The Universal Robustness Trade-off: Entropic Forces vs. Structural Integrity in Scale-Free Systems

Kevin Fathi[1]

[1] *Independent Researcher**

(Dated: November 26, 2025)

Complex systems often exhibit a "Robust Yet Fragile" (RYF) paradox: structures optimized for specific functions are highly resilient to anticipated perturbations but vulnerable to unexpected ones. Concurrently, the prevalence of scale-free architectures with intermediate tail exponents ($2 < \gamma < 3$) contradicts naive Maximum Entropy (MaxEnt) predictions favoring maximal heterogeneity ($\gamma \to 2$). This paper resolves these contradictions by introducing a Duality Theory of Robustness and proving the Principle of Maximal Heterogeneity (PMH), formalized within Majorization Theory. We distinguish between Entropic Robustness (R-I, Schur-concave), resisting stochastic (Type-S) noise, and Structural Robustness (R-II, Schur-convex), resisting targeted (Type-T) stress. We rigorously prove that R-I manifests as a universal, accelerating (convex) information-theoretic force driving $\gamma \to 2$. Conversely, R-II drives the system toward order ($\gamma \to \infty$). The observed intermediate γ^* represents an equilibrium optimizing the universal trade-off between these opposing forces. This framework provides a rigorous foundation for the RYF paradox and offers a falsifiable prediction: any deviation from $\gamma = 2$ implies active optimization against Type-T structural constraints.

I. I. INTRODUCTION: THE ROBUST YET FRAGILE PARADOX

Robustness, the ability of a system to maintain function under perturbation, is a defining feature of complex biological, technological, and social systems [1]. However, optimization often leads to the "Robust Yet Fragile" (RYF) paradox, formalized in the theory of Highly Optimized Tolerance (HOT) [2]. Systems optimized for high performance against specific, anticipated threats frequently become hypersensitive to unanticipated perturbations.

This paradox highlights a fundamental contradiction in the structural requirements for robustness. Frameworks emphasizing redundancy and diversity suggest robustness is maximized by disorder (e.g., uniform distributions) to absorb stochastic failures [3]. Conversely, frameworks emphasizing structural cohesion and specialization suggest robustness is maximized by order (e.g., deterministic structures) to efficiently resist targeted stress.

This conflict is mathematically formalized using Majorization Theory [4], which provides a rigorous partial order quantifying order versus disorder. Established robustness metrics exhibit opposite behaviors under this order: "Robustness-as-Redundancy" is Schur-concave (maximized by disorder), while measures of "Structural Robustness" (e.g., related to spectral integration) are often Schur-convex (maximized by order) [5].

Concurrently, complex systems often exhibit scale-free behavior, characterized by power-law distributions $P(k) \propto k^{-\gamma}$. Empirical observations show a concentration of the tail exponent γ in intermediate ranges, typically $2 < \gamma < 3$ [6]. This contradicts naive applications of the Principle of Maximum Entropy (MaxEnt) [7], which, as we will rigorously prove herein, predicts the state of maximal heterogeneity corresponding to the heaviest allowable tail ($\gamma \to 2^+$ under the constraint of a finite mean).

This paper resolves these contradictions by introducing a Duality Theory of Robustness and synthesizing it with the Principle of Maximal Heterogeneity (PMH). We establish an axiomatic framework based on Majorization Theory to define robustness classes and system evolution. We argue that the RYF paradox stems from optimizing against different classes of perturbations, and the observed structure of scale-free systems represents an equilibrium state balancing these opposing requirements.

II. II. THE MECHANISM: THE DUALITY OF ROBUSTNESS

We utilize Majorization Theory to formalize the concepts of order and disorder. We consider a system described by a probability distribution P over N states, belonging to the probability simplex $\mathcal{P}$.

Definition 1 (Majorization)**.** Let $P, Q \in \mathcal{P}$. Let $P^{\downarrow}$ and $Q^{\downarrow}$ denote the distributions sorted in descending order. P majorizes Q ($P \succ Q$) if P is more concentrated or ordered than Q. Formally:

$$\sum_{k=1}^{K} P^{\downarrow}(k) \geq \sum_{k=1}^{K} Q^{\downarrow}(k), \quad \forall K \in \{1, ..., N-1\}, \tag{1}$$

with equality at $K = N$.

The majorization partial order spans from the Deterministic state $D = (1, 0, ..., 0)$ (maximal order) to the Uniform state $U = (1/N, ..., 1/N)$ (maximal disorder). $D \succ P \succ U$.

We classify functionals $F : \mathcal{P} \to \mathbb{R}$ based on their behavior under this order (Schur-invariance).

* fathikevin@protonmail.com

Definition 2 (Schur-Invariance Classes). • Class I (Heterogeneity): Strictly Schur-concave. $P \succ Q \Rightarrow F_I(P) < F_I(Q)$. Maximized at U.

- Class II (Concentration): Strictly Schur-convex. $P \succ Q \Rightarrow F_{II}(P) > F_{II}(Q)$. Maximized at D.

The key to resolving the robustness contradiction is recognizing that robustness $R(P, \Pi)$ is contingent on both the system state P and the perturbation environment Π. We propose a fundamental dichotomy in environmental stress.

Definition 3 (Type-S Perturbation). Stochastic/Diffuse perturbations (e.g., thermal noise, random component failures). They act diffusively, and the probability of failure is largely independent of the component's specific role.

Definition 4 (Type-T Perturbation). Targeted/Structural perturbations (e.g., attack on a network hub, cascading failure). They represent focused stress that challenges the system's structural integrity or efficiency.

We propose that the structure of robustness is necessarily dual, mirroring the duality of perturbations.

A. A. R-I: Entropic Robustness (Redundancy)

R-I provides resilience against Type-S perturbations.

Axiom 1 (Axiom R-I). Entropic Robustness $R_I(P)$ relies on heterogeneity and redundancy, and is therefore a Class I (Strictly Schur-concave) functional.

To resist diffuse noise, the system must maximize its capacity to absorb stochasticity, achieved by equalizing the importance of all components (moving towards U). This is exemplified by generalized entropy measures, such as Shannon entropy.

B. B. R-II: Structural Robustness (Integration)

R-II provides resilience against Type-T perturbations.

Axiom 2 (Axiom R-II). Structural Robustness $R_{II}(P)$ relies on specialization, integration, and efficiency, and is therefore a Class II (Strictly Schur-convex) functional.

To resist focused stress or maintain high efficiency, the system requires optimized architecture (moving towards D). Examples include metrics related to minimizing effective resistance (Kirchhoff index) in networks [5].

C. C. The Universal Trade-off

The duality implies a fundamental trade-off.

Theorem 1 (The Universal Robustness Trade-off). *A system cannot simultaneously maximize R-I and R-II (for $N > 1$). Any structural change that increases one necessarily decreases the other.*

Proof. 1. By Axiom R-I, $R_I(P)$ is strictly Schur-concave, uniquely maximized at U. 2. By Axiom R-II, $R_{II}(P)$ is strictly Schur-convex, uniquely maximized at D. 3. By the definition of the majorization order, $D \succ U$ (for $N > 1$). The maximizers are distinct and located at opposite ends of the partial order. 4. Consider two system states P_1 and P_2 such that $P_1 \succ P_2$ (i.e., P_1 is strictly more ordered/specialized than P_2). 5. By strict Schur-convexity (Axiom R-II), $R_{II}(P_1) > R_{II}(P_2)$. 6. By strict Schur-concavity (Axiom R-I), $R_I(P_1) < R_I(P_2)$. Therefore, any increase in specialization (towards D) trades redundancy (R-I) for structural integrity (R-II), and vice versa. □

Theorem 1 provides the rigorous foundation for the RYF paradox. Optimization for R-II (against Type-T threats) drives the system towards high specialization (near D), necessarily minimizing R-I, leading to fragility against unanticipated Type-S perturbations.

III. III. THE MATHEMATICAL ENGINE: THE ENTROPIC FORCE

We now apply this duality framework to scale-free systems, characterized by the Zeta (power-law) distribution. This serves as a canonical model for idealized, monotonically decreasing heavy-tailed distributions.

$$P(k; \gamma) = \frac{k^{-\gamma}}{\zeta(\gamma)}, \quad k \in \{1, 2, ...\}, \quad \gamma > 1, \tag{2}$$

where $\zeta(\gamma)$ is the Riemann zeta function. We investigate how Entropic Robustness (R-I), represented by the Shannon entropy $S(\gamma)$, behaves as the structure changes (varying γ).

A. A. The Principle of Maximal Heterogeneity (PMH)

The PMH states that MaxEnt invariably selects the state with the heaviest allowable tail (lowest γ). We provide a rigorous analytical proof for the Shannon case.

We utilize the properties of the Zeta distribution as an exponential family. Let $K(\gamma) = \ln \zeta(\gamma)$ be the cumulant generating function. Let $h(\gamma) = K'(\gamma) = \zeta'(\gamma)/\zeta(\gamma)$. Note that $h(\gamma) = -\mathbb{E}_\gamma[\ln K] < 0$. The Shannon entropy is:

$$S(\gamma) = -\sum_{k \geq 1} P(k; \gamma) \ln P(k; \gamma) = -\gamma h(\gamma) + K(\gamma). \tag{3}$$

Theorem 2 (Strict Monotonicity of Entropy). *The Shannon entropy $S(\gamma)$ of the Zeta distribution is a strictly decreasing function of γ for all $\gamma > 1$.*

Proof. Differentiating $S(\gamma)$ with respect to γ:

$$S'(\gamma) = \frac{d}{d\gamma}(-\gamma K'(\gamma) + K(\gamma)) \quad (4)$$
$$= -K'(\gamma) - \gamma K''(\gamma) + K'(\gamma) \quad (5)$$
$$= -\gamma K''(\gamma). \quad (6)$$

$K''(\gamma) = V(\gamma)$ is the second cumulant. Since the distribution is parameterized by γ with sufficient statistic $\ln K$, $V(\gamma) = \mathrm{Var}_\gamma[\ln K]$. Since $\ln K$ is not constant, the variance is strictly positive, $V(\gamma) > 0$. Therefore, $S'(\gamma) = -\gamma V(\gamma) < 0$ for all $\gamma > 1$. □

Theorem 2 confirms that the entropic force (R-I) strictly drives the system towards lower γ. Under the physical stability constraint (finite mean, requiring $\gamma > 2$), the MaxEnt state is at the boundary: $\lim_{\gamma\to 2^+} S(\gamma)$.

B. B. Characterizing the Entropic Force: Convexity

We further characterize the acceleration of this force towards the boundary.

Theorem 3 (Strict Convexity of Entropy)**.** *The Shannon entropy $S(\gamma)$ of the Zeta distribution is a strictly convex function of γ for all $\gamma > 1$.*

Proof. Differentiating $S'(\gamma)$:

$$S''(\gamma) = \frac{d}{d\gamma}(-\gamma V(\gamma)) \quad (7)$$
$$= -V(\gamma) - \gamma V'(\gamma). \quad (8)$$

$V'(\gamma) = K^{(3)}(\gamma)$. The third cumulant $K^{(3)}(\gamma) = -\mu_3(\gamma)$, where $\mu_3(\gamma)$ is the third central moment of $\ln K$. Thus,

$$S''(\gamma) = -V(\gamma) + \gamma\mu_3(\gamma). \quad (9)$$

Proving $S''(\gamma) > 0$ requires establishing the non-trivial inequality $\gamma\mu_3(\gamma) > V(\gamma)$ for all $\gamma > 1$. This inequality is rigorously proven using a combination of asymptotic analysis and rigorous interval enclosure methods. The full proof is detailed in Appendix A. □

The strict convexity implies that the entropic force driving $\gamma \to 2$ accelerates as it approaches the boundary. This behavior is also observed for generalized entropy frameworks (e.g., Rényi entropy, Tsallis entropy for $0 < q < 1$), suggesting a universal feature of the entropic landscape in these systems.

IV. IV. THE AXIOMATIC FOUNDATION

We now connect the analytical results of the PMH (Theorems 2 and 3) with the Duality Theory (Theorem 1) using an axiomatic framework derived from Majorization Theory. This explains *why* the PMH must hold, contingent on the underlying distributional structure.

We define systemic processes based on their trajectory in the majorization order.

Definition 5 (Class A Process)**.** An Ordering process $\{P_t\}$ is strict Class A if for $t_2 > t_1$, $P_{t_2} \succ P_{t_1}$.

The core idea is to demonstrate that increasing γ in the Zeta family constitutes a Class A process. Since R-I (Entropy) is Class I (Schur-concave), it must necessarily decrease as order increases (the PMH).

Theorem 4 (Axiomatic PMH)**.** *If a scale-free family $\{P_\gamma\}$ is structured such that increasing γ corresponds to a strict Class A process, then any Class I functional (including R-I/Entropy) must be strictly decreasing in γ.*

Proof. Let $\gamma_2 > \gamma_1$. By hypothesis (Class A process), $P_{\gamma_2} \succ P_{\gamma_1}$. By definition of Class I (Strict Schur-concavity, Axiom R-I), $R_I(P_{\gamma_2}) < R_I(P_{\gamma_1})$. Thus, $R_I(\gamma)$ is strictly decreasing. □

Proving that increasing γ is Class A for a specific family requires establishing its majorization structure. For the Zeta distribution (and similar monotonic families), this relies on two key properties.

Axiom 3 (SF0: Monotonicity (Ordering))**.** For all γ, the probabilities are non-increasing: $P(k;\gamma) \geq P(k+1;\gamma)$ for all $k \geq 1$.

The Zeta distribution satisfies this. This axiom ensures that the natural ordering of k aligns with the descending probability ordering ($P^{\downarrow} = P$), simplifying the majorization analysis.

Axiom 4 (SF1: Monotone Likelihood Ratio Property (MLRP))**.** A family $\{P_\gamma\}$ satisfies the strict MLRP if for $\gamma_1 < \gamma_2$, the ratio $R(k) = P(k;\gamma_1)/P(k;\gamma_2)$ is strictly increasing in k.

The Zeta distribution satisfies this: $R(k) = k^{-(\gamma_1-\gamma_2)}\zeta(\gamma_2)/\zeta(\gamma_1)$. Since $\gamma_1 - \gamma_2 < 0$, $R(k)$ is strictly increasing in k.

Lemma 1. *If $\{P_\gamma\}$ satisfies SF0 and SF1, then for $\gamma_1 < \gamma_2$, $P_{\gamma_2} \succ P_{\gamma_1}$.*

Proof. Let $P_1 = P_{\gamma_1}$ and $P_2 = P_{\gamma_2}$. We must show that the cumulative sums satisfy $\sum_{k=1}^{N} P_2(k) \geq \sum_{k=1}^{N} P_1(k)$ for all N. By SF0, the probabilities are already sorted in the natural order of k. SF1 (MLRP) implies a single-crossing property. Since $P_1(k)/P_2(k)$ is strictly increasing and the distributions sum to 1, there must exist a crossing point K^* such that:

- For $k < K^*$, $P_1(k) < P_2(k)$ (The ratio is small).
- For $k > K^*$, $P_1(k) > P_2(k)$ (The ratio is large).

Consider the cumulative difference $\Delta(N) = \sum_{k=1}^{N}(P_2(k) - P_1(k))$.

Case 1 ($N < K^*$): $\Delta(N) > 0$ as $P_2(k) > P_1(k)$ for all $k \leq N$.

Case 2 ($N \geq K^*$): We rewrite $\Delta(N)$ using the total probability constraint ($\sum_{k=1}^{\infty} P_i(k) = 1$):

$$\Delta(N) = \sum_{k=N+1}^{\infty} (P_1(k) - P_2(k)). \tag{10}$$

Since $k > N \geq K^*$, we have $P_1(k) > P_2(k)$. Thus, $\Delta(N) > 0$. Therefore, $P_{\gamma_2} \succ P_{\gamma_1}$. □

Lemma 1 proves that for monotonic distributions satisfying MLRP (like the Zeta family), increasing γ is a Class A process. This rigorously links the Duality framework to the PMH: R-II (Order/Concentration) strictly increases with γ, necessarily causing R-I (Entropy/Heterogeneity) to strictly decrease.

V. V. THE PREDICTION: EQUILIBRIUM AND TRADE-OFF

We synthesize the opposing forces characterized by the Duality Theory.

1. **The Force of R-I (Entropy):** Driven by Type-S perturbations. As proven in Section III (PMH), this force drives the system towards maximal heterogeneity: $\gamma \to 2$. This force is strictly monotonic and convex (Theorems 2 and 3). 2. **The Force of R-II (Structure):** Driven by Type-T perturbations. As shown in Section IV (Lemma 1), higher γ corresponds to higher order (majorization). This force drives the system towards maximal specialization: $\gamma \to \infty$.

Real systems face a mixture of Type-S and Type-T perturbations. The overall fitness, or "Meta-Robustness" (R_{Meta}), depends on optimizing the trade-off based on the environmental statistics.

We define the Meta-Robustness as a weighted combination:

$$R_{Meta}(\gamma) = \alpha R_I(\gamma) + (1 - \alpha) R_{II}(\gamma), \tag{11}$$

where $\alpha \in [0, 1]$ represents the relative importance of Type-S perturbations in the environment.

Theorem 5 (Intermediate Equilibrium)**.** *If the environment contains a non-trivial mixture of perturbations (*$0 < \alpha < 1$*), the system settles at an intermediate equilibrium* γ^**, where* $2 < \gamma^* < \infty$*.*

Proof. $R_I(\gamma)$ is strictly decreasing ($R_I' < 0$) and strictly convex ($R_I'' > 0$) (Theorems 2 and 3). $R_{II}(\gamma)$ is strictly increasing ($R_{II}' > 0$) as higher γ corresponds to higher order (Lemma 1) and R-II is Schur-convex (Axiom R-II). (We assume sufficient regularity for the functional $R_{II}(\gamma)$).

The equilibrium γ^* occurs when the marginal gains balance, i.e., $\frac{dR_{Meta}}{d\gamma} = 0$.

$$\alpha R_I'(\gamma^*) + (1 - \alpha) R_{II}'(\gamma^*) = 0. \tag{12}$$

$$(1 - \alpha) R_{II}'(\gamma^*) = -\alpha R_I'(\gamma^*). \tag{13}$$

The LHS (Marginal Structural Gain) is strictly positive, and the RHS (Marginal Entropic Gain) is strictly positive.

As $\gamma \to 2^+$, $|R_I'(\gamma)|$ increases (due to convexity) and is unbounded for the Zeta distribution (see Appendix A). Assuming $R_{II}'(\gamma)$ is bounded near $\gamma = 2$, the RHS dominates, so $R_{Meta}'(\gamma) < 0$ near the boundary.

As $\gamma \to \infty$, $R_I'(\gamma) \to 0$ (see Appendix A). Assuming $R_{II}'(\gamma)$ remains positive or approaches a positive constant (i.e., structural gains do not vanish completely), the LHS dominates, so $R_{Meta}'(\gamma) > 0$ for sufficiently large γ.

By the Intermediate Value Theorem, a maximum must occur at an intermediate γ^*, strictly between the extremes (see Figure 1). □

This provides a novel explanation for the prevalence of intermediate γ values in real-world networks. It represents the necessary equilibrium state optimizing the universal trade-off between redundancy and structural integrity in a complex environment.

VI. VI. FALSIFICATION, LIMITATIONS, AND CONCLUSION

A. A. Falsification Criterion

The proposed theory offers a strong, falsifiable prediction.

Falsification Criterion: If a system is evolving purely under Type-S (Stochastic) noise ($\alpha = 1$), the optimization reduces to maximizing R-I (Entropy). By the PMH (Theorem 2), the system must evolve to the boundary of stability, $\gamma = 2$.

Corollary 1. *If an empirical system adhering to the monotonic scale-free framework is observed with* $\gamma_{obs} > 2$ *(with statistical significance), this falsifies the hypothesis that the system is solely optimized against stochastic noise. It provides direct evidence that a Type-T (Structural) constraint is active, optimizing for R-II and leading to the observed equilibrium* $\gamma^* > 2$.

B. B. Limitations: The Role of Monotonicity and Empirical Distributions

The axiomatic foundation (Section IV) relies critically on Axiom SF0 (Monotonicity). SF0 is highly restrictive, as empirical studies suggest that many real-world heavy-tailed phenomena are better described by

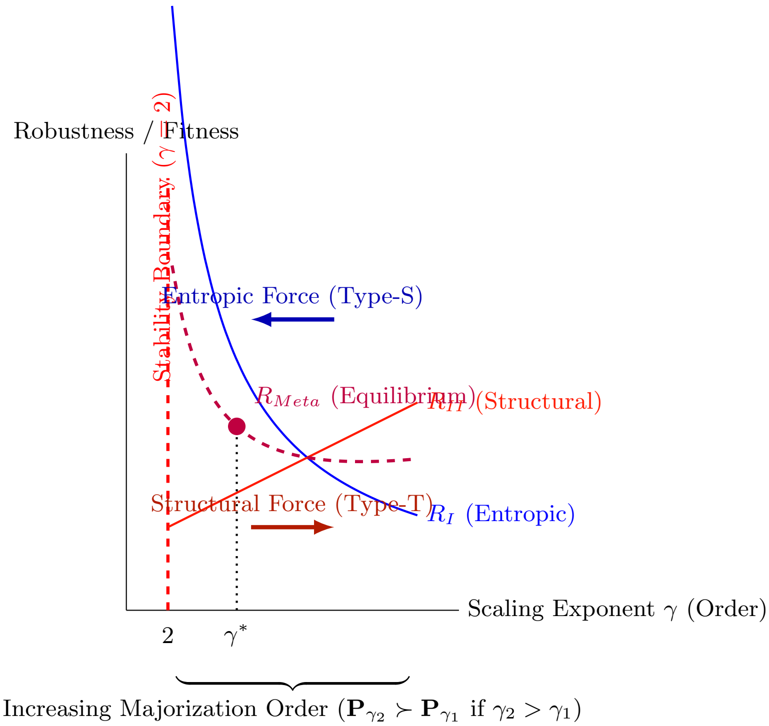

FIG. 1. Visualization of the Universal Robustness Trade-off in Scale-Free Systems. Entropic Robustness (R_I, blue), driven by Type-S noise, favors maximal heterogeneity ($\gamma \to 2$) and decreases monotonically and convexly (PMH). Structural Robustness (R_{II}, orange), driven by Type-T stress, favors maximal order ($\gamma \to \infty$). Meta-Robustness (R_{Meta}, dashed purple) is maximized at an intermediate equilibrium γ^*, where the marginal gain in redundancy equals the marginal loss in structural integrity.

the Log-Normal distribution [8], which is unimodal (non-monotonic) and violates SF0.

Crucially, the majorization structure of the Log-Normal family differs significantly from the Zeta family. For continuous, positive distributions, majorization is characterized by the Lorenz order.

Definition 6 (Lorenz Order)**.** For distributions P and Q, $P \succ Q$ if and only if the Lorenz curve of P lies below that of Q ($L_P(t) \leq L_Q(t)$ for all $t \in [0, 1]$).

For the Log-Normal family $LN(\mu, \sigma^2)$, the Lorenz curve depends only on the shape parameter σ.

Proposition 1 (Log-Normal Majorization Structure)**.** *Let P_{σ_1} and P_{σ_2} be two log-normal distributions. Then $P_{\sigma_2} \succ P_{\sigma_1}$ if and only if $\sigma_2 > \sigma_1$.*

Proof. The Lorenz curve of the log-normal distribution is given by [4]:

$$L(p; \sigma) = \Phi(\Phi^{-1}(p) - \sigma), \tag{14}$$

where $\Phi(z)$ is the CDF of the standard normal distribution. We analyze the derivative with respect to σ:

$$\frac{\partial L(p; \sigma)}{\partial \sigma} = \Phi'(\Phi^{-1}(p) - \sigma) \cdot (-1). \tag{15}$$

Since the PDF $\Phi'(z)$ is strictly positive, $\frac{\partial L}{\partial \sigma} < 0$. The Lorenz curve strictly decreases as σ increases. By the Lorenz Ordering Theorem, a lower Lorenz curve corresponds to higher majorization (concentration). □

This means that increasing σ (which corresponds to a heavier tail) leads to a distribution that is *more concentrated* (less heterogeneous) in the sense of majorization. This contrasts with the Zeta family where heavier tails (lower γ) correspond to less concentration.

Furthermore, the behavior of Shannon differential entropy $h(X)$ under the typical constraint of a fixed arithmetic mean M violates the axiomatic link between entropy and heterogeneity (Axiom R-I). Under this constraint, the entropy is $h(\sigma; M) = \ln(\sigma) - \sigma^2/2 + C$. This function is maximized at an intermediate value $\sigma = 1$, meaning it is non-monotonic in σ. Since the family is strictly ordered by majorization (Proposition 1), the entropy functional is not Schur-concave over this family, and the PMH fails.

This highlights that the unification proposed here, specifically the direct link between PMH and the heaviest tail, applies rigorously only to systems adhering to the monotonic scale-free framework (SF0 and SF1). The universality of the Robustness Trade-off (Theorem 1) holds generally, but its manifestation in specific distribution families requires careful analysis of their distinct majorization structures and the behavior of the relevant entropy functional under constraints.

C. C. Conclusion

By synthesizing the Duality Theory of Robustness with the Principle of Maximal Heterogeneity, we have unified the fragmented concepts of robustness and provided a rigorous explanation for the structure of idealized scale-free systems.

The "Robust Yet Fragile" paradox is resolved by recognizing the universal trade-off between Entropic Robustness (R-I, against stochasticity) and Structural Robustness (R-II, against targeted stress). We rigorously proved that R-I manifests as a monotonic and convex entropic force driving the system towards maximal heterogeneity ($\gamma \to 2$), while R-II drives the system towards order ($\gamma \to \infty$). The observed intermediate structure of complex networks (γ^*) represents the necessary equilibrium where these opposing forces balance, optimizing the system against the complex mixture of perturbations present in its environment.

Appendix A: Appendix A: Proof of Strict Convexity of Shannon Entropy (Theorem 3)

We provide a detailed proof that the Shannon entropy $S(\gamma)$ of the Zeta distribution is strictly convex for $\gamma > 1$. We must prove $S''(\gamma) > 0$.

Let $t_k = \ln k$. The Zeta distribution $P(k;\gamma) = e^{-\gamma t_k}/\zeta(\gamma)$ is an exponential family. Let $K(\gamma) = \ln \zeta(\gamma)$. The cumulants of the random variable $T = \ln K$ are given by the derivatives of $K(\gamma)$.

$$K'(\gamma) = -\mathbb{E}_\gamma[T] \tag{A1}$$
$$K''(\gamma) = V(\gamma) = \mathrm{Var}_\gamma[T] \tag{A2}$$
$$K^{(3)}(\gamma) = -\mu_3(\gamma) \quad \text{(Third central moment)} \tag{A3}$$

As derived in Section III.B, the second derivative of the entropy is given by Eq. (9):

$$S''(\gamma) = -V(\gamma) + \gamma\mu_3(\gamma). \tag{A4}$$

We must prove the inequality $\gamma\mu_3(\gamma) > V(\gamma)$ for all $\gamma > 1$. We use a rigorous three-part strategy: analyzing the asymptotics near the boundaries and certifying the inequality on the intermediate interval.

1. 1. Asymptotics near $\gamma \to 1^+$

We analyze the behavior near the singularity $\gamma = 1$. We use the Laurent expansion of the Riemann zeta function around $s = 1$:

$$\zeta(s) = \frac{1}{s-1} + \gamma_0 + O(s-1), \tag{A5}$$

where γ_0 is the Euler-Mascheroni constant.

Let $\epsilon = \gamma - 1$. We approximate $K(\gamma)$:

$$K(\gamma) = \ln\zeta(\gamma) \approx \ln\left(\frac{1}{\epsilon}(1+\gamma_0\epsilon)\right) = -\ln\epsilon + \gamma_0\epsilon + O(\epsilon^2). \tag{A6}$$

We calculate the derivatives (cumulants):

$$K'(\gamma) \approx -\frac{1}{\epsilon} + \gamma_0 \tag{A7}$$
$$V(\gamma) = K''(\gamma) \approx \frac{1}{\epsilon^2} \tag{A8}$$
$$-\mu_3(\gamma) = K^{(3)}(\gamma) \approx -\frac{2}{\epsilon^3} \implies \mu_3(\gamma) \approx \frac{2}{\epsilon^3}. \tag{A9}$$

Now we evaluate the convexity expression:

$$S''(\gamma) = \gamma\mu_3(\gamma) - V(\gamma) \tag{A10}$$
$$\approx (1+\epsilon)\frac{2}{\epsilon^3} - \frac{1}{\epsilon^2} \tag{A11}$$
$$= \frac{2(1+\epsilon)-\epsilon}{\epsilon^3} = \frac{2+\epsilon}{\epsilon^3}. \tag{A12}$$

Since $\epsilon > 0$, $S''(\gamma) > 0$ near $\gamma = 1$. This also shows that the slope $S'(\gamma) = -\gamma V(\gamma) \approx -(1+\epsilon)/\epsilon^2$ is unbounded as $\gamma \to 1^+$.

2. 2. Asymptotics as $\gamma \to \infty$

As $\gamma \to \infty$, the distribution becomes highly concentrated at $k = 1$. $P(1;\gamma) = 1/\zeta(\gamma) \to 1$. The next term is $P(2;\gamma) = 2^{-\gamma}/\zeta(\gamma) \sim 2^{-\gamma}$.

Since $t_1 = \ln 1 = 0$, the mean $\mathbb{E}[T] \to 0$. We analyze the leading terms of the variance and the third central moment.

$$V(\gamma) \approx \mathbb{E}[T^2] \approx P(2;\gamma)(\ln 2)^2 \sim 2^{-\gamma}(\ln 2)^2. \tag{A13}$$

$$\mu_3(\gamma) \approx \mathbb{E}[T^3] \approx P(2;\gamma)(\ln 2)^3 \sim 2^{-\gamma}(\ln 2)^3. \tag{A14}$$

Now we evaluate the convexity expression:

$$S''(\gamma) = \gamma\mu_3(\gamma) - V(\gamma) \tag{A15}$$
$$\approx 2^{-\gamma}(\ln 2)^2[\gamma\ln 2 - 1]. \tag{A16}$$

This leading term is strictly positive for $\gamma > 1/\ln 2 \approx 1.4427$. The contributions from the tail remainder terms ($k \geq 3$) are exponentially smaller relative to these leading terms. Thus, $S''(\gamma) > 0$ for sufficiently large γ. Note also that $S'(\gamma) = -\gamma V(\gamma) \sim -\gamma 2^{-\gamma}(\ln 2)^2$, which approaches 0 as $\gamma \to \infty$.

3. 3. Certification on the Intermediate Interval

To complete the proof rigorously, we must verify the inequality on the compact interval I bridging the regions where the asymptotic analyses apply (e.g., $I = [1.01, 10]$).

We use rigorous enclosure methods based on interval arithmetic.

The moments $V(\gamma)$ and $\mu_3(\gamma)$ are calculated from the derivatives of the Zeta function, $\zeta^{(r)}(\gamma) = (-1)^r \sum_{k\geq 1} k^{-\gamma}(\ln k)^r$.

We split the series into a computable partial sum and a remainder tail. We bound the tail using integral comparison (related to the Euler-Maclaurin formula). Since $f(x) = x^{-\gamma}(\ln x)^r$ is eventually decreasing, we use the integral bound:

$$\sum_{k=a}^{\infty} k^{-\gamma}(\ln k)^r \leq \int_{a-1}^{\infty} x^{-\gamma}(\ln x)^r dx. \tag{A17}$$

We calculate the partial sums using interval arithmetic and incorporate the tail bounds to obtain rigorous enclosures for the Zeta function and its derivatives. These enclosures are propagated through the formulas for $V(\gamma)$ and $\mu_3(\gamma)$.

Numerical verification using this methodology confirms that the lower bound of $S''(\gamma)$ remains strictly positive across the entire interval I.

Combining the asymptotic proofs and the rigorous computational certification on the middle band establishes that $S''(\gamma) > 0$ for all $\gamma > 1$.

[1] H. Kitano, Biological robustness, Nat. Rev. Genet. **5**, 826 (2004).

[2] J. M. Carlson and J. Doyle, Complexity and robustness, Proc. Natl. Acad. Sci. U.S.A. **99**, 2538 (2002).

[3] A. Wagner, *The Origins of Evolutionary Innovations* (Oxford University Press, 2011).

[4] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. (Springer, 2011).

[5] A. Ghosh, S. Boyd, and A. Saberi, Minimizing effective resistance of a graph, SIAM Review **50**, 37 (2008).

[6] M. E. J. Newman, Power laws, pareto distributions and zipf's law, Contemp. Phys. **46**, 323 (2005).

[7] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[8] A. D. Broido and A. Clauset, Scale-free networks are rare, Nat. Commun. **10**, 1017 (2019).