# Algorithms, Combinatorics, Information, and Beyond*

Wojciech Szpankowski
Purdue University
W. Lafayette, IN 47907

August 2, 2011

## NSF STC Center for Science of Information

Plenary ISIT, St. Petersburg, 2011

## Dedicated to PHILIPPE FLAJOLET

*Research supported by NSF Science & Technology Center, and Humboldt Foundation.

# Outline

1. Shannon Legacy
2. Analytic Combinatorics + IT = Analytic Information Theory
3. The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources
   (c) Universal Markov Sources
4. Post-Shannon Challenges: Beyond Shannon
5. Science of Information: NSF Science and Technology Center

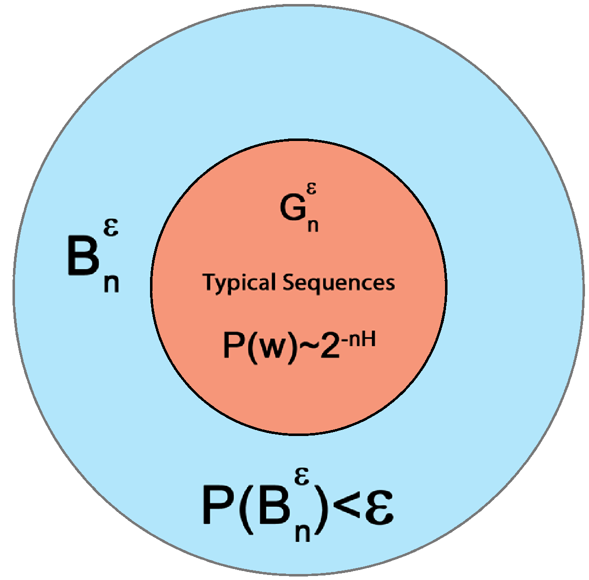
| | |
|---|---|
| **Algorithms**: | are at the heart of virtually all computing technologies; |
| **Combinatorics**: | provides indispensable tools for finding patterns and structures; |
| **Information**: | permeates every corner of our lives and shapes our universe. |

# Shannon Legacy: Three Theorems of Shannon

**Theorem 1 & 3**. (**Shannon 1948; Lossless & Lossy Data Compression**)

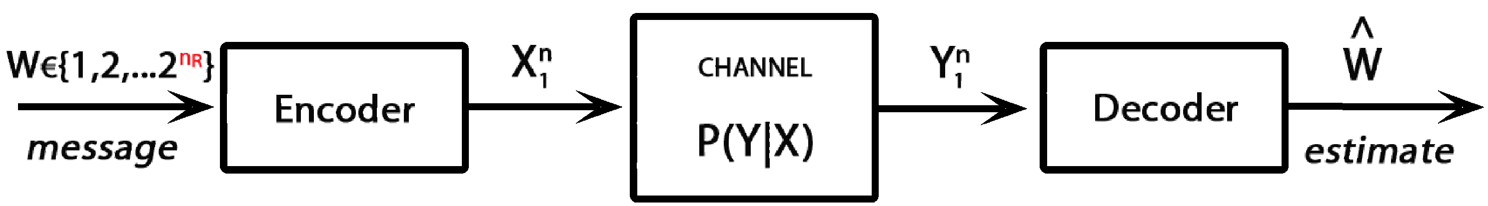
compression bit rate $\geq$ source entropy $H(X)$

for distortion level $D$:
lossy bit rate $\geq$ rate distortion function $R(D)$




**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.



# Post-Shannon Challenges

1. **Back off from infinity** (Ziv'97): Extend Shannon findings to finite size data structures (i.e., sequences, graphs), that is, develop information theory of various data structures beyond first-order asymptotics.

> **Claim**: Many interesting information-theoretic phenomena appear in the second-order terms.

2. **Science of Information**: Information Theory needs to meet new challenges of current applications in

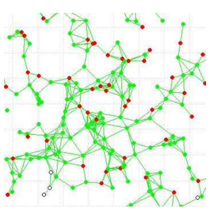
biology, communication, knowledge extraction, economics, . . .

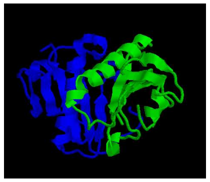
to understand new aspects of information in:

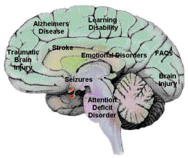
structure, time, space, and semantics,

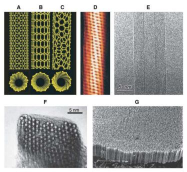
and

dynamic information, limited resources, complexity, representation-invariant information, and cooperation & dependency.







# Outline Update

1. Shannon Legacy
2. Analytic Information Theory
3. Source Coding: The Redundancy Rate Problem
4. Post-Shannon Information
5. NSF Science and Technology Center

# Analytic Combinatorics+Information Theory=Analytic Information Theory

- In the **1997 Shannon Lecture** Jacob Ziv presented compelling arguments for "backing off" from first-order asymptotics in order to predict the behavior of real systems with finite length description.

- To **overcome** these difficulties, one may replace first-order analyses by non-asymptotic analysis, however, we propose to develop full asymptotic expansions and more precise analysis (e.g., large deviations, CLT).

- Following **Hadamard's precept**[1], we study information theory problems using techniques of complex analysis such as generating functions, combinatorial calculus, Rice's formula, Mellin transform, Fourier series, sequences distributed modulo 1, saddle point methods, analytic poissonization and depoissonization, and singularity analysis.

- This program, which applies complex-analytic tools to information theory, constitutes **analytic information theory**.[2]

---

[1] The shortest path between two truths on the real line passes through the complex plane.

[2] Andrew Odlyzko: "Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision."

# Some Successes of Analytic Information Theory

- Wyner-Ziv Conjecture concerning the longest match in the WZ'89 compression scheme (W.S., 1993).
- Ziv's Conjecture on the distribution of the number of phrases in the LZ'78 (Jacquet & W.S., 1995, 2011).
- Redundancy of the LZ'78 (Savari, 1997, Louchard & W.S., 1997).
- Steinberg-Gutman Conjecture regarding lossy pattern matching compression (Luczak & W.S., 1997; Kieffer & Yang, 1998; Kontoyiannis, 2003).
- Precise redundancy of Huffman's Code (W.S., 2000) and redundancy of a fixed-to-variable no prefix free code (W.S. & Verdu, 2010).
- Minimax Redundancy for memoryless sources (Xie &Barron, 1997; W.S., 1998; W.S. & Weinberger, 2010), Markov sources (Risannen, 1998; Jacquet & W.S., 2004), and renewal sources (Flajolet & W.S., 2002; Drmota & W.S., 2004).
- Analysis of variable-to-fixed codes such as Tunstall and Khodak codes (Drmota, Reznik, Savari, & W.S., 2006, 2008, 2010).
- Entropy of hidden Markov processes and the noisy constrained capacity (Jacquet, Seroussi, & W.S., 2004, 2007, 2010; Han & Marcus, 2007).
- ...

# Outline Update

1. Shannon Legacy
2. Analytic Information Theory
3. Source Coding: The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Markov Sources
   (c) Universal Renewal Sources

# Source Coding and Redundancy

Source coding aims at finding codes $C : \mathcal{A}^* \to \{0,1\}^*$ of the shortest length $L(C,x)$, either on *average* or for *individual sequences*.

**Known Source** $P$: The pointwise and maximal redundancy are:

$$
\begin{aligned}
R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \log P(x_1^n) \\
R^*(C_n, P) &= \max_{x_1^n} [L(C_n, x_1^n) + \log P(x_1^n)]
\end{aligned}
$$

where $P(x_1^n)$ is the probability of $x_1^n = x_1 \cdots x_n$.

**Unknown Source** $P$: Following Davisson, the maximal minimax redundancy $R_n^*(\mathcal{S})$ for a family of sources $\mathcal{S}$ is:

$$
R_n^*(\mathcal{S}) = \min_{C_n} \sup_{P \in \mathcal{S}} \max_{x_1^n} [L(C_n, x_1^n) + \log P(x_1^n)].
$$

# Source Coding and Redundancy

Source coding aims at finding codes $C : \mathcal{A}^* \to \{0,1\}^*$ of the shortest length $L(C,x)$, either on *average* or for *individual sequences*.

**Known Source** $P$: The pointwise and maximal redundancy are:

$$\begin{aligned} R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \log P(x_1^n) \\ R^*(C_n, P) &= \max_{x_1^n}[L(C_n, x_1^n) + \log P(x_1^n)] \end{aligned}$$

where $P(x_1^n)$ is the probability of $x_1^n = x_1 \cdots x_n$.

**Unknown Source** $P$: Following Davisson, the maximal minimax redundancy $R_n^*(\mathcal{S})$ for a family of sources $\mathcal{S}$ is:

$$R_n^*(\mathcal{S}) = \min_{C_n} \sup_{P\in\mathcal{S}} \max_{x_1^n}[L(C_n, x_1^n) + \log P(x_1^n)].$$

**Shtarkov's Bound**:

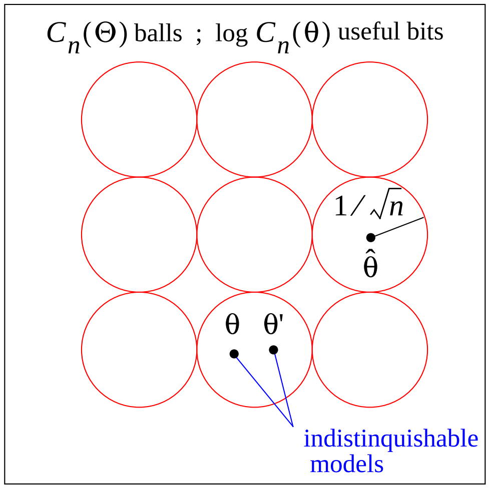
$$d_n(\mathcal{S}) := \log \sum_{x_1^n \in \mathcal{A}^n} \sup_{P\in\mathcal{S}} P(x_1^n) \le R_n^*(\mathcal{S}) \le \underbrace{\log \sum_{x_1^n \in \mathcal{A}^n} \sup_{P\in\mathcal{S}} P(x_1^n)}_{D_n(\mathcal{S})} + 1$$

# Learnable Information and Redundancy

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

# Learnable Information and Redundancy

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

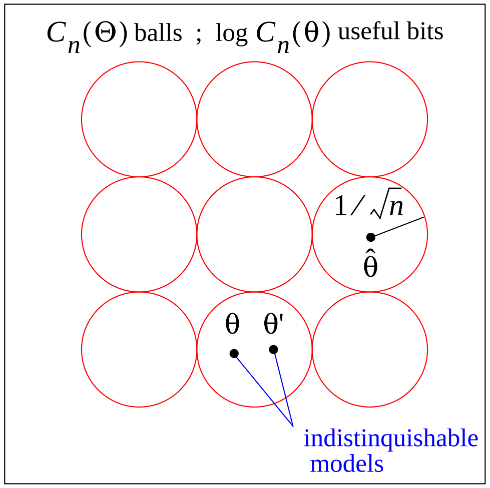



**2.** Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$ are indistinguishable if the ML estimator $\hat{\theta}$ with high probability declares both models are the same.

**3.** The number of distinguishable distributions (i.e, $\hat{\theta}$), $C_n(\Theta)$, summarizes then learnable information, $I(\Theta) = \log_2 C_n(\Theta)$.

# Learnable Information and Redundancy

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




**2.** Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$ are indistinguishable if the ML estimator $\hat{\theta}$ with high probability declares both models are the same.

**3.** The number of distinguishable distributions (i.e, $\hat{\theta}$), $C_n(\Theta)$, summarizes then learnable information, $I(\Theta) = \log_2 C_n(\Theta)$.

**4.** 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 the number of distinguishable balls $C_n(\Theta)$ of radius $O(1/\sqrt{n})$ is asymptotically equal to the minimax redundancy:

$$\textbf{Learnable Information} = \log C_n(\Theta) = \inf_{\theta\in\Theta}\max_{x^n} \log\frac{P_{\hat{\theta}}}{P_\theta} = R_n^*(\mathcal{M}^k)$$

# Outline Update

1. Shannon Legacy

2. Analytic Information Theory

3. Source Coding: The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources
   (c) Universal Markov Sources

# Maximal Minimax for Memoryless Sources

For a **memoryless source** over the alphabet $\mathcal{A} = \{1, 2, \ldots, m\}$ we have

$$P(x_1^n) = {p_1}^{k_1} \cdots {p_m}^{k_m}, \qquad k_1 + \cdots + k_m = n.$$

Then

$$\begin{aligned}
D_n(\mathcal{M}_0) &:= \sum_{x_1^n} \sup_{P(x_1^n)} P(x_1^n) \\
&= \sum_{x_1^n} \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} \\
&= \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1, \ldots, k_m} \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} \\
&= \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1, \ldots, k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}.
\end{aligned}$$

since the (unnormalized) **likelihood distribution** is

$$\sup_{P(x_1^n)} P(x_1^n) = \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} = \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}$$

# Generating Function for $D_n(\mathcal{M}_0)$

We write

$$D_n(\mathcal{M}_0) = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1,\ldots,k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m} = \frac{n!}{n^n} \sum_{k_1+\cdots+k_m=n} \frac{k_1^{k_1}}{k_1!} \cdots \frac{k_m^{k_m}}{k_m!}$$

Let us introduce a tree-generating function

$$B(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k = \frac{1}{1-T(z)}, \qquad T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k$$

where $T(z) = ze^{T(z)}$ $(= -W(-z)$, **Lambert's** $W$-function) that enumerates all rooted labeled trees. Let now

$$D_m(z) = \sum_{n=0}^{\infty} z^n \frac{n^n}{n!} D_n(\mathcal{M}_0).$$

Then by the convolution formula

$$\boxed{D_m(z) = [B(z)]^m - 1.}$$

# Asymptotics for FINITE $m$

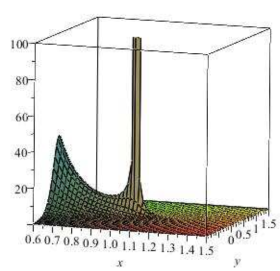
The function $B(z)$ has an algebraic singularity at $z = e^{-1}$, and

$$\beta(z) = B(z/e) = \frac{1}{\sqrt{2(1-z)}} + \frac{1}{3} + O(\sqrt{(1-z)}.$$

By Cauchy's coefficient formula

$$D_n(\mathcal{M}_0) = \frac{n!}{n^n}[z^n][B(z)]^m = \sqrt{2\pi n}(1+O(1/n))\frac{1}{2\pi i}\oint \frac{\beta(z)^m}{z^{n+1}}dz.$$

For finite $m$, the singularity analysis of Flajolet and Odlyzko implies



$$[z^n](1-z)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}, \quad \alpha \notin \{0,-1,-2,\ldots\}$$

that finally yields (cf. Clarke & Barron, 1990, W.S., 1998)

$$R_n^*(\mathcal{M}_0) = \frac{m-1}{2}\log\left(\frac{n}{2}\right) + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + \frac{\Gamma(\frac{m}{2})m}{3\Gamma(\frac{m}{2}-\frac{1}{2})}\cdot\frac{\sqrt{2}}{\sqrt{n}}$$

$$+ \left(\frac{3+m(m-2)(2m+1)}{36} - \frac{\Gamma^2(\frac{m}{2})m^2}{9\Gamma^2(\frac{m}{2}-\frac{1}{2})}\right)\cdot\frac{1}{n} + \cdots$$

# Redundancy for LARGE $m$

Now assume that $m$ is unbounded and may vary with $n$. Then

$$D_{n,m}(\mathcal{M}_0) = \sqrt{2\pi n}\frac{1}{2\pi i}\oint \frac{\beta(z)^m}{z^{n+1}}dz = \sqrt{2\pi n}\frac{1}{2\pi i}\oint e^{g(z)}dz$$

where $g(z) = m\ln\beta(z) - (n+1)\ln z$.

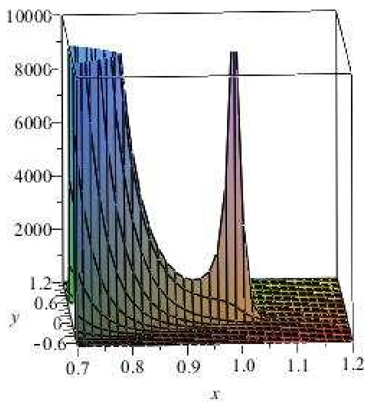
The saddle point $z_0$ is a solution of $g'(z_0) = 0$, that is,

$$g(z) = g(z_0) + \frac{1}{2}(z - z_0)^2 g''(z_0) + O(g'''(z_0)(z - z_0)^3).$$

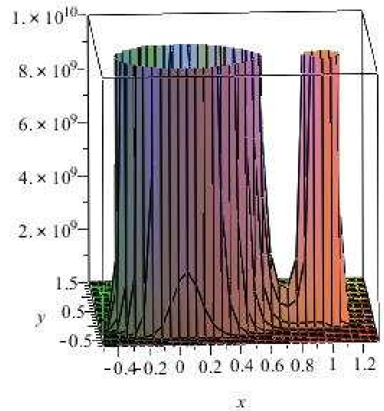
Under mild conditions satisfied by our $g(z)$ (e.g., $z_0$ is real and unique), the saddle point method leads to:

$$D_{n,m}(\mathcal{M}_0) = \frac{e^{g(z_0)}}{\sqrt{2\pi|g''(z_0)|}} \times \left(1 + O\left(\frac{g'''(z_0)}{(g''(z_0))^\rho}\right)\right),$$

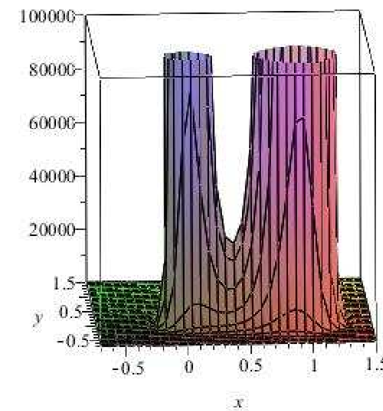
for some $\rho < 3/2$.




$m = o(n)$




$m = n$




$n = o(m)$

**Theorem 1** (Orlitsky and Santhanam, 2004, and W.S. and Weinberger, 2010).
(i) *For* $m = o(n)$

$$R^*_{n,m}(\mathcal{M}_0) = \frac{m-1}{2}\log\frac{n}{m} + \frac{m}{2}\log e + \frac{m\log e}{3}\sqrt{\frac{m}{n}\frac{1}{2}} - O\left(\sqrt{\frac{m}{n}}\right)$$

(ii) *For* $m = \alpha n + \ell(n)$, *where* $\alpha$ *is a positive constant and* $\ell(n) = o(n)$,

$$R^*_{n,m}(\mathcal{M}_0) \quad = \quad n\log B_\alpha + \ell(n)\log C_\alpha - \log\sqrt{A_\alpha} + O(\ell(n)^2/n)$$

*where* $C_\alpha := 0.5 + 0.5\sqrt{1+4/\alpha}$, $A_\alpha := C_\alpha + 2/\alpha$, $B_\alpha = \alpha C_\alpha^{\alpha+2} e^{-\frac{1}{C_\alpha}}$.
(iii) *For* $n = o(m)$

$$R^*_{n,m}(\mathcal{M}_0) = n\log\frac{m}{n} + \frac{3}{2}\frac{n^2}{m}\log e - \frac{3}{2}\frac{n}{m}\log e + O\left(\frac{1}{\sqrt{n}} + \frac{n^3}{m^2}\right).$$

# Outline Update

1. Shannon Legacy

2. Analytic Information Theory

3. Source Coding: The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources
   (c) Universal Markov Sources

# Renewal Sources (Virtual Large Alphabet)

The **renewal process** $\mathcal{R}_0$ (introduced in 1996 by Csiszár and Shields) defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$.
- In a **binary renewal sequence** the positions of the $1$'s are at the renewal epochs $T_0, T_0 + T_1, \ldots$ with runs of zeros of lengths $T_1 - 1, T_2 - 1, \ldots$.

# Renewal Sources (Virtual Large Alphabet)

The **renewal process** $\mathcal{R}_0$ (introduced in 1996 by Csiszár and Shields) defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$.
- In a **binary renewal sequence** the positions of the $1$'s are at the renewal epochs $T_0, T_0 + T_1, \ldots$ with runs of zeros of lengths $T_1 - 1, T_2 - 1, \ldots$.

For a sequence

$$x_0^n = 10^{\alpha_1} 10^{\alpha_2} 1 \cdots 10^{\alpha_n} 1 \underbrace{0 \cdots 0}_{k^*}$$

define $k_m$ as the number of $i$ such that $\alpha_i = m$. Then

$$P(x_1^n) = [Q(0)]^{k_0} [Q(1)]^{k_1} \cdots [Q(n-1)]^{k_{n-1}} \Pr\{T_1 > k^*\}.$$

# Renewal Sources (Virtual Large Alphabet)

The **renewal process** $\mathcal{R}_0$ (introduced in 1996 by Csiszár and Shields) defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$.
- In a **binary renewal sequence** the positions of the $1$'s are at the renewal epochs $T_0, T_0 + T_1, \ldots$ with runs of zeros of lengths $T_1 - 1, T_2 - 1, \ldots$.

For a sequence

$$x_0^n = 10^{\alpha_1}10^{\alpha_2}1 \cdots 10^{\alpha_n}1\underbrace{0 \cdots 0}_{k^*}$$

define $k_m$ as the number of $i$ such that $\alpha_i = m$. Then

$$P(x_1^n) = [Q(0)]^{k_0}[Q(1)]^{k_1} \cdots [Q(n-1)]^{k_{n-1}}\Pr\{T_1 > k^*\}.$$

**Theorem 2** (Flajolet and W.S., 1998). *Consider the class of renewal processes. Then*

$$R_n^*(\mathcal{R}_0) = \frac{2}{\log 2}\sqrt{cn} + O(\log n).$$

*where* $c = \frac{\pi^2}{6} - 1 \approx 0.645$.

# Maximal Minimax Redundancy

It can be proved that $r_{n+1} - 1 \le D_n(\mathcal{R}_0) \le \sum_{m=0}^{n} r_m$

$$r_n = \sum_{k=0}^{n} r_{n,k}, \quad r_{n,k} = \sum_{\mathcal{I}(n,k)} \binom{k}{k_0 \ldots k_{n-1}} \left(\frac{k_0}{k}\right)^{k_0} \left(\frac{k_1}{k}\right)^{k_1} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}$$

where $\mathcal{I}(n, k)$ is is the integer partition of $n$ into $k$ terms, i.e.,

$$n = k_0 + 2k_1 + \cdots + nk_{n-1}, \quad k = k_0 + \cdots + k_{n-1}.$$

But we shall study $s_n = \sum_{k=0}^{n} s_{n,k}$ where

$$s_{n,k} = e^{-k} \sum_{\mathcal{I}(n,k)} \frac{k^{k_0}}{k_0!} \cdots \frac{k^{k_{n-1}}}{k_{n-1}!}$$

since $S(z, u) = \sum_{k,n} s_{n,k} u^k z^n = \prod_{i=1}^{\infty} \beta(z^i u)$.

# Maximal Minimax Redundancy

It can be proved that $r_{n+1} - 1 \le D_n(\mathcal{R}_0) \le \sum_{m=0}^{n} r_m$

$$r_n = \sum_{k=0}^{n} r_{n,k}, \quad r_{n,k} = \sum_{\mathcal{I}(n,k)} \binom{k}{k_0 \ldots k_{n-1}} \left(\frac{k_0}{k}\right)^{k_0} \left(\frac{k_1}{k}\right)^{k_1} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}$$

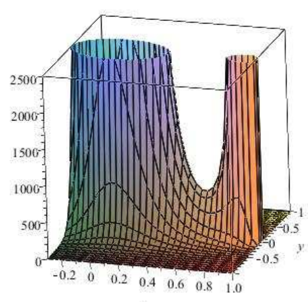
where $\mathcal{I}(n,k)$ is is the integer partition of $n$ into $k$ terms, i.e.,

$$n = k_0 + 2k_1 + \cdots + nk_{n-1}, \quad k = k_0 + \cdots + k_{n-1}.$$

But we shall study $s_n = \sum_{k=0}^{n} s_{n,k}$ where

$$s_{n,k} = e^{-k} \sum_{\mathcal{I}(n,k)} \frac{k^{k_0}}{k_0!} \cdots \frac{k^{k_{n-1}}}{k_{n-1}!}$$

since $S(z,u) = \sum_{k,n} s_{n,k} u^k z^n = \prod_{i=1}^{\infty} \beta(z^i u)$.




$$s_n = [z^n]S(z,1) = [z^n] \exp\left(\frac{c}{1-z} + a \log \frac{1}{1-z}\right)$$

**Theorem 3** (Flajolet and W.S., 1998). *We have the following asymptotics*

$$s_n \sim \exp\left(2\sqrt{cn} - \frac{7}{8}\log n + O(1)\right), \quad \log r_n = \frac{2}{\log 2}\sqrt{cn} - \frac{5}{8}\log n + \frac{1}{2}\log\log n + O(1).$$

# Outline Update

1. Shannon Legacy

2. Analytic Information Theory

3. Source Coding: The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources
   (c) Universal Markov Sources

# Maximal Minimax for Markov Sources

Markov source over $|\mathcal{A}| = m$ (FINITE) with transition matrix $P = \{p_{ij}\}_{i,j=1}^{m}$ :

$$D_n(\mathcal{M}_1) = \sum_{x_1^n} \sup_P \; p_{11}^{k_{11}} \cdots p_{mm}^{k_{mm}} = \sum_{\mathbf{k} \in \mathcal{P}_n} |\mathcal{T}_n(\mathbf{k})| \left(\frac{k_{11}}{k_1}\right)^{k_{11}} \cdots \left(\frac{k_{mm}}{k_m}\right)^{k_{mm}},$$

where $\mathcal{P}_n(m)$ denotes Markov types, while $\mathcal{T}_n(\mathbf{k}) := \mathcal{T}_n(x_1^n)$ denotes the number of sequences of type $\mathbf{k} = [k_{ij}]$, with $k_{ij}$ representing the numbers of pairs $(ij)$ in $x_1^n$.

# Maximal Minimax for Markov Sources

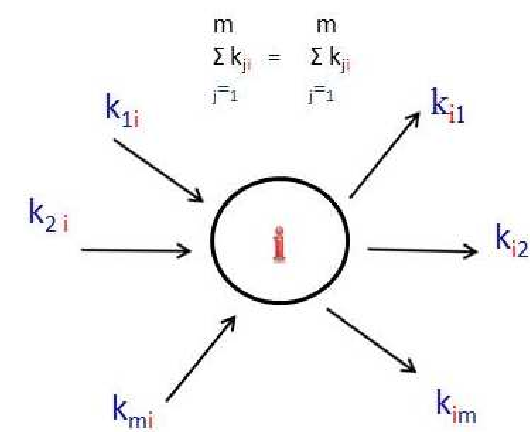
Markov source over $|\mathcal{A}| = m$ (FINITE) with transition matrix $P = \{p_{ij}\}_{i,j=1}^m$ :

$$D_n(\mathcal{M}_1) = \sum_{x_1^n} \sup_P \; p_{11}^{k_{11}} \cdots p_{mm}^{k_{mm}} = \sum_{\mathbf{k} \in \mathcal{P}_n} |\mathcal{T}_n(\mathbf{k})| \left(\frac{k_{11}}{k_1}\right)^{k_{11}} \cdots \left(\frac{k_{mm}}{k_m}\right)^{k_{mm}},$$

where $\mathcal{P}_n(m)$ denotes Markov types, while $\mathcal{T}_n(\mathbf{k}) := \mathcal{T}_n(x_1^n)$ denotes the number of sequences of type $\mathbf{k} = [k_{ij}]$, with $k_{ij}$ representing the numbers of pairs $(ij)$ in $x_1^n$.

For circular strings (i.e., after the $n$th symbol we re-visit the first symbol of $x_1^n$), the matrix $\mathbf{k} = [k_{ij}]$ satisfies the following constraints denoted as $\mathcal{F}_n(m)$:

$$\sum_{1 \le i,j \le m} k_{ij} = n, \qquad \sum_{j=1}^m k_{ij} = \sum_{j=1}^m k_{ji}$$




For example: m=3

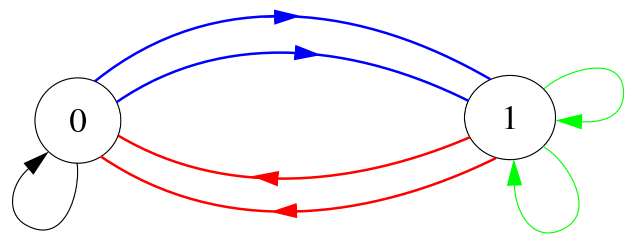
$k_{11} + k_{12} + k_{13} + k_{21} + k_{22} + k_{23} + k_{31} + k_{32} + k_{33} = n$
$k_{12} + k_{13} = k_{21} + k_{31}$
$k_{12} + k_{32} = k_{21} + k_{23}$
$k_{13} + k_{23} = k_{31} + k_{32}$

Matrices satisfying $\mathcal{F}_n(m)$ are called balanced matrices, and $|\mathcal{F}_n(m)|$ denotes the number of balanced matrices over $\mathcal{A}$.

# Markov Types and Eulerian Cycles

Example: Let $\mathcal{A} = \{0, 1\}$ and

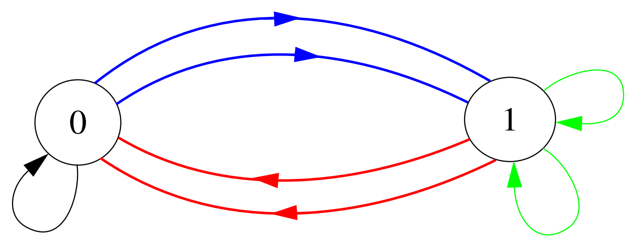
$$\mathbf{k} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$



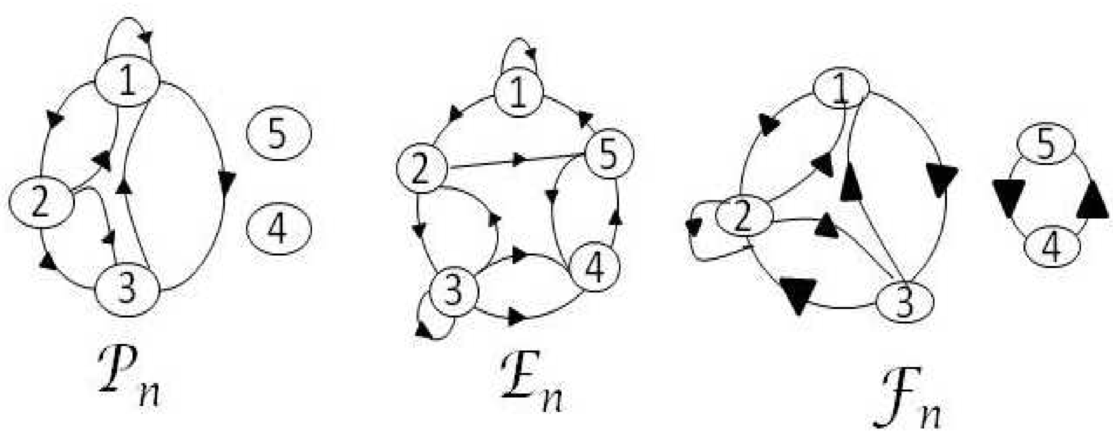
# Markov Types and Eulerian Cycles

Example: Let $\mathcal{A} = \{0, 1\}$ and

$$\mathbf{k} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$




$\mathcal{P}_n(m)$ – Markov types but also . . .
a set of all connected Eulerian di-graphs $G = (V(G)], E(G))$ such that $V(G) \subseteq \mathcal{A}$ and $|E(G)| = n$.
$\mathcal{E}_n(m)$ – set of connected Eulerian digraphs on $\mathcal{A}$.
$\mathcal{F}_n(m)$ – balanced matrices but also . . .
set of (not necessary connected) Eulerian digraphs on $\mathcal{A}$.




Asymptotic equivalence: $|\mathcal{P}_n(m)| = |\mathcal{F}_n(m)| + O(n^{m^2-3m+3}) \sim |\mathcal{E}_n(m)|$.

# Markov Types – Main Results

**Theorem 4** (Knessl, Jacquet, and W.S., 2010). (i) *For fixed* $m$ *and* $n \to \infty$

$$|\mathcal{P}_n(m)| = d(m)\frac{n^{m^2-m}}{(m^2-m)!} + O(n^{m^2-m-1})$$

$$d(m) = \frac{1}{(2\pi)^{m-1}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(m-1)-fold} \prod_{j=1}^{m-1} \frac{1}{1+\varphi_j^2} \cdot \prod_{k \neq \ell} \frac{1}{1+(\varphi_k - \varphi_\ell)^2} \, d\varphi_1 d\varphi_2 \cdots d\varphi_{m-1}.$$

(ii) *When* $m \to \infty$ *we find that provided that* $m^4 = o(n)$

$$|\mathcal{P}_n(m)| \sim \frac{\sqrt{2}m^{3m/2}e^{m^2}}{m^{2m^2}2^m\pi^{m/2}} \cdot n^{m^2-m}$$

Some numerics: $|\mathcal{P}_n(3)| \sim \frac{1}{12}\frac{n^6}{6!}$, $|\mathcal{P}_n(4)| \sim \frac{1}{96}\frac{n^{12}}{12!}$, $|\mathcal{P}_n(5)| \sim \frac{37}{34560}\frac{n^{20}}{20!}$.

# Markov Types – Main Results

**Theorem 4** (Knessl, Jacquet, and W.S., 2010). (i) *For fixed* $m$ *and* $n \to \infty$

$$|\mathcal{P}_n(m)| = d(m)\frac{n^{m^2-m}}{(m^2-m)!} + O(n^{m^2-m-1})$$

$$d(m) = \frac{1}{(2\pi)^{m-1}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(m-1)-fold} \prod_{j=1}^{m-1} \frac{1}{1+\varphi_j^2} \cdot \prod_{k \neq \ell} \frac{1}{1+(\varphi_k - \varphi_\ell)^2} \, d\varphi_1 d\varphi_2 \cdots d\varphi_{m-1}.$$

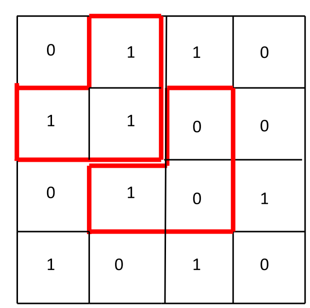
(ii) *When* $m \to \infty$ *we find that provided that* $m^4 = o(n)$

$$|\mathcal{P}_n(m)| \sim \frac{\sqrt{2} m^{3m/2} e^{m^2}}{m^{2m^2} 2^m \pi^{m/2}} \cdot n^{m^2-m}$$

Some numerics: $|\mathcal{P}_n(3)| \sim \frac{1}{12}\frac{n^6}{6!}$, $|\mathcal{P}_n(4)| \sim \frac{1}{96}\frac{n^{12}}{12!}$, $|\mathcal{P}_n(5)| \sim \frac{37}{34560}\frac{n^{20}}{20!}$.

**Open Question**:

Counting types and sequences of a given type in a **Markov field**. (Answer: $|\mathcal{P}_n(2)| \sim \frac{1}{12}\frac{n^5}{5!}$ ?)



# Markov Redundancy: Main Results

**Theorem 5** (Rissanen, 1996, Jacquet & W.S., 2004). *Let* $\mathcal{M}_1$ *be a Markov source over an* $m$*-ry alphabet. Then*

$$R_n^*(\mathcal{M}_1) = \frac{m(m-1)}{2} \log\left(\frac{n}{2\pi}\right) + \log A_m + \log(1/n)$$

*with*

$$A_m = \int_{\mathcal{K}(1)} m F_m(y_{ij}) \prod_i \frac{\sqrt{\sum_j y_{ij}}}{\prod_j \sqrt{y_{ij}}} d[y_{ij}]$$

*where* $\mathcal{K}(1) = \{y_{ij} : \sum_{ij} y_{ij} = 1\}$ *and* $F_m(\cdot)$ *is a polynomial.*

In particular, for $m = 2$ $A_2 = 16 \times$ Catalan where Catalan is Catalan's constant $\sum_i \frac{(-1)^i}{(2i+1)^2} \approx 0.915965594$.

**Theorem 6.** *Let* $\mathcal{M}_r$ *be a Markov source of order* $r$. *Then*

$$R_n^*(\mathcal{M}_r) = \frac{m^r(m-1)}{2} \log\left(\frac{n}{2\pi}\right) + \log A_m^r + O(1/n)$$

*where* $A_m^r$ *is a constant defined in a similar fashion as* $A_m$ *above.*

# Outline Update

1. Shannon Information Theory
2. Source Coding
3. The Redundancy Rate Problem
4. Post-Shannon Information (Science of Information)
5. NSF Science and Technology Center

# Post-Shannon Challenges: Time & Space

We need to extend information theory to include structure, **time & space**, and . . ..

**1.** Coding Rate for Finite Blocklength: Polyanskiy, Poor & Verdu, 2010

$$\frac{1}{n}\log M^*(n, , \varepsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\varepsilon)$$

where $C$ is the capacity, $V$ is the channel dispersion, $n$ is the block length, $\varepsilon$ error probability, and $Q$ is the complementary Gaussian distribution.

# Post-Shannon Challenges: Time & Space

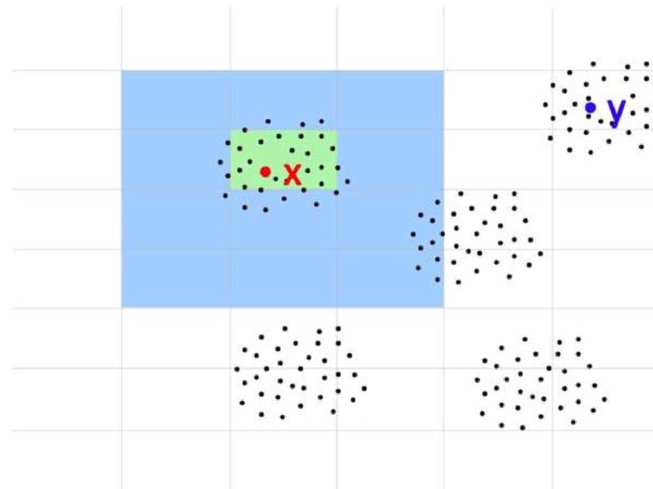
We need to extend information theory to include structure, **time & space**, and . . ..

1. Coding Rate for Finite Blocklength: Polyanskiy, Poor & Verdu, 2010

$$\frac{1}{n}\log M^*(n,,\varepsilon) \approx C - \sqrt{\frac{V}{n}}Q^{-1}(\varepsilon)$$

where $C$ is the capacity, $V$ is the channel dispersion, $n$ is the block length, $\varepsilon$ error probability, and $Q$ is the complementary Gaussian distribution.

**Time & Space**:
Classical Information Theory is at its weakest in dealing with problems of delay (e.g., information arriving late maybe useless or has less value).



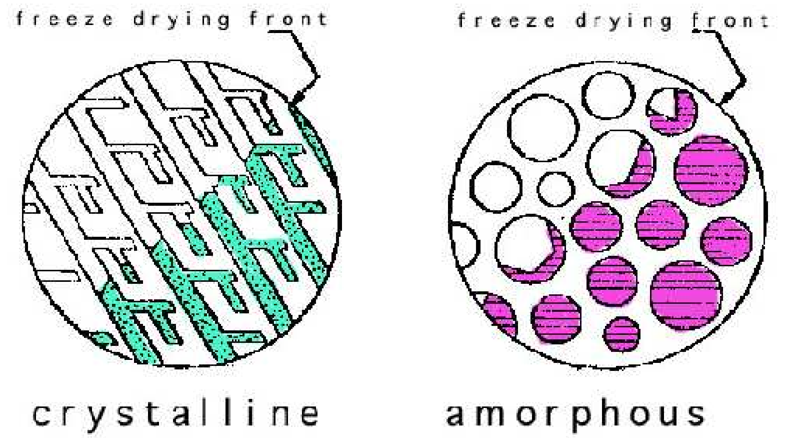

2. Speed of Information in DTN: Jacquet et al., 2008.

3. Real-Time Communication over Unreliable Wireless Channels with Hard Delay Constraints: P.R. Kumar & Hou, 2011.

4. The Quest for Fundamental Limits in MANET, Goldsmith, et al., 2011.

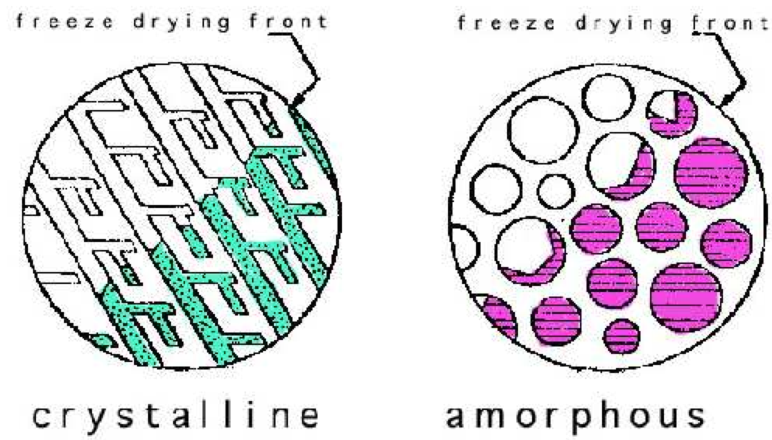
# Structure

**Structure**:
Measures are needed for quantifying information embodied in structures
(e.g., material structures, nanostructures, biomolecules, gene regulatory networks protein interaction networks, social networks, financial transactions).

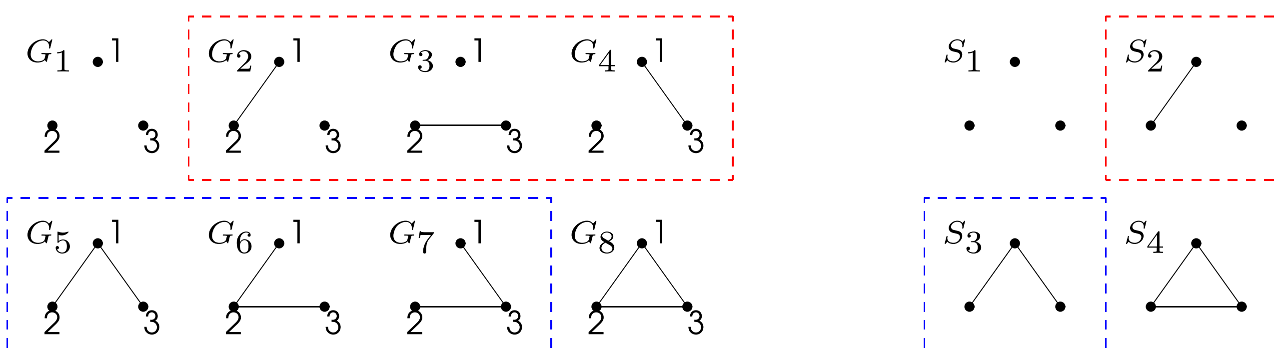


# Structure

**Structure**:
Measures are needed for quantifying information embodied in structures (e.g., material structures, nanostructures, biomolecules, gene regulatory networks protein interaction networks, social networks, financial transactions).




**Information Content of Unlabeled Graphs**:
A random structure model $\mathcal{S}$ of a graph $\mathcal{G}$ is defined for an unlabeled version. Some labeled graphs have the same structure.



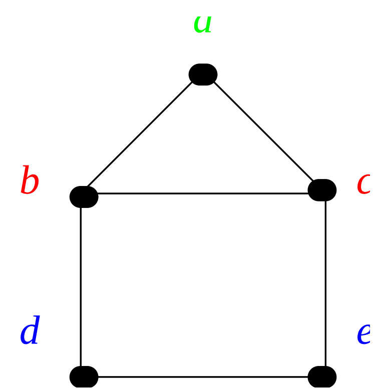

$$H_{\mathcal{G}} = \mathbf{E}[-\log P(G)] = -\sum_{G \in \mathcal{G}} P(G) \log P(G),$$

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

# Automorphism and Erdös-Rényi Graph Model

**Graph Automorphism**:

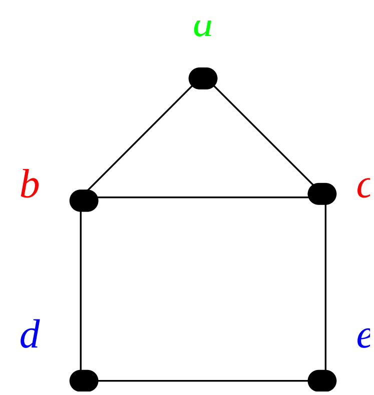
For a graph $G$ its automorphism
is adjacency preserving permutation
of vertices of $G$.



# Automorphism and Erdös-Rényi Graph Model

**Graph Automorphism**:

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




**Erdös and Rényi** model: $\mathcal{G}(n,p)$ generates graphs with $n$ vertices, where edges are chosen independently with probability $p$. If $G$ has $k$ edges, then

$$P(G) = p^k (1-p)^{\binom{n}{2}-k}.$$

**Theorem 7** (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.),*

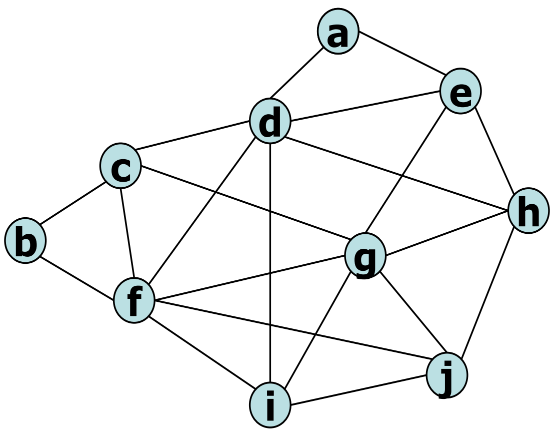
$$H_{\mathcal{S}} = \binom{n}{2} h(p) - \log n! + o(1) = \binom{n}{2} h(p) - n \log n + n \log e - \frac{1}{2} \log n + O(1),$$

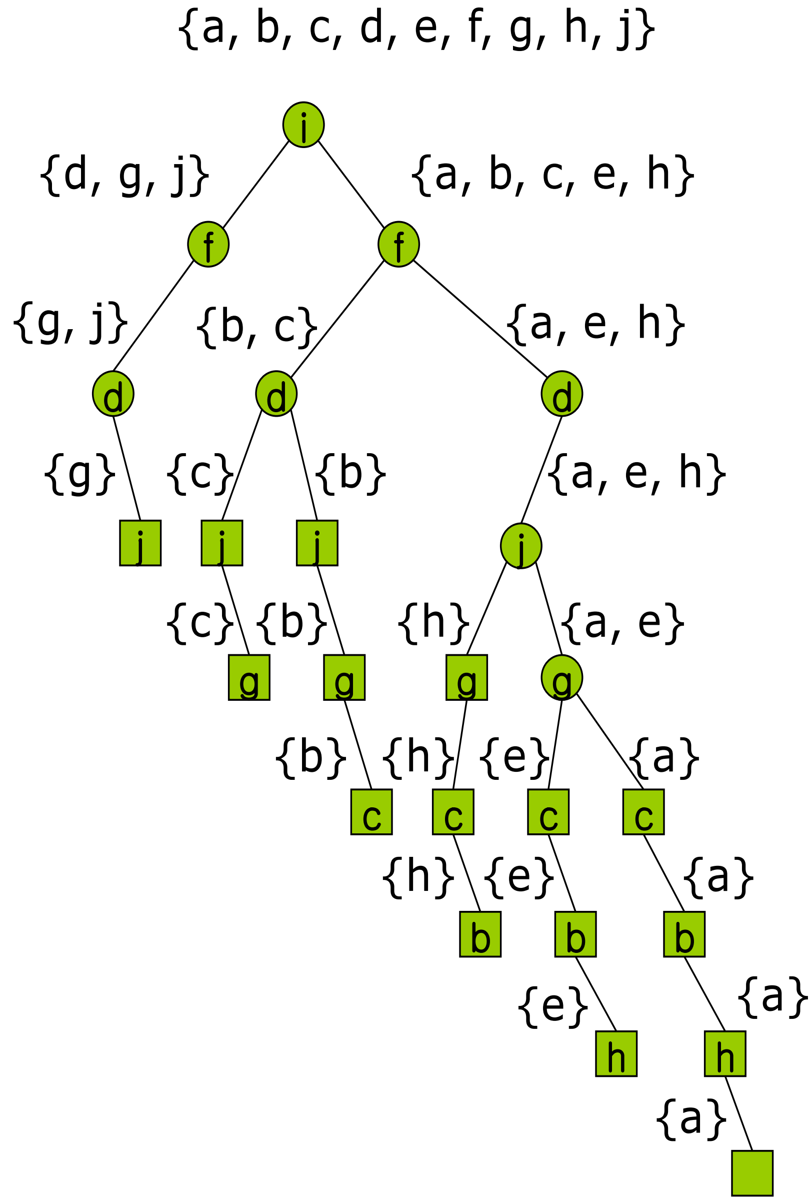
*where $h(p) = -p \log p - (1-p) \log(1-p)$ is the entropy rate.*

**AEP** *for structures:* $\quad 2^{-\binom{n}{2}(h(p)+\varepsilon)+\log n!} \le P(S) \le 2^{-\binom{n}{2}(h(p)-\varepsilon)+\log n!}.$

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

# Structural Zip (SZIP) Algorithm




B1 = 0100110100001110101

B2 = 100101100000101

# Asymptotic Optimality of SZIP

**Theorem 8** (Choi, W.S., 2008). *Let* $L(S)$ *be the length of the code.*

(i) *For large* $n$*,*

$$\mathbf{E}[L(S)] \le \binom{n}{2} h(p) - n \log n + n\,(c + \Phi(\log n)) + o(n),$$

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

(ii) *Furthermore, for any* $\varepsilon > 0$*,*

$$P\,(L(S) - \mathbf{E}[L(S)] \le \varepsilon n \log n) \ge 1 - o(1).$$

(iii) *The algorithm runs in* $O(n + e)$ *on average, where* $e$ *# edges.*

# Asymptotic Optimality of SZIP

**Theorem 8** (Choi, W.S., 2008). *Let $L(S)$ be the length of the code.*

(i) *For large $n$,*

$$\mathbf{E}[L(S)] \le \binom{n}{2} h(p) - n \log n + n \left(c + \Phi(\log n)\right) + o(n),$$

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

(ii) *Furthermore, for any $\varepsilon > 0$,*

$$P\left(L(S) - \mathbf{E}[L(S)] \le \varepsilon n \log n\right) \ge 1 - o(1).$$

(iii) *The algorithm runs in $O(n+e)$ on average, where $e$ # edges.*

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 | 37,0 18,710 | 729,848 | 310,569 |
| | Internet (AS level) | 25,881 | 52,407 | 301,148 | 334,900,140 | 1,572, 210 | 396,060 |

# Outline Update

1. Shannon Information Theory
2. Source Coding
3. The Redundancy Rate Problem
4. Post-Shannon Information
5. NSF Science and Technology Center on Science of Information

# NSF Center for Science of Information

In 2010 National Science Foundation established $25M

**Science and Technology Center for Science of Information**
(http: soihub.org)

to advance science and technology through a new quantitative understanding of the representation, communication and processing of information in biological, physical, social and engineering systems.

The center is located at Purdue University and partner istitutions include: Berkeley, MIT, Princeton, Stanford, UIUC, UCSD and Bryn Mawr & Howard U.

# NSF Center for Science of Information

In 2010 National Science Foundation established $25M

**Science and Technology Center for Science of Information**
(http: soihub.org)

to advance science and technology through a new quantitative understanding of the representation, communication and processing of information in biological, physical, social and engineering systems.

The center is located at Purdue University and partner istitutions include: Berkeley, MIT, Princeton, Stanford, UIUC, UCSD and Bryn Mawr & Howard U.

**Specific Center's Goals**:

- define core theoretical principles governing transfer of information.
- develop meters and methods for information.
- apply to problems in physical and social sciences, and engineering.
- offer a venue for multi-disciplinary long-term collaborations.
- transfer advances in research to education and industry.

# Acknowledgments

**My French Connection**:

**Philippe Flajolet** (1948-2011)
Analytic Combinatorics

# Acknowledgments

**My French Connection**:

**Philippe Flajolet** (1948-2011)
Analytic Combinatorics

**Philippe Jacquet**
My *copain* and accomplice for the last thirty something years in the analytic information theory adventure.

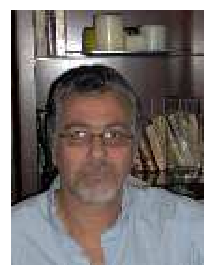
# Acknowledgments

**My French Connection**:

**Philippe Flajolet** (1948-2011)
Analytic Combinatorics

**Philippe Jacquet**
My *copain* and accomplice for the last thirty something years in the analytic information theory adventure.

**My Palo Alto Connection**:



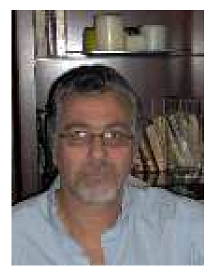
# Acknowledgments

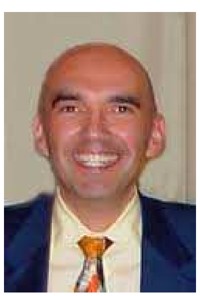
**My French Connection**:

**Philippe Flajolet** (1948-2011)
Analytic Combinatorics

**Philippe Jacquet**
My *copain* and accomplice for the last thirty something years in the analytic information theory adventure.

**My Palo Alto Connection**:



**Finally:**
The Master!

.... and ALL my co-authors.

# That's IT