k steps is

$$T_k = A_{k-1} A_{k-2} \cdots A_0.$$

Under standard integrability conditions, the **Multiplicative Ergodic Theorem** guarantees the existence of Lyapunov exponents

$$\lambda_1 > \lambda_2 > \cdots > \lambda_m$$

and an invariant decomposition of idea space

$$V = E_1 \oplus E_2 \oplus \ldots \oplus E_m.$$

Each subspace E_i corresponds to directions that grow or decay exponentially at rate λ_i.

Define the **neutral subspace**

$$E_0 = v \in V : \lambda(v) = 0.$$

Vectors in this subspace satisfy

$$|T_k v| \sim |v|$$

for large k. Thus, they are neither amplified nor suppressed by interpretation dynamics. Conceptually, these directions correspond to **epistemic content that remains stable under repeated reinterpretation**.

A conceptual feature $q(v)$ is said to be **conserved under epistemic transport** if

$$q(T_k v) = q(v)$$

for all k. Such quantities represent conceptual structures that survive across citation chains without systematic distortion.

Assume that epistemic quantities are represented by linear functionals

$$q(v) = w^\top v.$$

Conservation requires

$$w^\top T_k v = w^\top v$$

for all v. This implies

$$T_k^{\top} w = w.$$

Thus conserved quantities correspond to **invariant directions of the dual transport operator**.

Suppose v lies in a Lyapunov subspace with exponent λ. Then

$$|T_k v| \sim e^{\lambda k}.$$

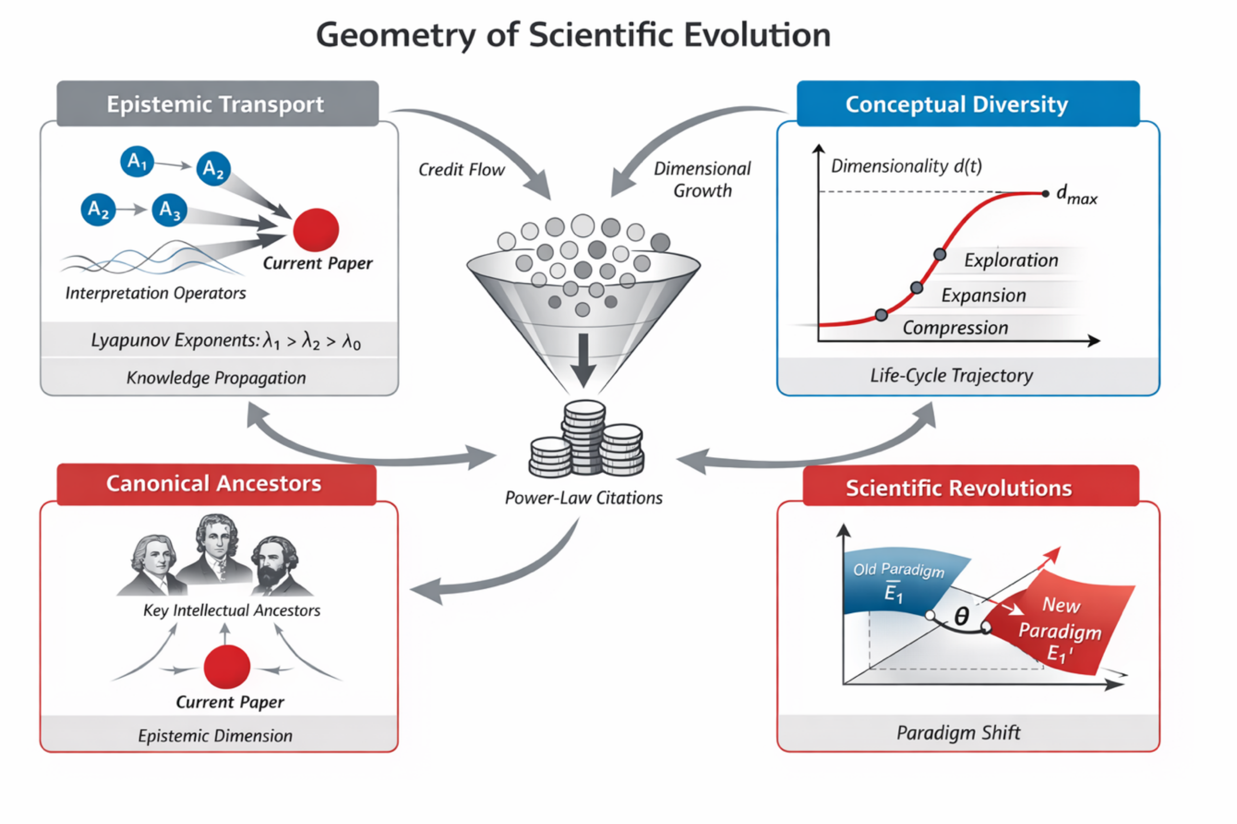
If $\lambda > 0$, the direction grows exponentially and therefore cannot correspond to a conserved quantity. If $\lambda < 0$, the direction decays and eventually vanishes. Only directions with $\lambda = 0$ remain stable under repeated transport. Therefore, conserved epistemic quantities must lie in the **neutral Lyapunov subspace**. We can now state the main result.

Neutral Subspace Dimension Theorem. Let A_k be a stationary sequence of interpretation operators satisfying the assumptions of the Multiplicative Ergodic Theorem. Let E_0 denote the neutral Lyapunov subspace. Then the number of independent conserved epistemic quantities equals the dimension of E_0:

$$N_{conserved} = dim(E_0).$$

Tables and figures

Figure 1. Mathematical properties of epistemic transport generating phenomena in scientific fields



References

Von Neumann, J., & Morgenstern, O. (1947). Theory of games and economic behavior, 2nd rev.

Shapley, L. S. (1953). A value for n-person games.

Lyapunov, A. M. (1992). The general problem of the stability of motion. *International journal of control*, *55*(3), 531-534.

Chung, F. R. (1996). Lectures on spectral graph theory. *CBMS Lectures, Fresno*, *6*(92), 17-21.

Michel, A., Hou, L., & Liu, D. (2008). Stability of Dynamical Systems. Birkhäuser.

Oseledets, V. I. (1968). "Мультипликативная эргодическая теорема. Характеристические показатели Ляпунова динамических систем" [Multiplicative ergodic theorem: Characteristic Lyapunov exponents of dynamical systems]. Trudy MMO. ***19****: 179–210.*

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425-440.

Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196). PMLR.

Beltagy, I., Lo, K., & Cohan, A. (2019, November). SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 3615-3620)

Strogatz, S. H. (2001). Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering (studies in nonlinearity) (Vol. 1). Westview press.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago press.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association*, *69*(345), 118-121.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56-63.

Nica, B. (2018). *A brief introduction to spectral graph theory* (Vol. 3). Zürich: European Mathematical Society.

Mohar, B., Alavi, Y., Chartrand, G., & Oellermann, O. (1991). The Laplacian spectrum of graphs. *Graph theory, combinatorics, and applications*, *2*(871-898), 12.

Raza, A., Munir, M., Hussain, M., & Tasgera, F. (2024). A spectrum-based approach to network analysis utilizing laplacian and signless laplacian spectra to torus networks. *IEEE Access*, *12*, 52016-52029.

Ruelle, D. (1979)."Ergodic theory of differentiable dynamic systems". IHES Publ. Math. **50** (1): 27–58. doi:10.1007/BF02684768. S2CID 56389695.

Walters, P. (1993). A dynamical proof of the multiplicative ergodic theorem. *Transactions of the American Mathematical Society*, *335*(1), 245-257.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., ... & Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics reports*, *874*, 1-92.

Holme, P., & Saramäki, J. (Eds.). (2019). *Temporal network theory* (Vol. 2). New York: Springer.

Barrat, A., Barthelemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge university press.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*(45), 17268-17272.

Hirsch, M., Smale, S., & Devaney, R. (2013). *Differential Equations, Dynamical Systems, and an Introduction to Chaos* (3rd ed.). Academic Press.

Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, *64*(1), 016132.

Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences*, *111*(34), 12325-12330.

Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127-132.

Pastor-Satorras, R., & Castellano, C. (2016). Distinct types of eigenvector localization in networks. *Scientific reports*, *6*(1), 18847.

Sontag, E. D. (1998). Mathematical Control Theory: Deterministic Finite Dimensional Systems. Springer.

Slotine, J. J., & Li, W. (1991). *Applied Nonlinear Control*. Prentice Hall.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford infolab.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, *102*(46), 16569-16572.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... & Barabási, A. L. (2018). Science of science. *Science*, *359*(6379), eaao0185.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Chen, H., Lundberg, S. M., & Lee, S. I. (2022). Explaining a series of models by propagating Shapley values. *Nature communications*, *13*(1), 4512.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Meiss, J. D. (2007). *Differential dynamical systems*. Society for Industrial and Applied Mathematics.

Perko, L. (2013). *Differential equations and dynamical systems* (Vol. 7). Springer Science & Business Media.

Wiggins, S. (2003). *Introduction to applied nonlinear dynamical systems and chaos*. New York, NY: Springer New York.

Khalil, H. K., & Grizzle, J. W. (2002). *Nonlinear systems* (Vol. 3, p. 126). Upper Saddle River, NJ: Prentice hall.

Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

Porter, M. A., & Gleeson, J. P. (2016). Dynamical systems on networks. *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, *4*, 29.

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.

Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137-146).

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, *99*(9), 5766-5771.