Facets of Information*

W. Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907

May 25, 2009

AofA and **IT** logos

POZNAN 2009

---

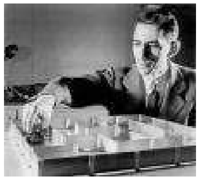
*Thanks to Y. Choi, Purdue, P. Jacquet, INRIA, France and J. Konorski, Gdansk, Poland.

Outline

1. Shannon Information
2. Beyond Shannon (space, time, structure, semantics)
3. What is Information? (this is the question)
4. **Learnable Information** – Fundamental Limit of Information Extraction
5. **Transfer of Spatio-Temporal Information** – Speed of Information
6. **Structural Information**
   (a) Graphical Compression and Fundamental Limit
   (b) Remote Spanner – Topology Compression in Ad-hoc Networks
7. Science of Information

Shannon Information

In 1948 C. Shannon created a powerful and beautiful **theory of information** that served as the backbone to nowadays digital communications.



C. Shannon:
Shannon information quantifies the extent to which a recipient of data can reduce its statistical uncertainty.

Some aspects of Shannon information:

objective: statistical ignorance of the recipient;
statistical uncertainty of the recipient.

cost: # binary decisions to describe E;
= $-\log P(E)$; $P(E)$ being the probability of E.

Context: "semantic aspects of communication are irrelevant"

Self-information for E_i: $\mathrm{info}(E_i) = -\log P(E_i)$.
Average information: $H(P) = -\sum_i P(E_i)\log P(E_i)$
Entropy of $X = \{E_1, \ldots\}$: $H(X) = -\sum_i P(E_i)\log P(E_i)$
Mutual Information: $I(X;Y) = H(Y) - H(Y|X)$, (faulty channel).

Three Jewels of Shannon

Theorem 1. (**Shannon 1948; Lossless Data Compression**)

compression bit rate $\geq$ source entropy $H(X)$.

(There exists a codebook of size 2^{nR} of universal codes of length n with

$$R > H(X)$$

and probability of error smaller than any $\varepsilon > 0$.)

Theorem 2. (**Shannon 1948; Channel Coding**)
In Shannon's words:

It is possible to send information at the capacity through the channel with as small a frequency of errors as desired by proper (**long**) encoding. This statement is not true for any rate greater than the capacity.

(The maximum codebook size $N(n, \varepsilon)$ for codelength n and error probability ε is asymptotically equal to: $N(n, \varepsilon) \sim 2^{nC}$.)

Theorem 3. (**Shannon 1948; Lossy Data Compression**).
For distortion level D:

lossy bit rate $\geq$ rate distortion function $R(D)$.

Outline Update

1. Shannon Information
2. Beyond Shannon
3. What is Information?
4. Learnable Information – Fundamental Limits
5. Spatio-Temporal Information
6. Structural Information
7. Science of Information

Beyond Shannon

Participants of **2005/2008 Information Beyond Shannon** workshops realize:

Delay: Delay incurred is a issue not adequately addressed in information theory (e.g., information arriving late maybe useless).

Space: In networks the spatially distributed components raise fundamental issues of limitations in information exchange since the available resources must be shared, allocated and re-used. **Information** is exchanged in space and time for decision making, thus timeliness of information delivery along with reliability and complexity constitute the basic objective.

Structure: We still lack measures and meters to define and appraise the amount of information embodied in structure and organization.

Semantics. In many scientific contexts, one is interested in signals, without knowing precisely what these signals represent. Is there a general way to account for the meaning of signals in a given context?

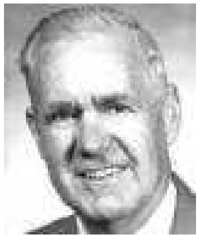
Limited Computational Resources: In many scenarios, information is limited by available computational resources (e.g., cell phone, living cell).

Representation-invariant of information. How to know whether two representations of the same information are information equivalent?

Outline Update

1. Shannon Information
2. Beyond Shannon
3. What is Information?
4. Learnable Information – Fundamental Limits
5. Spatio-Temporal Information
6. Structural Information
7. Science of Information

What is Information?



C. F. Von Weizsäcker:
"Information is only that which produces information" (relativity).
"Information is only that which is understood" (rationality)
"Information has no absolute meaning".

Informally Speaking: A piece of data carries **information** if it can impact a **recipient's ability** to achieve the **objective** of some **activity** in a given **context** within limited available resources.

Using the event-driven paradigm, we may formally define:

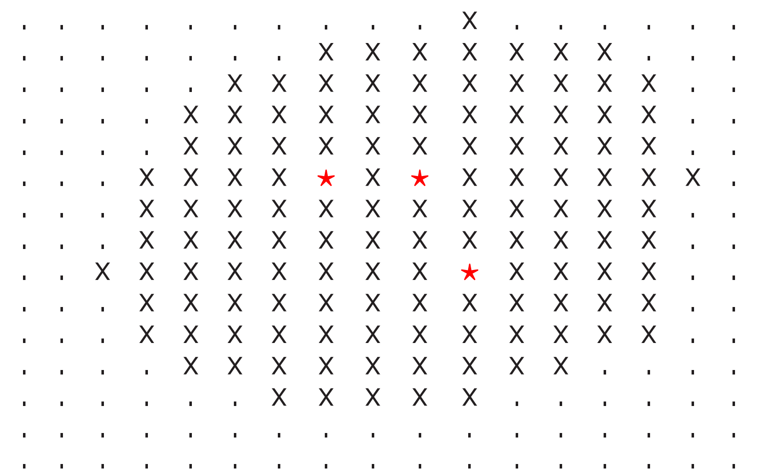
Definition 1 (Konorski, W.S., 2006). *The* **amount of information** *(in a faultless scenario)* info*(E) carried by the event E in the context C as measured for a system with the rules of conduct R is*

$$\mathrm{info}_{R,C}(E) = \mathbf{cost}[\mathrm{objective}_R(C(E)), \mathrm{objective}_R(C(E) + E)]$$

where the **cost** *(weight, distance) is taken according to the ordering of points in the space of objectives.*

Example: Distributed Information

1. In an N-threshold secret sharing scheme, N subkeys of the decryption key roam among $A \times A$ stations.

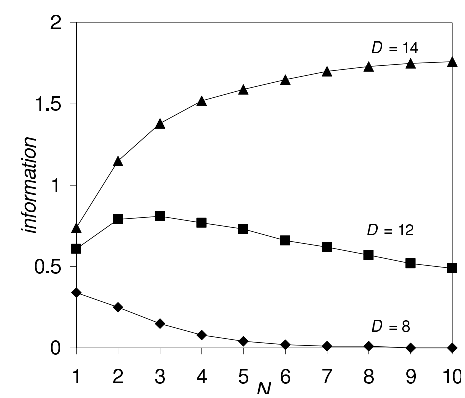


2. By protocol P a station has access:

- only it sees all N subkeys.
- it is within a distance D from all subkeys.

3. Assume that the larger N, the more valuable the secrets.
We define the amount of information as
(cf. J. Konorski and W.S.)

info= $N \times \{\#\ \text{of stations having access}\}$.



Outline Update

1. Shannon Information
2. What is Information?
3. Beyond Shannon
4. Learnable Information – Fundamental Limits
5. Spatio-Temporal Information
6. Structural Information
7. Science of Information

Learnable Information

How much useful information can be extracted from a given data set (then we can ask about amount of knowledge in Google database)?.

Learnable Information

How much useful information can be extracted from a given data set (then we can ask about amount of knowledge in Google database)?.

Learnable Information:

1. For a fixed n a sequence $x^n = x_1 \ldots x_n$ is given to us.

2. Summarizing Property: Let S be a set representing useful information, structure, regularity or summarizing properties of x^n
(e.g., S could be the number of 1 in x^n).

3. We can represent x^n be describing the set S – we denote it by $I(S)$ and call it **useful information** – and position of x^n in S that requires $\log |S|$ bits and represents its **complexity** of x^n.

4. Choose $\hat{S}$ with the smallest $I(S)$; call $I(\hat{S})$ the **learnable information**.

Learnable Information

How much useful information can be extracted from a given data set (then we can ask about amount of knowledge in Google database)?.

Learnable Information:

1. For a fixed n a sequence $x^n = x_1 \dots x_n$ is given to us.

2. Summarizing Property: Let S be a set representing useful information, structure, regularity or summarizing properties of x^n
(e.g., S could be the number of 1 in x^n).

3. We can represent x^n be describing the set S – we denote it by $I(S)$ and call it **useful information** – and position of x^n in S that requires $\log |S|$ bits and represents its **complexity** of x^n.

4. Choose $\hat{S}$ with the smallest $I(S)$; call $I(\hat{S})$ the **learnable information**.

Kolmogorov Information: Define

$$K(x^n) = K(\hat{S}) + \log |\hat{S}|.$$

Example: For x^n being a binary sequence, let S be the type of x^n that requires $K(\hat{S}) = \frac{1}{2} \log n$ bits, and

$$\log |S| = \log \binom{n}{k} = nH(n/k) \text{ bits.}$$

Computable Learnable Information (Rissanen MDL)

Statistically Learnable/Useful Information

1. $\mathcal{M}_k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions.
Let $\hat{\theta}(x^n) = \arg\max_{\theta\in\Theta} - \log P_\theta(x^n)$ be the ML estimator.

Computable Learnable Information (Rissanen MDL)

Statistically Learnable/Useful Information

1. $\mathcal{M}_k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions.
Let $\hat{\theta}(x^n) = \arg\max_{\theta \in \Theta} - \log P_\theta(x^n)$ be the ML estimator.

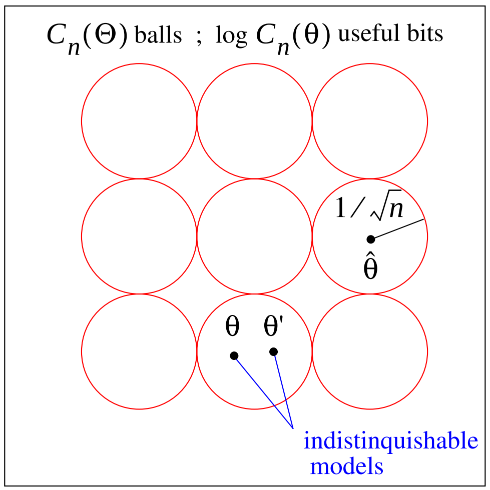
2. The summarizing property is here represented by $\mathcal{M}_k$.

Computable Learnable Information (Rissanen MDL)

Statistically Learnable/Useful Information

1. $\mathcal{M}_k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions.
Let $\hat{\theta}(x^n) = \arg\max_{\theta\in\Theta} - \log P_\theta(x^n)$ be the ML estimator.

2. The summarizing property is here represented by $\mathcal{M}_k$.



3. Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$ are indistinguishable (have the same useful information) if the ML estimator $\hat{\theta}$ with high probability declares both models are the same (i.e., θ and θ' are close).

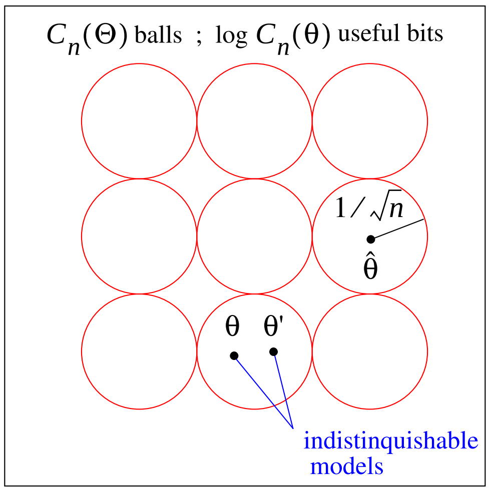
4. The number of distinguishable distributions (i.e, $\hat{\theta}$) $C_n(\Theta)$ summarizes then the learnable information and we denote it as $I(\Theta) = \log_2 C_n(\Theta)$.

Computable Learnable Information (Rissanen MDL)

Statistically Learnable/Useful Information

1. $\mathcal{M}_k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions.
Let $\hat{\theta}(x^n) = \arg\max_{\theta\in\Theta} - \log P_\theta(x^n)$ be the ML estimator.

2. The summarizing property is here represented by $\mathcal{M}_k$.



3. Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$
are indistinguishable (have the same useful information)
if the ML estimator $\hat{\theta}$ with high probability
declares both models are the same
(i.e., θ and θ' are close).

4. The number of distinguishable distributions (i.e, $\hat{\theta}$)
$C_n(\Theta)$ summarizes then the learnable information
and we denote it as $I(\Theta) = \log_2 C_n(\Theta)$.

5. Consider the following expansion of the Kullback-Leibler (KL) divergence

$$D(P_{\hat{\theta}}||P_\theta) := \mathbf{E}[\log P_{\hat{\theta}}(X^n)] - \mathbf{E}[\log P_\theta(X^n)] \sim \frac{1}{2}(\theta-\hat{\theta})^T I(\hat{\theta})(\theta-\hat{\theta}) \asymp d_I^2(\theta,\hat{\theta})$$

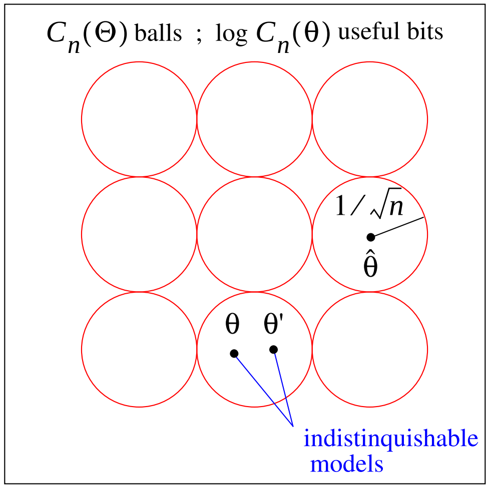
where $I(\theta) = \{I_{ij}(\theta)\}_{ij}$ is the *Fisher information matrix* and $d_I(\theta,\hat{\theta})$ is a rescaled Euclidean distance known as Mahalanobis distance.

Computable Learnable Information (Rissanen MDL)

Statistically Learnable/Useful Information

1. $\mathcal{M}_k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions. Let $\hat{\theta}(x^n) = \arg\max_{\theta\in\Theta} - \log P_\theta(x^n)$ be the ML estimator.

2. The summarizing property is here represented by $\mathcal{M}_k$.



3. Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$ are indistinguishable (have the same useful information) if the ML estimator $\hat{\theta}$ with high probability declares both models are the same (i.e., θ and θ' are close).

4. The number of distinguishable distributions (i.e, $\hat{\theta}$) $C_n(\Theta)$ summarizes then the learnable information and we denote it as $I(\Theta) = \log_2 C_n(\Theta)$.

5. Consider the following expansion of the Kullback-Leibler (KL) divergence

$$D(P_{\hat{\theta}}||P_\theta) := \mathbf{E}[\log P_{\hat{\theta}}(X^n)] - \mathbf{E}[\log P_\theta(X^n)] \sim \frac{1}{2}(\theta - \hat{\theta})^T I(\hat{\theta})(\theta - \hat{\theta}) \asymp d_I^2(\theta, \hat{\theta})$$

where $I(\theta) = \{I_{ij}(\theta)\}_{ij}$ is the *Fisher information matrix* and $d_I(\theta, \hat{\theta})$ is a rescaled Euclidean distance known as Mahalanobis distance.

5. Balasubramanian proved that the number of distinguishable balls $C_n(\Theta)$ of radius $O(1/\sqrt{n})$ is equal to (i.e., minimax maximal regret)

$$I(\Theta) = \log C_n(\Theta) = \inf_{\theta\in\Theta} \max_{x^n} \log \frac{P_{\hat{\theta}}}{P_\theta} = \log \sum_{x^n} P_{\hat{\theta}}(x^n).$$

Memoryless Sources

Consider the **minimax regret**=**useful information** for memoryless sources of m-ary alphabet, thus with $k = m - 1$. Observe that

$$C_n(\Theta) = \sum_{x_1^n} \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1,\ldots,k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}.$$

which becomes:

$$C_n(\Theta) = \frac{n!}{n^n} \sum_{k_1+\cdots+k_m=n} \frac{k_1^{k_1}}{k_1!} \cdots \frac{k_m^{k_m}}{k_m!}.$$

Using tree-generating functions and analytic information theory tools, we find (cf. Clarke & Barron, 1990, W.S., 1998)

$$\begin{aligned} I_n(\Theta) &= \frac{m-1}{2}\log\left(\frac{n}{2}\right) + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + \frac{\Gamma(\frac{m}{2})m}{3\Gamma(\frac{m}{2}-\frac{1}{2})} \cdot \frac{\sqrt{2}}{\sqrt{n}} \\ &+ \left(\frac{3+m(m-2)(2m+1)}{36} - \frac{\Gamma^2(\frac{m}{2})m^2}{9\Gamma^2(\frac{m}{2}-\frac{1}{2})}\right) \cdot \frac{1}{n} + \cdots \end{aligned}$$

Outline Update

1. Shannon Information
2. What is Information?
3. Beyond Shannon
4. Learnable Information – Fundamental Limits
5. Spatio-Temporal Information – Speed of Information
6. Structural Information
7. Science of Information

Transfer of Information in Ubiquitous Networks

Information Theory, born 50 years ago, needs a recharge if it is to meet new challenges of ubiquitous networks.

Fundamental New Problems:

1. Future networks will transport information not data.

2. Information is only useful when delivered in a timely fashion (e.g., new resource scheduling in inherently unreliable wireless environment).

3. To design scalable networks, node must cooperate (e.g., interference can be turn into useful signals thru distributive multiantenna processing; mobility may diffuse traffic but will cause large delays).

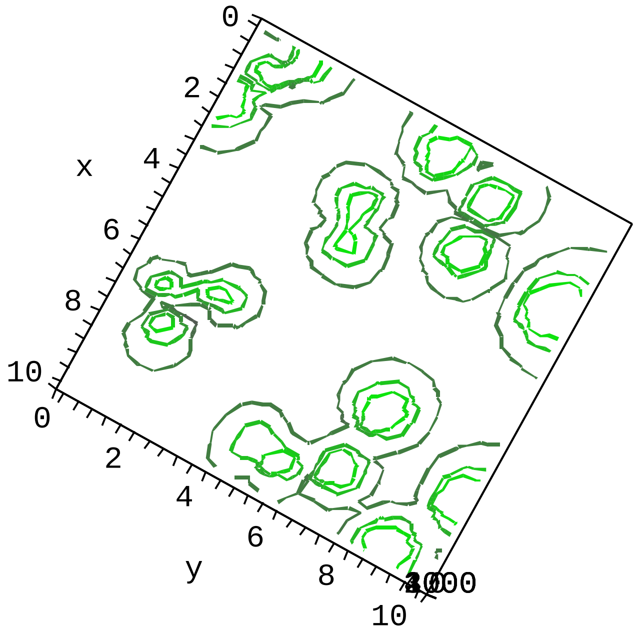
4. New Information Theory of dependence is needed to design more energy efficient communication (i.e., how fast? at what cost?).

5. To turn it into reality, we must seriously consider:
selfishness (it is in each node's self-interest to cooperate);
channel capacity (to turn interference into useful signals);
delay (mobility can diffuse traffic for large delays).

Speed of Information

Based on P. Jacquet, B. Mans and G. Rodolakis, ISIT, 2008

Intermittently Connected Mobile Networks (ICN):



1. Nodes move in space with uniform density $\nu > 0$.

2 Nodes do random walks with speed v and turn rate τ.

3. Connectivity is achieved in a unit disk.

4. Radio propagation speed is infinite.

Problem statement:
At time $t = 0$ a node at the origin broadcasts a beacon and nodes retransmit beacon immediately to neighbors in the ICN network.

Question: At what time T node at distance L from the origin will receive the beacon? **Propagation speed** is $\frac{L}{T}$.

Journey Analysis Through the Laplace Transform

The beacon undergoes a journey C from the origin to some point $\mathbf{z}$. Let $\mathbf{z}(C)$ be the destination point reached at time $t(C)$.

Let $p(\mathbf{z}, t)$ be the space-time density of paths C that reaches location $\mathbf{z}(C)$ at time t.

Journey Analysis Through the Laplace Transform

The beacon undergoes a journey C from the origin to some point $\mathbf{z}$. Let $\mathbf{z}(C)$ be the destination point reached at time $t(C)$.

Let $p(\mathbf{z}, t)$ be the space-time density of paths C that reaches location $\mathbf{z}(C)$ at time t.

(Probabilistic) Information speed is the smallest σ_0 such that for all $\sigma > \sigma_0$

$$\lim p\left(\mathbf{z}, \frac{|\mathbf{z}|}{\sigma}\right) = 0.$$

Example: For $p(\mathbf{z}, t) = O(\exp(-A|\mathbf{z}| + Bt + C))$, then $\sigma_0 = B/A$.

Journey Analysis Through the Laplace Transform

The beacon undergoes a journey C from the origin to some point $\mathbf{z}$. Let $\mathbf{z}(C)$ be the destination point reached at time $t(C)$.

Let $p(\mathbf{z}, t)$ be the space-time density of paths C that reaches location $\mathbf{z}(C)$ at time t.

(Probabilistic) Information speed is the smallest σ_0 such that for all $\sigma > \sigma_0$

$$\lim p\left(\mathbf{z}, \frac{|\mathbf{z}|}{\sigma}\right) = 0.$$

Example: For $p(\mathbf{z}, t) = O(\exp(-A|\mathbf{z}| + Bt + C))$, then $\sigma_0 = B/A$.

The bivariate Laplace transform of $\mathbf{z}(C)$ and $t(C)$ is

$$\mathbf{E}[\exp(-\zeta\mathbf{z}(\mathbf{C}) - \theta t(C))] = \frac{1}{D(|\zeta|, \theta)}$$

with

$$D(\rho, \theta) = \sqrt{(\theta + \tau)^2 - \rho^2 v^2} - \frac{4\pi\nu v I_0(\rho)}{1 - \pi\nu\frac{2}{\rho}I_1(\rho)}$$

with I_k modified Bessel functions of order k.

In order to find $p(\mathbf{z}, t)$ one needs to inverse the Laplace transform through the saddle point method.

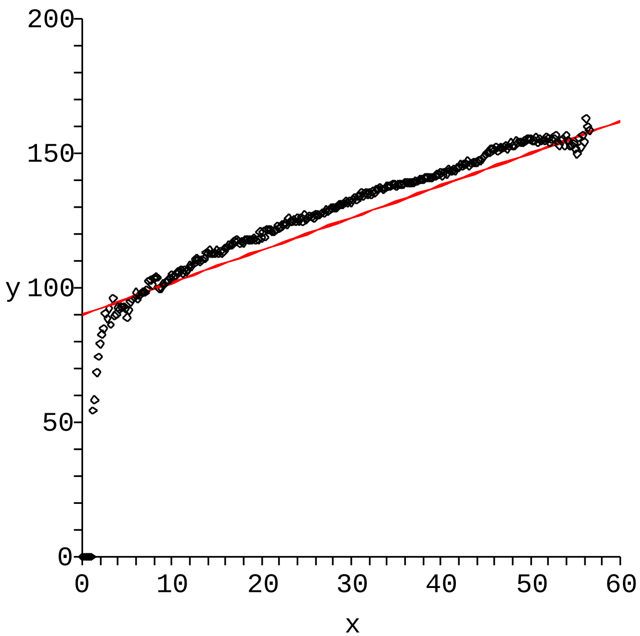
Main Result on Information Speed

Let $\mathcal{K}$ be the set (ρ, θ) of all roots of $D(\rho, \theta) = 0$.

Theorem 1 (Jacquet, et. al., 2008). *The information speed is not greater than the smallest ratio*

$$\frac{\theta}{\rho}$$

where (ρ, θ) belongs to $\mathcal{K}$.



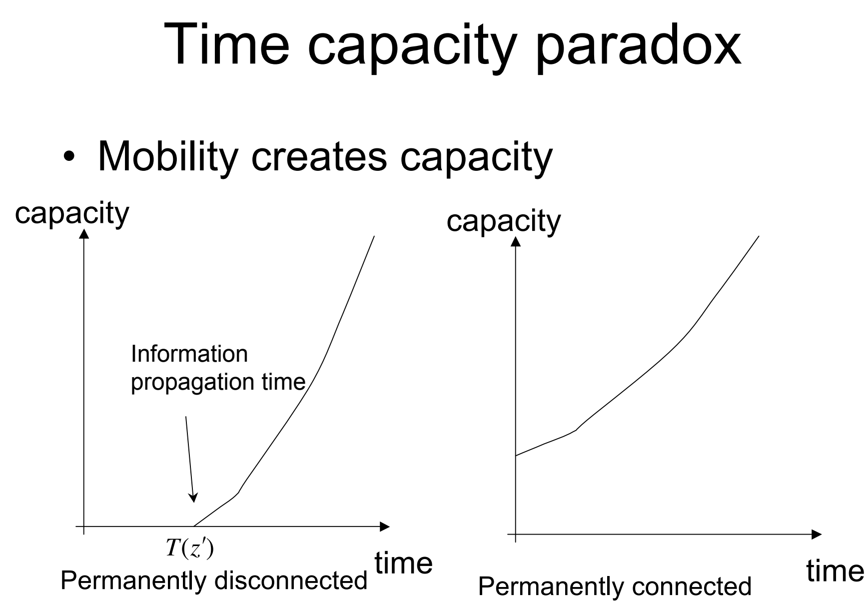


Figure 1: (LEFT) Time versus distance for $\nu = 0.1$, $v = 1$ and $\tau = 0.25$; (RIGHT) Impact on the network capacity.

Outline Update

1. Shannon Information
2. What is Information?
3. Beyond Shannon
4. Learnable Information – Fundamental Limits
5. Spatio-Temporal Information
6. Structural Information
   (a) Graphical Compression and Fundamental Limit
   (b) Remote Spanner – Topology Compression in Ad-hoc Networks
7. Science of Information

Structural Information

F. Brooks, jr. (JACM, 50, 2003, "Three Great Challenges for . . . CS "):

We have **no theory** that gives us a metric for the Information embodied in **structure**. This is the most fundamental gap in the theoretical underpinning of Information. . . . A young information theory scholar willing to spend years on a deeply fundamental problem need look no further."

Random graph model:

A graph process $\mathcal{G}$ generates a set of graphs $G = (V, E)$, that produces a probability distribution on graphs.

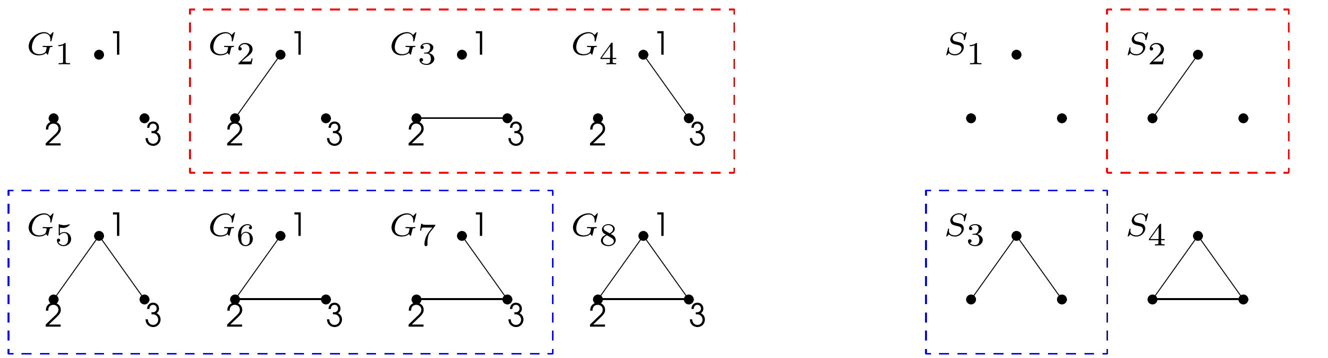
The (descriptive) entropy of a random (labeled) graph process $\mathcal{G}$ is defined as

$$H_{\mathcal{G}} = \mathbf{E}[-\log P(G)] = -\sum_{G \in \mathcal{G}} P(G) \log P(G),$$

where $P(G)$ is the probability of a graph G.

Random Structure Model

A random structure model is defined for an unlabeled version. Some labeled graphs have the same structure.



Random Graph Model   Random Structure Model

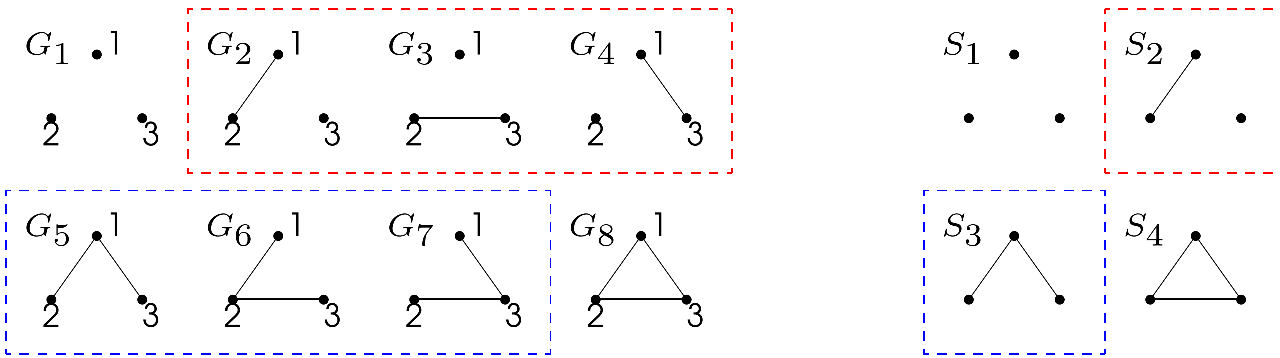
The probability of a structure S is

$$P(S) = N(S) \cdot P(G)$$

$N(S)$ is the number of different labeled graphs having the same structure.

Random Structure Model

A random structure model is defined for an unlabeled version. Some labeled graphs have the same structure.



Random Graph Model Random Structure Model

The probability of a structure S is

$$P(S) = N(S) \cdot P(G)$$

$N(S)$ is the number of different labeled graphs having the same structure.

The entropy of a random structure $\mathcal{S}$ can be defined as

$$H_{\mathcal{S}} = \mathbf{E}[-\log P(S)] = -\sum_{S \in \mathcal{S}} P(S) \log P(S),$$

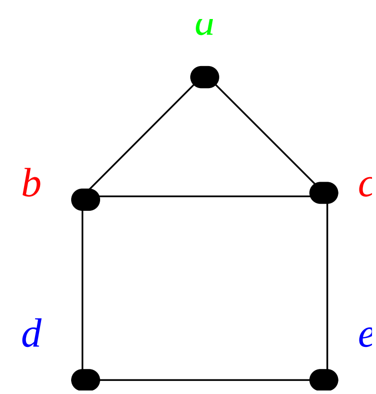
where the summation is over all distinct structures.

Relationship between $H_{\mathcal{G}}$ and $H_{\mathcal{S}}$

Two labeled graphs G_1 and G_2 are called *isomorphic* if and only if there is a one-to-one map from $V(G_1)$ onto $V(G_2)$ which preserves the adjacency.

Graph Automorphism:

For a graph G its automorphism is adjacency preserving permutation of vertices of G.



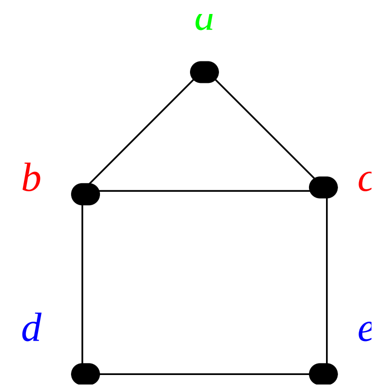
The collection $\mathrm{Aut}(G)$ of all automorphism of G is called *the automorphism group* of G.

Relationship between $H_{\mathcal{G}}$ and $H_{\mathcal{S}}$

Two labeled graphs G_1 and G_2 are called *isomorphic* if and only if there is a one-to-one map from $V(G_1)$ onto $V(G_2)$ which preserves the adjacency.

Graph Automorphism:

For a graph G its automorphism is adjacency preserving permutation of vertices of G.



The collection $\mathrm{Aut}(G)$ of all automorphism of G is called *the automorphism group* of G.

Lemma 1. *If all isomorphic graphs have the same probability, then*

$$H_{\mathcal{S}} = H_{\mathcal{G}} - \log n! + \sum_{S \in \mathcal{S}} P(S) \log |\mathrm{Aut}(S)|,$$

where $\mathrm{Aut}(S)$ *is the automorphism group of* S.

Proof idea: Using the fact that

$$N(S) = \frac{n!}{|\mathrm{Aut}(S)|}.$$

Erdös-Rényi Graph Model

Our random structure model is the unlabeled version of the binomial random graph model known also as the Erdös and Rényi model.

The binomial random graph model $\mathcal{G}(n, p)$ generates graphs with n vertices, where edges are chosen independently with probability p.

If a graph G in $\mathcal{G}(n, p)$ has k edges, then $P(G) = p^k q^{\binom{n}{2}-k}$, where $q = 1 - p$.

Theorem 2 (Y. Choi and W.S., 2008). *For large n and all p satisfying $\frac{\ln n}{n} \ll p$ and $1 - p \gg \frac{\ln n}{n}$ (i.e., the graph is connected w.h.p.),*

$$H_{\mathcal{S}} = \binom{n}{2} h(p) - \log n! + o(1),$$

where $h(p) = -p \log p - (1 - p) \log (1 - p)$ is the entropy rate. Thus

$$H_{\mathcal{S}} = \binom{n}{2} h(p) - n \log n + n \log e - \frac{1}{2} \log n - \frac{1}{2} \log (2\pi) + o(1).$$

Proof idea: **1**. $H_{\mathcal{S}} = H_{\mathcal{G}} - \log n! + \sum_{S \in \mathcal{S}} P(S) \log |\mathrm{Aut}(S)|$.
2. $H_{\mathcal{G}} = \binom{n}{2} h(p)$
3. $\sum_{S \in \mathcal{S}} P(S) \log |\mathrm{Aut}(S)| = o(1)$ by asymmetry of $\mathcal{G}(n, p)$.

Compression Algorithm

Compression Algorithm called Structural zip, in short SZIP – Demo.

Compression Algorithm

Compression Algorithm called Structural zip, in short SZIP – Demo.

We can prove the following estimate on the compression ratio of $S(p,n)$ for our algorithm SZIP.

Theorem 3 (Y. Choi and W.S., 2008). *The total expected length of encoding produced by our algorithm* SZIP *for a structure S from $\mathcal{S}(n,p)$, is at most*

$$\binom{n}{2} h(p) - n \log n + n\,(c + \Phi(\log n)) + O(n^{1-\eta}),$$

where $h(p) = -p \log p - (1-p) \log(1-p)$, c is an explicitly computable constant, η is a positive constant, and $\Phi(x)$ is a fluctuating function with a small amplitude or zero.

Our algorithm is asymptotically optimal up to the second largest term, and works quite fine in practise.

Experimental Results

Real-world and random graphs.

Table 1: The length of encodings (in bits)

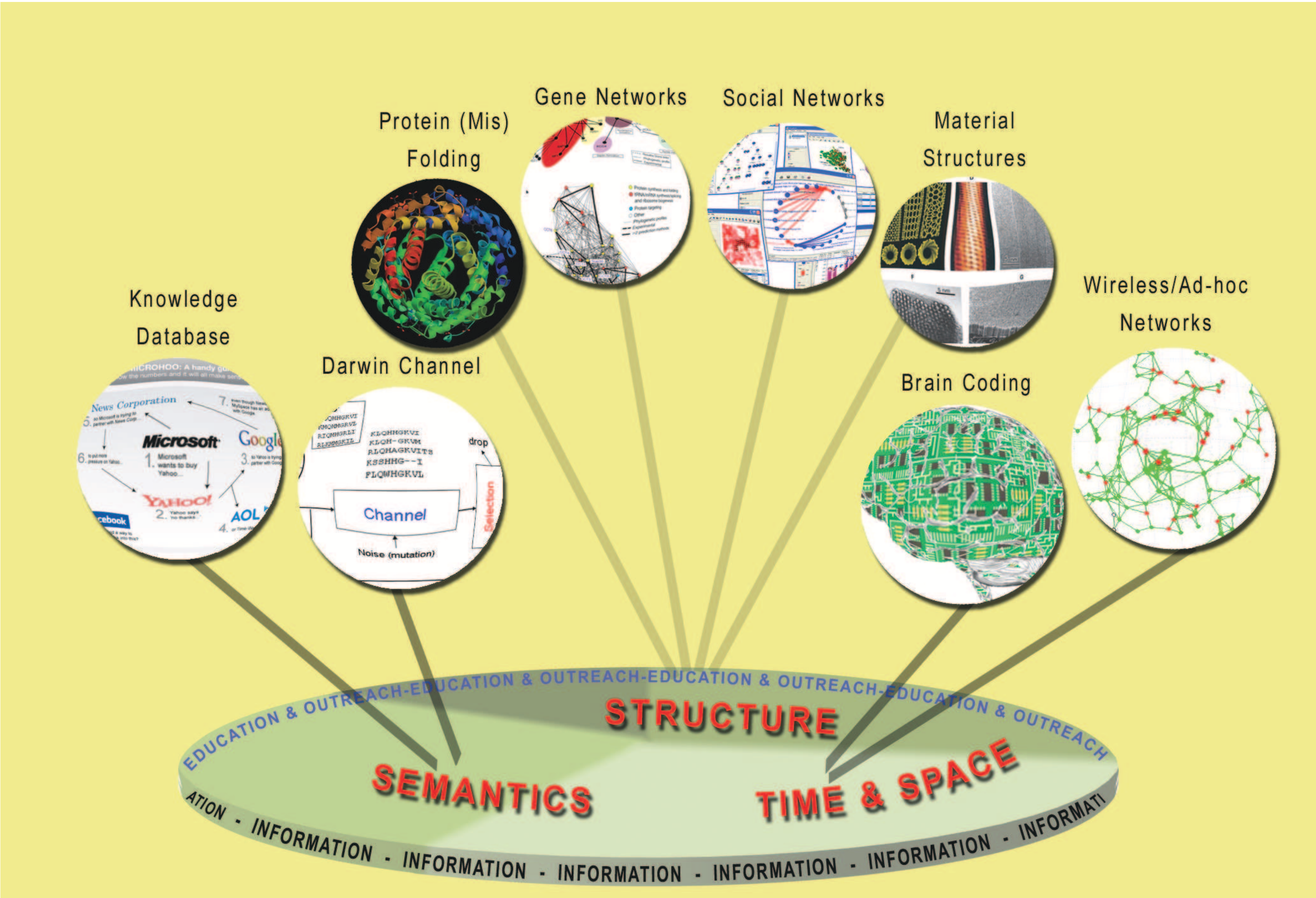
	Networks	# of nodes	# of edges	our algorithm	adjacency matrix	adjacency list	arithmetic coding
Real-world	US Airports	332	2,126	8,118	54,946	38,268	12,991
	Protein interaction (Yeast)	2,361	6,646	46,912	2,785,980	1 59,504	67,488
	Collaboration (Geometry)	6,167	21,535	115,365	19,012,861	55 9,910	241,811
	Collaboration (Erdös)	6,935	11,857	62,617	24,043,645	308,2 82	147,377
	Genetic interaction (Human)	8,605	26,066	221,199	37,018,710	729,848	310,569
	Internet (AS level)	25,881	52,407	301,148	334,900,140	1,572, 210	396,060
Random	$\mathcal{S}(n,p)$	1,000	$p = 0.01$	34,361	499,500	99,900	40,350
	$\mathcal{S}(n,p)$	1,000	$p = 0.1$	227,236	499,500	999,999	234,392
	$\mathcal{S}(n,p)$	1,000	$p = 0.3$	432,692	499,500	2,997,99 9	440,252
					$\binom{n}{2}$	$2e\lceil\log n\rceil$	$\binom{n}{2}h(p)$

- n : number of vertices
- e : number of edges
- Adjacency matrix : $\binom{n}{2}$ bits
- Adjacency list : $2e\lceil\log n\rceil$ bits
- Arithmetic coding : $\sim \binom{n}{2}h(p)$ bits (compressing the adjacency matrix)

Outline Update

1. Shannon Information
2. What is Information?
3. Beyond Shannon
4. Learnable Information – Fundamental Limits
5. Spatio-Temporal Information
6. Structural Information
   (a) Graphical Compression and Fundamental Limit
   (b) Remote Spanner – Topology Compression in Ad-hoc Networks
7. Science of Information

Science of Information



Institute for Science of Information

In 2008 at Purdue we launched the

Institute for Science of Information

integrating research and teaching activities aimed at investigating the role of **information** from various viewpoints: from the fundamental theoretical underpinnings of information to the science and engineering of novel information substrates, biological pathways, communication networks, economics, and complex social systems.

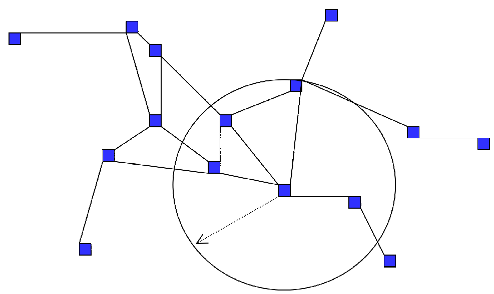
The specific means and goals for the Center are:

- Prestige Science Lecture Series on Information to collectively ponder short and long term goals;
- organize meetings and workshops (e.g., Information Beyond Shannon, Orlando 2005, and Venice 2008).
- encourage and facilitate interdisciplinary collaborations (NSF STC with Berkeley, MIT, Princeton, and Stanford).
- provide scholarships and fellowships for the best students, and support the development of new interdisciplinary courses.
- initiate similar centers around the world to support research on information.

That’s It

THANK YOU

Mobile Ad-hoc Networks



Low bandwidth ($\leq$ 10 Mbps).
High mobility: topology may change every second.
Traffic control floods the network (with classic protocols).

Example: Stadium Network. Undirected network graph $G(V, E)$ with $|V| = 10,000$ mobile nodes, each with 1,000 neighbors: $|E| = 10^7$ links;
Classic protocols would need $|E|^2 = 10^{14}$ link advertisements per second.

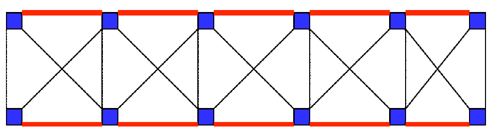
Shortest Path Routing:
Limits the traffic overhead,
Optimizes the number of packet retransmissions.

Information (and structural compression) issues arise here again, however, this time the objective function is different, namely to *minimize* the *number of retransmissions*.

Topological Compression

1. Consider an undirected graph $G(V, E)$. The set $S \subset E$ is a *spanner* if S connects all nodes and every node u computes its shortest path to all nodes based on S.

2. A spanner is *unstretched* if the shortest path in $G(V, S)$ is also in $G(V, E)$.



3. $S \subset E$ is a *remote spanner* if $\forall u \in V : S \cup N(u)$ connects to all nodes, where $N(u)$ is the neighborhood of u.

4. S is an *unstretched remote spanner* if $\forall(u, v) \in V^2$ there exists a shortest path from u to all v in $S \cup N(u)$.

5. Topological Compression: $\frac{|S|}{|E|}$ is called the topology compression ratio.

6. **Theorem**: The set $S \subset E$ is an unstretched remote spanner iff $\forall u \in V$ $N(u) \cap S$ s the dominating set of the two hop neighborhood (i.e., $N^2(u)$).

Compression Ratio for Two Graph Models

$T = N(u) \cap S$ is called the MultiPoint Relay Star of node u.
Then $S = \bigcup_{u \in V} T$. Observe that:

Bad news: Computing optimal T is NP-complete.
Good News: Greedy algorithm within factor $1 + \log |V|$ of the optimal T (Chvátal, 1979).

Erdös-Rényi Graph Model: Edges are added with probability p.

Theorem 4 (Jacquet and Viennot, 2009). *In Erdös-Rényi random graph model the average topology compression ratio is asymptotically equal to* $\frac{\log |V|}{p^2 |V|}$ *when* $|V| \to \infty$.

Geometric Graph Model: Nodes are uniformly distributed with density ν.

Theorem 5 (Jacquet and Viennot, 2009). *In geometric random graph model, the average topology compression is asymptotically equal to*

$$\frac{3}{\nu^{\frac{2}{3}}}$$

as $\nu \to \infty$.