Module 7 — Information Flow

Systems Thinking

hbar.university

Shannon Entropy

Position in Knowledge Graph

This node formalizes Shannon entropy as the foundational quantitative measure of information.

It builds on: - Information-Theoretic Complexity (6.3)

It prepares for: - Channel Capacity (7.2) - Signal Propagation (7.3)

Shannon entropy provides the fundamental unit for analyzing information flow in systems.

Formal Statement / Definition / Construction

Let X be a discrete random variable with finite alphabet $\mathcal{X}$.

Let probability mass function:

$$p(x) = \mathbb{P}(X = x), \quad x \in \mathcal{X}.$$

Definition (Shannon Entropy)

The Shannon entropy of X is:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Logarithm base determines units: - base 2 $\rightarrow$ bits, - base $e \rightarrow$ nats.

Joint Entropy

For random variables X, Y:

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y).$$

Conditional Entropy

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y).$$

Equivalent identity:

$$H(X|Y) = H(X,Y) - H(Y).$$

Chain Rule

$$H(X,Y) = H(X) + H(Y|X).$$

Assumptions and Conditions

1. Alphabet $\mathcal{X}$ finite.
2. $0 \log 0$ defined as 0 by continuity.
3. All sums finite.
4. Logarithm base fixed.

Continuous variables require differential entropy (not treated here).

Proof or Mathematical Development

Maximum Entropy

Suppose $|\mathcal{X}| = m$.

Maximize entropy subject to:

$$\sum_x p(x) = 1.$$

Using Lagrange multipliers:

$$\mathcal{L} = -\sum_x p(x) \log p(x) + \lambda \left(\sum_x p(x) - 1 \right).$$

Take derivative:

$$\frac{\partial \mathcal{L}}{\partial p(x)} = -(1 + \log p(x)) + \lambda = 0.$$

Thus:

$$p(x) = e^{\lambda - 1}.$$

Since sum equals 1:

$$p(x) = \frac{1}{m}.$$

Therefore:

$$H(X) = \log m.$$

Uniform distribution maximizes entropy.

Non-Negativity

Since $0 \le p(x) \le 1$, we have $-\log p(x) \ge 0$, thus:

$$H(X) \ge 0.$$

Equality occurs if and only if X is deterministic.

Structural Role

Shannon entropy:

- Quantifies uncertainty.
- Measures dispersion of probability mass.
- Provides basis for information transmission limits.
- Connects randomness to complexity.

All information flow analysis builds upon entropy identities.

Structural Compression

Invariant

Entropy:

$$H(X) = -\sum_x p(x) \log p(x).$$

Mechanism

Logarithmic weighting of probability produces additive information measure.

Cross-System Echo

- Physics: thermodynamic entropy.
- Economics: uncertainty in markets.
- Biology: genetic diversity.
- Networks: communication capacity limits.

Channel Capacity

Position in Knowledge Graph

This node formalizes the maximum reliable information transmission rate through a noisy channel.

It builds on: - Shannon Entropy (7.1)

It prepares for: - Signal Propagation (7.3) - Information Integration (7.4)

Channel capacity defines the fundamental limit of information flow.

Formal Statement / Definition / Construction

Let $X \in \mathcal{X}$ be channel input and $Y \in \mathcal{Y}$ be channel output.

A **discrete memoryless channel (DMC)** is defined by conditional probabilities:

$$p(y|x), \quad x \in \mathcal{X},\ y \in \mathcal{Y}.$$

Joint distribution:

$$p(x, y) = p(x)p(y|x).$$

Definition (Mutual Information)

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

Definition (Channel Capacity)

The **channel capacity** C is:

$$C = \max_{p(x)} I(X; Y).$$

The maximization is over all input distributions $p(x)$.

Assumptions and Conditions

1. Channel is memoryless.
2. Alphabets $\mathcal{X}, \mathcal{Y}$ finite.
3. Logarithm base fixed.
4. Codewords of length n allowed for transmission.

Asymptotic reliability refers to block length $n \to \infty$.

Proof or Mathematical Development

Achievability (Sketch)

For block length n:

Generate 2^{nR} codewords independently from optimal input distribution $p^*(x)$.

Using typical-set decoding:

If

$$R < I(X;Y),$$

then probability of decoding error satisfies:

$$P_e^{(n)} \to 0 \quad \text{as } n \to \infty.$$

Thus rates below capacity are achievable.

Converse

Suppose reliable transmission at rate R exists.

Let M denote transmitted message.

Then:

$$H(M) = nR.$$

Using Fano's inequality:

$$H(M|Y^n) \le 1 + P_e^{(n)} nR.$$

If $P_e^{(n)} \to 0$,

$$H(M|Y^n) \to 0.$$

Thus:

$$nR = I(M;Y^n) + H(M|Y^n) \le I(M;Y^n) + \epsilon_n.$$

Since channel memoryless:

$$I(M;Y^n) \le \sum_{i=1}^{n} I(X_i;Y_i) \le nC.$$

Therefore:

$$R \le C.$$

Binary Symmetric Channel Example

Let crossover probability p.

Channel capacity:

$$C = 1 - H_2(p),$$

where binary entropy:

$$H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p).$$

Structural Role

Channel capacity:

- Defines maximum sustainable information flow.
- Separates feasible from infeasible communication regimes.
- Connects entropy to reliability.
- Applies to biological, economic, and technological systems.

Information transfer is bounded by mutual information.

Structural Compression

Invariant

Capacity:

$$C = \max_{p(x)} I(X;Y).$$

Mechanism

Reliable transmission possible iff transmission rate $R < C$.

Cross-System Echo

- Communications: bandwidth limits.
- Biology: genetic information transmission.
- Economics: market signal distortion.
- Networks: throughput bounds.

Signal Propagation

Position in Knowledge Graph

This node formalizes how signals and perturbations propagate through networks and dynamical systems.

It builds on: - Graph Foundations (4.1) - Shannon Entropy (7.1) - Channel Capacity (7.2)

It prepares for: - Information Integration (7.4) - Information Constraints in Systems (7.5)

Signal propagation connects topology, dynamics, and information flow.

Formal Statement / Definition / Construction

Let $G = (V, E)$ be a directed graph with adjacency matrix $A \in \mathbb{R}^{n\times n}$.

Consider linear network dynamics:

$$\dot{x}(t) = Ax(t) + Bu(t),$$

where:

- $x(t) \in \mathbb{R}^n$ is node state vector,
- $u(t) \in \mathbb{R}^m$ is external input,
- $B \in \mathbb{R}^{n\times m}$ is input matrix.

Impulse Response

Let $u(t) = \delta(t)e_k$ (impulse at node k).

Then response:

$$x(t) = e^{At}Be_k.$$

Thus propagation governed by matrix exponential e^{At}.

Path Expansion

Using series expansion:

$$e^{At} = \sum_{j=0}^{\infty} \frac{(At)^j}{j!}.$$

Hence:

$$x(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!} A^j Be_k.$$

Term A^j corresponds to propagation along length-j paths.

Assumptions and Conditions

1. A constant (time-invariant).
2. System is well-posed (matrix exponential defined).
3. Network finite.
4. For probabilistic propagation, assume noise-free linear channel.

Nonlinear propagation requires separate treatment.

Proof or Mathematical Development

Propagation Speed in Linear Diffusion

Consider consensus-type dynamics:

$$\dot{x} = -Lx,$$

where L is graph Laplacian.

Solution:

$$x(t) = e^{-Lt}x(0).$$

Since Laplacian eigenvalues satisfy:

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n,$$

convergence rate determined by spectral gap:

$$\lambda_2.$$

Thus information mixing time approximately:

$$\tau \sim \frac{1}{\lambda_2}.$$

Entropy Flow Interpretation

Let $X(t)$ denote random state.

Entropy evolution:

$$\frac{d}{dt}H(X(t)) = -\sum_x \dot{p}(x,t)\log p(x,t).$$

Under diffusion-like dynamics, entropy typically increases toward equilibrium.

Thus signal disperses over network.

Structural Role

Signal propagation:

- Links adjacency structure to dynamic influence.
- Determines speed of synchronization and diffusion.
- Connects spectral properties to mixing times.
- Provides quantitative model for information spread.

Propagation is governed by eigenstructure of network operator.

Structural Compression

Invariant

Signal response governed by:

$$x(t) = e^{At}x(0).$$

Mechanism

Matrix exponential expansion encodes path-based propagation:

$$e^{At} = \sum_{j=0}^{\infty} \frac{A^j t^j}{j!}.$$

Cross-System Echo

- Physics: heat equation diffusion.
- Economics: shock transmission across markets.
- Biology: neural signal propagation.
- Networks: rumor spreading dynamics.

Information Integration

Position in Knowledge Graph

This node formalizes how distributed components combine information into coherent collective structure.

It builds on: - Shannon Entropy (7.1) - Signal Propagation (7.3)

It prepares for: - Information Constraints in Systems (7.5)

Information integration quantifies the degree to which system components form a unified informational structure.

Formal Statement / Definition / Construction

Let $X = (X_1, \dots, X_n)$ be a collection of discrete random variables.

Joint entropy:

$$H(X) = H(X_1, \dots, X_n).$$

Define total independent entropy:

$$H_{\text{ind}} = \sum_{i=1}^{n} H(X_i).$$

Definition (Multi-Information)

The **multi-information** (total correlation) is:

$$I_{\text{tot}}(X) = \sum_{i=1}^{n} H(X_i) - H(X_1, \dots, X_n).$$

Equivalently:

$$I_{\text{tot}}(X) = D\left(p(x_1, \dots, x_n) \parallel \prod_{i=1}^{n} p(x_i)\right),$$

where $D(\cdot\|\cdot)$ is KL divergence.

Definition (Integrated Information — Structural Form)

For a bipartition $\{A, B\}$ of indices:

$$\Phi_{A|B} = I(X_A; X_B).$$

System exhibits integration if:

$$\min_{A|B} \Phi_{A|B} > 0.$$

This measures minimal mutual dependence across partitions.

Assumptions and Conditions

1. Variables discrete with finite support.
2. Joint distribution well-defined.
3. Partition is nontrivial (both subsets nonempty).
4. Logarithm base fixed.

No claim about specific physical interpretation beyond informational dependence.

Proof or Mathematical Development

Non-Negativity

Since:

$$I_{\mathrm{tot}}(X) = D\left(p(x) \parallel \prod_i p(x_i)\right),$$

and KL divergence satisfies:

$$D(p\|q) \geq 0,$$

we have:

$$I_{\mathrm{tot}}(X) \geq 0.$$

Equality holds iff variables are mutually independent.

Decomposition

For $n = 2$:

$$I_{\mathrm{tot}}(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2).$$

For larger n, multi-information generalizes pairwise mutual information.

Interpretation in Networked Systems

If network state X arises from coupling:

$$\dot{x} = F(x),$$

then interaction induces statistical dependence among components.

Thus:

$$I_{\text{tot}}(X) > 0$$

reflects informational integration induced by coupling.

Structural Role

Information integration:

- Quantifies collective informational cohesion.
- Distinguishes independent subsystems from unified systems.
- Connects entropy to network coupling.
- Provides bridge to distributed computation and consciousness theories (formal aspects only).

It measures how much the whole exceeds sum of parts in informational terms.

Structural Compression

Invariant

Multi-information:

$$I_{\text{tot}}(X) = \sum_i H(X_i) - H(X).$$

Mechanism

KL divergence between joint distribution and product of marginals.

Cross-System Echo

- Physics: correlation functions.
- Biology: coordinated neural activity.
- Economics: systemic dependence.
- Networks: coupling-induced coherence.

Information Constraints in Systems

Position in Knowledge Graph

This node formalizes fundamental limits imposed by information theory on control, coordination, and adaptation.

It builds on: - Channel Capacity (7.2) - Information Integration (7.4)

It prepares for: - Institutions as Constraint Systems (8.1)

Information constraints determine what systems can know, transmit, and coordinate.

Formal Statement / Definition / Construction

Let a dynamical system be:

$$\dot{x}(t) = f(x(t), u(t)),$$

where:

- $x(t) \in \mathbb{R}^n$,
- $u(t)$ is control input,
- control is transmitted through a noisy channel.

Let communication channel have capacity:

$$C = \max_{p(x)} I(X; Y).$$

Theorem (Data-Rate Constraint for Stabilization — Informal Statement)

For a linear unstable system:

$$\dot{x} = Ax + Bu,$$

with unstable eigenvalues λ_i satisfying $\mathrm{Re}(\lambda_i) > 0$,

a necessary condition for stabilization over a noiseless digital channel of capacity C is:

$$C \geq \sum_{\mathrm{Re}(\lambda_i) > 0} \mathrm{Re}(\lambda_i).$$

This bound is known as the data-rate theorem.

Assumptions and Conditions

1. System is linear time-invariant.
2. Control signals transmitted over digital channel.
3. Channel has finite capacity C.
4. Stabilization means:

$$\lim_{t \to \infty} x(t) = 0.$$

5. Eigenvalues counted with multiplicity.

Full proof requires stochastic control theory; only structural derivation given.

Proof or Mathematical Development

Instability Growth Rate

For unstable mode:

$$x_i(t) \sim e^{\lambda_i t}.$$

To counteract instability, controller must receive state information at sufficient rate.

Information required to describe state with precision ϵ scales as:

$$\log \frac{1}{\epsilon}.$$

If state grows as $e^{\lambda t}$, precision requirement grows as:

$$\lambda t.$$

Thus minimum information rate required:

$$\frac{d}{dt}(\lambda t) = \lambda.$$

Summing over unstable modes yields lower bound:

$$C \geq \sum_{\mathrm{Re}(\lambda_i) > 0} \mathrm{Re}(\lambda_i).$$

If channel rate is below this threshold, controller cannot transmit corrective information fast enough.

Information–Integration Constraint

Let system partitioned into subsystems A, B.

If:

$$I(X_A; X_B) = 0,$$

then subsystems evolve independently.

Thus coordination across partitions requires:

$$I(X_A; X_B) > 0.$$

Limited channel capacity bounds achievable integration:

$$I(X_A; X_B) \leq C.$$

Structural Role

Information constraints:

- Bound achievable control performance.
- Limit coordination across subsystems.
- Connect eigenvalue instability to data rate.
- Formalize tradeoff between dynamics and communication.

Systems cannot overcome instability without sufficient informational throughput.

Structural Compression

Invariant

Minimum rate for stabilization:

$$C \geq \sum_{\mathrm{Re}(\lambda_i)>0} \mathrm{Re}(\lambda_i).$$

Mechanism

Instability generates exponential growth of uncertainty requiring proportional information rate.

Cross-System Echo

- Control Theory: data-rate theorem.
- Economics: limits of information aggregation.
- Biology: signaling bandwidth constraints.
- Networks: throughput limits in distributed coordination.