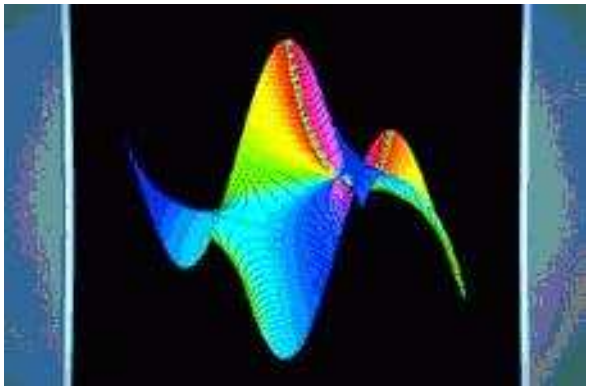
Information Transfer in Biological Systems*

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

July 13, 2008

AofA and **IT** logos



*Joint work with A. Grama, P. Jacquet, M. Koyoturk, and G. Seroussi.

Outline

1. Information Theory in Biology
2. Beyond Shannon
3. Information Transfer: Darwin Channel
   - Model
   - Capacity of the Noisy Constrained Channel
4. Information in Structures: Network Motifs
   - Biological Networks
   - Finding Biologically Significant Structures

Shannon Information in Biology

C. Shannon:
"These semantic aspects of communication are irrelevant . . ."

1. In 1949 Henry Quastler launched information theory in biology in "*The Information Content and Error Rate of Living Things*".
2. Henry Linschitz argued that these attempts were rather unsuccessful since there are difficulties in defining information of a system composed of functionally interdependent units and channel information (entropy) to produce a functioning cell.

> **Life** is a delicate interplay of energy, entropy, and information; essential functions of living beings correspond to the generation, consumption, processing, preservation, and duplication of **information**.

M. Eigen

"The differentiable characteristic of the living systems is Information. Information assures the controlled reproduction of all constituents, thereby ensuring conservation of viability . . . . Information theory, pioneered by **Claude Shannon**, cannot answer this question . . .
in principle, the answer was formulated 130 years ago by **Charles Darwin**.

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.

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

"Shannon performed an inestimable serviceby giving us Information.
We have **no theory** however that gives us a metric
for the Information embodied in **structure** . . .
this is the most fundamental gap in the theoretical underpinning of Information and computer science.

Information Transfer in Biology:

- how information is generated and transferred through underlying mechanisms of variation and selection?
- how information in biomolecules (sequences and structures) relates to the organization of the cell?
- whether there are error correcting mechanisms (codes) in biomolecules?
- and how organisms survive and thrive in noisy environments?

Beyond Shannon

Participants of the **2005 Information Beyond Shannon** workshop realize:

Time: When information is transmitted over networks of gene regulation, protein interactions, the associated delay is an important factor.
(e.g., timely information exchange in cells may be responsible for bidirectional microtubule-based transport in cells).

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. What is semantic information and how to characterize it? How much more semantic information is there when when compared with its syntactic information?

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

Physics of Information: Information is physical (J. Wheeler).

Outline Update

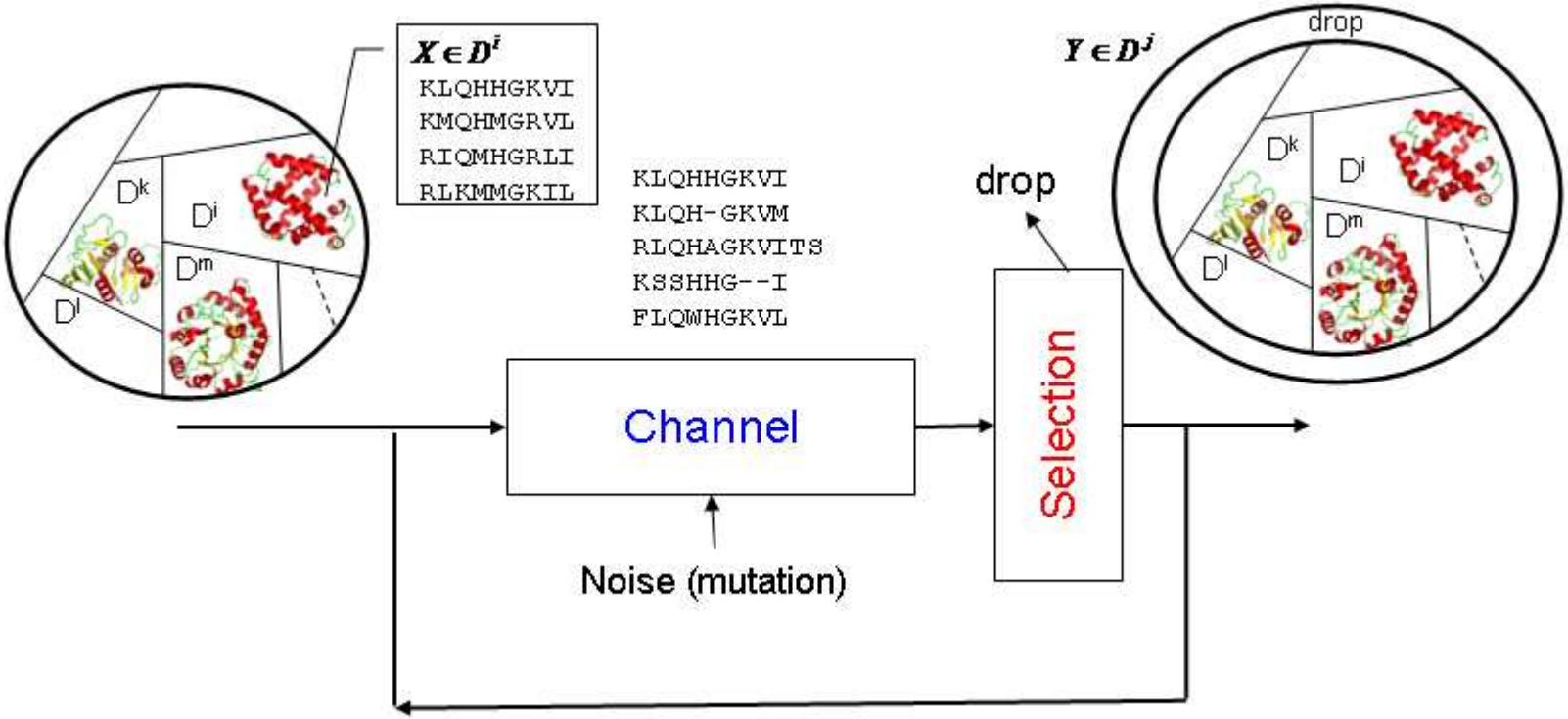
1. What is Information?
2. Beyond Shannon Information
3. Information Transfer: Darwin Channel
   - Model
   - Capacity of the Noisy Constrained Channel
4. Information in Structures: Network Motifs

Darwin Channel

Biomolecular structures, species, and in general biodiversity, have gone through significant metamorphosis over eons through mutation and natural selection, which we model by constrained sequences/channels.

To capture mutation and natural selection we introduce

Darwin channel.



Noisy Constrained Channel

1. Binary Symmetric Channel (BSC):
(i) crossover probability ε,
(ii) **constrained set of inputs** (Darwin preselected) that can be modeled by a **Markov Process**,
(ii) $\mathcal{S}_n$ denotes the set of binary constrained sequences of length n.

2. Channel Input and Output:
Input: Stationary process $X = \{X_k\}_{k\geq 1}$ supported on $\mathcal{S} = \bigcup_{n>0} \mathcal{S}_n$.
Channel Output: **Hidden Markov Process** (HMP)

$$Z_i = X_i \oplus E_i$$

where $\oplus$ denotes addition modulo 2, and $E = \{E_k\}_{k\geq 1}$, independent of X, with $P(E_i = 1) = \varepsilon$ is a Bernoulli process (noise).

Note: To focus, we illustrate our results on

$$\mathcal{S}_n = \{\text{(d,k) sequences}\}$$

i.e., no sequence in $\mathcal{S}_n$ contains a run of zeros of length ishorter than d or longer than k. Such sequences can model neural spike trains (no two spikes in a short time).

Noisy Constrained Capacity

$C(\varepsilon)$ – conventional BSC channel capacity $C(\varepsilon) = 1 - H(\varepsilon)$, where $H(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$.

$C(\mathcal{S}, \varepsilon)$ – noisy constrained capacity defined as

$$C(\mathcal{S}, \varepsilon) = \sup_{X \in \mathcal{S}} I(X; Z) = \lim_{n \to \infty} \frac{1}{n} \sup_{X_1^n \in \mathcal{S}_n} I(X_1^n, Z_1^n),$$

where the suprema are over all stationary processes supported on $\mathcal{S}$ and $\mathcal{S}_n$, respectively. **This is an open problem since Shannon.**

Mutual information

$$I(X; Z) = H(Z) - H(Z|X)$$

where $H(Z|X) = H(\varepsilon)$.

Thus, we must find the entropy $H(Z)$ of a **hidden Markov process**! (e.g., (d, k) sequence can be generated as an output of a kth order Markov process).

Entropy Rate as a Lyapunov Exponent

Theorem 1 (Furstenberg and Kesten, 1960). *Let* $\mathbf{M}_1, \ldots, \mathbf{M}_n$ *form a stationary ergodic sequence and* $\mathbf{E}[\log^+ ||\mathbf{M}_1||] < \infty$ *Then*

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{E}[\log ||\mathbf{M}_1 \cdots \mathbf{M}_n||] = \lim_{n\to\infty} \frac{1}{n} \log ||\mathbf{M}_1 \cdots \mathbf{M}_n|| = \mu \quad \text{a.s.}$$

where μ *is called top Lyapunov exponent.*

Corollary 1. *Consider the HMP* Z *as defined above. The entropy rate*

$$\begin{aligned} h(Z) &= \lim_{n\to\infty} \mathbf{E}[\, -\frac{1}{n} \log P(Z_1^n)\,] \\ &= \lim_{n\to\infty} \frac{1}{n}\mathbf{E}[-\log\left(\mathbf{p}_1 \mathbf{M}(Z_1, Z_2) \cdots \mathbf{M}(Z_{n-1}, Z_n)\mathbf{1}^t\right)] \end{aligned}$$

is a top Lyapunov exponent of some random matrices $\mathbf{M}(Z_1, Z_2) \cdots \mathbf{M}(Z_{n-1}, Z_n)$, *as shown in Jacquet, Seroussi, W.S., (2004, 2008).*

Unfortunately, it is notoriously difficult to compute top Lyapunov exponents as proved in Tsitsiklis and Blondel. Therefore, in next we derive an explicit asymptotic expansion of the entropy rate $h(Z)$.

Asymptotic Expansion

We now assume that $P(E_i = 1) = \varepsilon \to 0$ is small (e.g., $\varepsilon = 10^{-12}$ for mutation).

Theorem 2 (Seroussi, Jacquet and W.S., 2004). *Assume rth order Markov. If the conditional probabilities in the Markov process X satisfy*

$$P(a_{r+1}|a_1^r) > 0 \quad \textbf{IMPORTANT!}$$

for all $a_1^{r+1} \in \mathcal{A}^{r+1}$, then the entropy rate of Z for small ε is

$$h(Z) = \lim_{n\to\infty} \frac{1}{n} H_n(Z^n) = h(X) + f_1(P)\varepsilon + O(\varepsilon^2),$$

where

$$f_1(P) = \sum_{z_1^{2r+1}} P_X(z_1^{2r+1}) \log \frac{P_X(z_1^{2r+1})}{P_X(\bar{z}_1^{2r+1})} = \mathbb{D}\left(P_X(z_1^{2r+1}) || P_X(\bar{z}_1^{2r+1})\right),$$

where $\bar{z}^{2r+1} = z_1 \ldots z_r \bar{z}_{r+1} z_{r+2} \ldots z_{2r+1}$. In the above, $h(X)$ is the entropy rate of the Markov process X, $\mathbb{D}$ denotes the Kullback-Liebler divergence.

Examples

Example 1. Consider a Markov process with symmetric transition probabilities $p_{01} = p_{10} = p$, $p_{00} = p_{11} = 1-p$. This process has stationary probabilities $P_X(0) = P_X(1) = \frac{1}{2}$. Then

$$h(Z) = h(X) + f_1(p)\varepsilon + f_2(p)\varepsilon^2 + O(\varepsilon^3)$$

where

$$f_1(p) = 2(1-2p)\log\frac{1-p}{p}, \qquad f_2(p) = -f_1(p) - \frac{1}{2}\left(\frac{2p-1}{p(1-p)}\right)^2 .$$

Example 2. (Degenerate Case.) Consider the following Markov process

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ 1 & 0 \end{bmatrix}$$

where $0 \le p \le 1$.
Ordentlich and Weissman (2004) proved for this case

$$H(Z) = H(P) - \frac{p(2-p)}{1+p}\varepsilon\log\varepsilon + O(\varepsilon)$$

(e.g., $(11\ldots)$ will not be generated by MC, but can be outputed by HMM with probability $O(\varepsilon^\kappa)$).

Main Asymptotic Results

We observe (cf. Han and Marcus (2007))

$$H(Z) = H(P) - f_0(P)\varepsilon \log \varepsilon + f_1(P)\varepsilon + o(\varepsilon)$$

for explicitly computable $f_0(P)$ and $f_1(P)$.

Let $P^{\max}$ be the maxentropic maximizing $H(P)$. Then

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) - (1 - f_0(P^{\max}))\varepsilon \log \varepsilon + (f_1(P^{\max}) - 1)\varepsilon + o(\varepsilon)$$

where $C(\mathcal{S})$ is known capacity of a noiseless channel.

Example: For (d, k) sequences, we can prove:

(i) for $k \leq 2d$

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) + A \cdot \varepsilon + O(\varepsilon^2 \log \varepsilon)$$

(ii) For $k > 2d$

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) + B \cdot \varepsilon \log \varepsilon + O(\varepsilon),$$

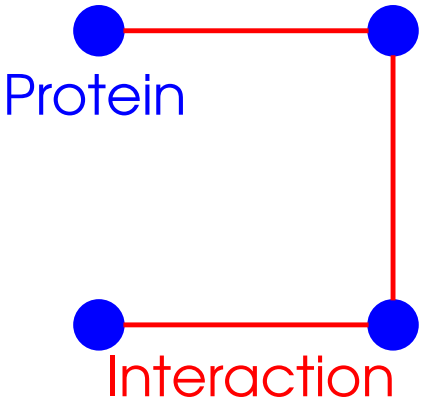
where $A \& B$ are computable constants (cf. also Han and Marcus (2007)).

Outline Update

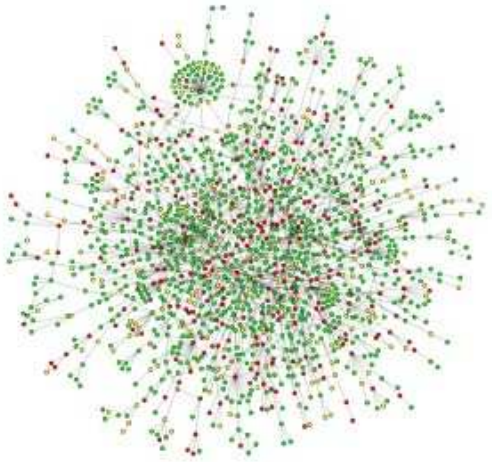
1. What is Information?
2. Information Transfer: Darwin Channel
3. Information in Structures: Network Motifs
   - Biological Networks
   - Finding Biologically Significant Structures

Protein Interaction Networks

- Molecular Interaction Networks: Graph theoretical abstraction for the organization of the cell
- Protein-protein interactions (PPI Network)
  - Proteins signal to each other, form complexes to perform a particular function, transport each other in the cell...
  - It is possible to detect interacting proteins through high-throughput screening, small scale experiments, and *in silico* predictions



Undirected Graph Model

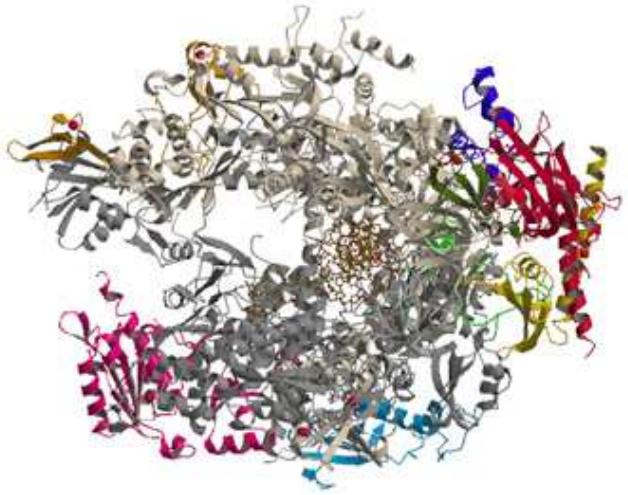


S. Cerevisiae PPI network hspace0.3in

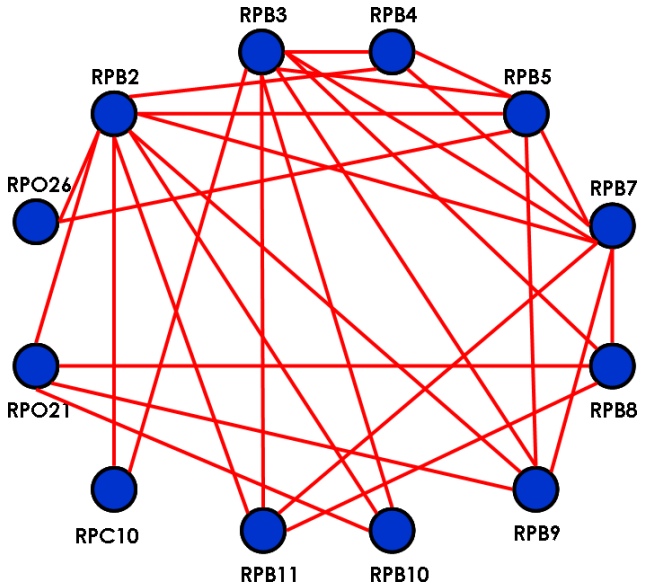
(Jeong et al., *Nature*, 2001)

Modularity in PPI Networks

- A functionally modular group of proteins (*e.g.* a protein complex) is likely to induce a dense subgraph
- Algorithmic approaches target identification of dense subgraphs
- An important problem: How do we define dense?
  - Statistical approach: What is significantly dense?



RNA Polymerase II Complex



Corresponding induced subgraph

Significance of Dense Subgraphs

- A subnet of r proteins is said to be ρ-dense if the number of interactions, $F(r)$, between these r proteins is $\geq \rho r^2$, that is,

$$F(r) \geq \rho r^2$$

- What is the expected size, R_ρ, of the largest ρ-dense subgraph in a random graph?
  - Any ρ-dense subgraph with larger size is statistically significant!
  - Maximum clique is a special case of this problem ($\rho = 1$)

- $G(n, p)$ model
  - n proteins, each interaction occurs with probability p
  - Simple enough to facilitate rigorous analysis

- Piecewise $G(n, p)$ model
  - Captures the basic characteristics of PPI networks

- Power-law model

Largest Dense Subgraph on $G(n,p)$

Theorem 4. *If G is a random graph with n nodes, where every edge exists with probability p, then*

$$\lim_{n\to\infty} \frac{R_\rho}{\log n} = \frac{1}{H_p(\rho)} \qquad (pr.),$$

where

$$H_p(\rho) = \rho \log \frac{\rho}{p} + (1-\rho) \log \frac{1-\rho}{1-p}$$

denotes divergence. More precisely,

$$P(R_\rho \geq r_0) \leq O\left(\frac{\log n}{n^{1/H_p(\rho)}}\right),$$

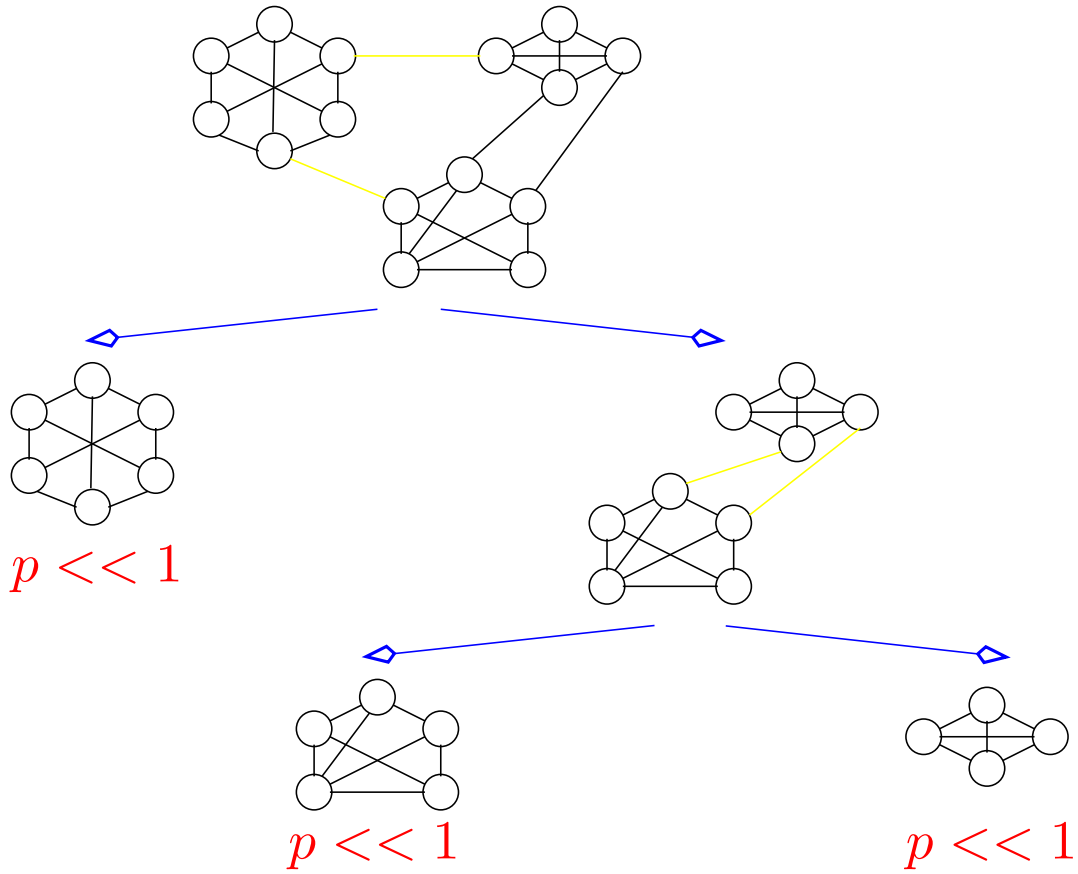
where

$$r_0 = \frac{\log n - \log\log n + \log H_p(\rho)}{H_p(\rho)}$$

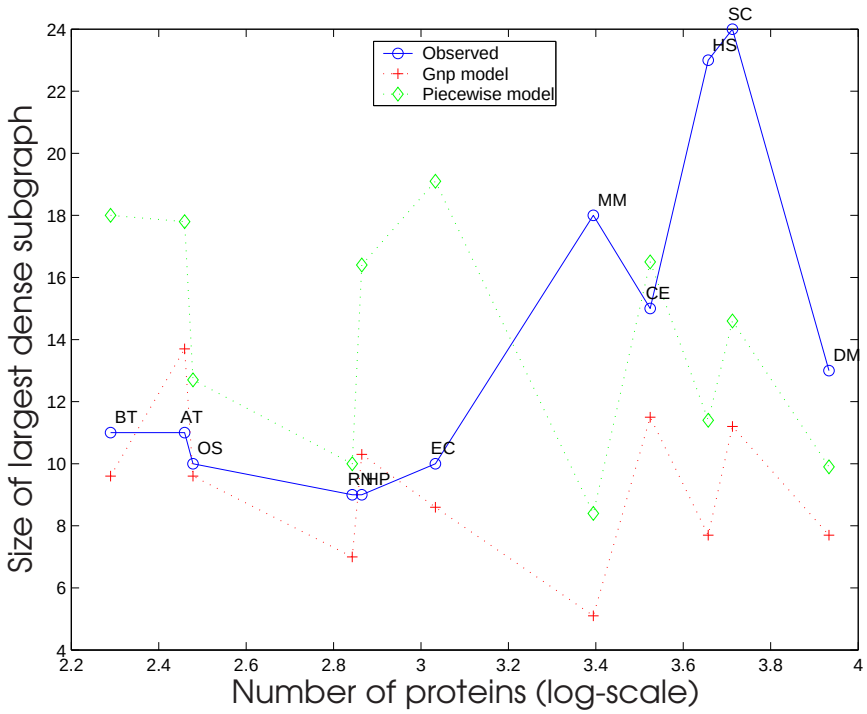
for large n.

SIDES

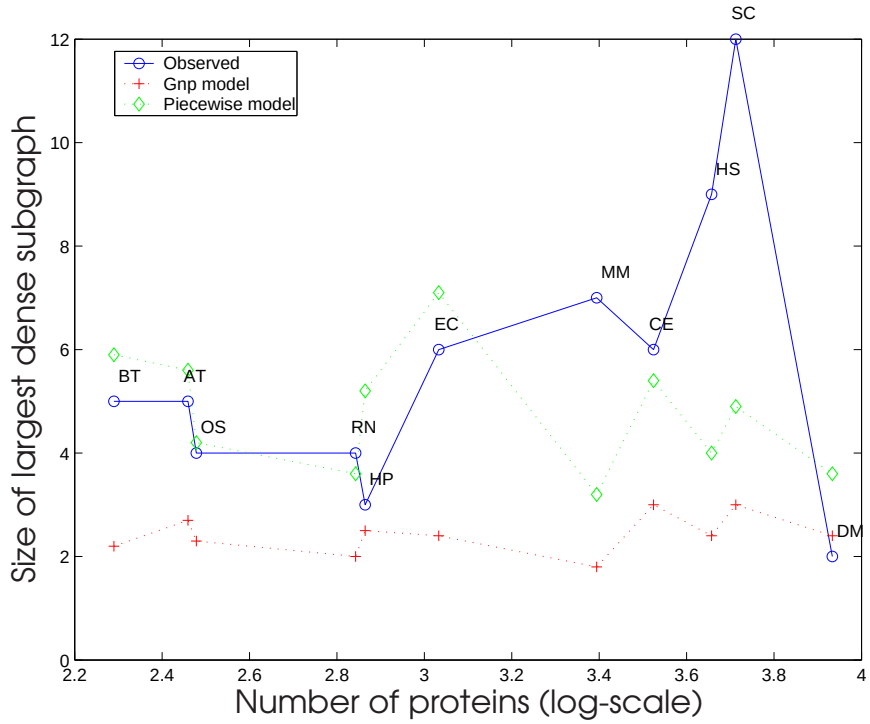
- An algorithm for identification of Significantly Dense Subgraphs (SIDES)
  - Based on Highly Connected Subgraphs algorithm (Hartuv & Shamir, 2000)
  - Recursive min-cut partitioning heuristic
  - We use statistical significance as stopping criterion



Behavior of Largest Dense Subgraph Across Species



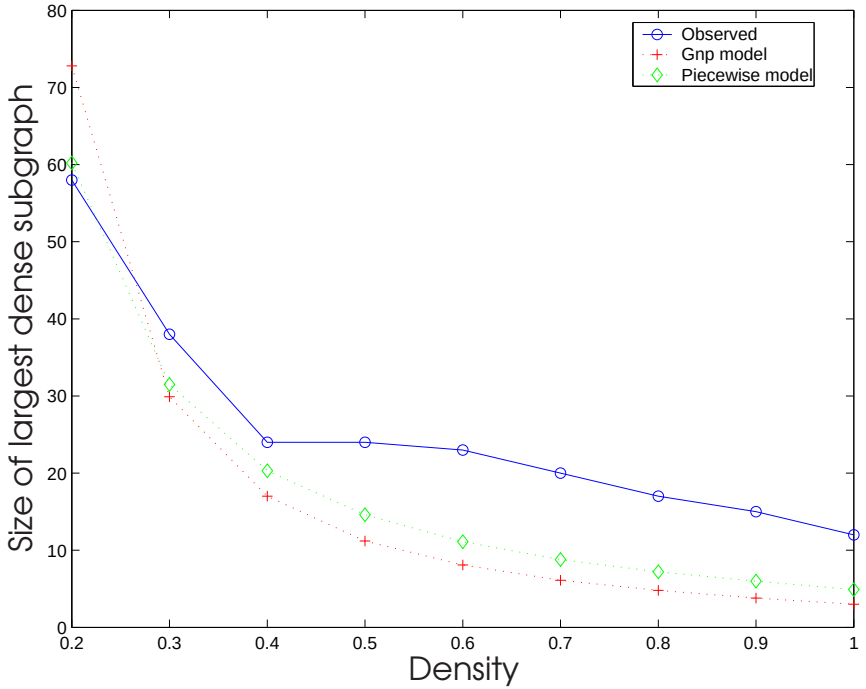
$\rho = 0.5$



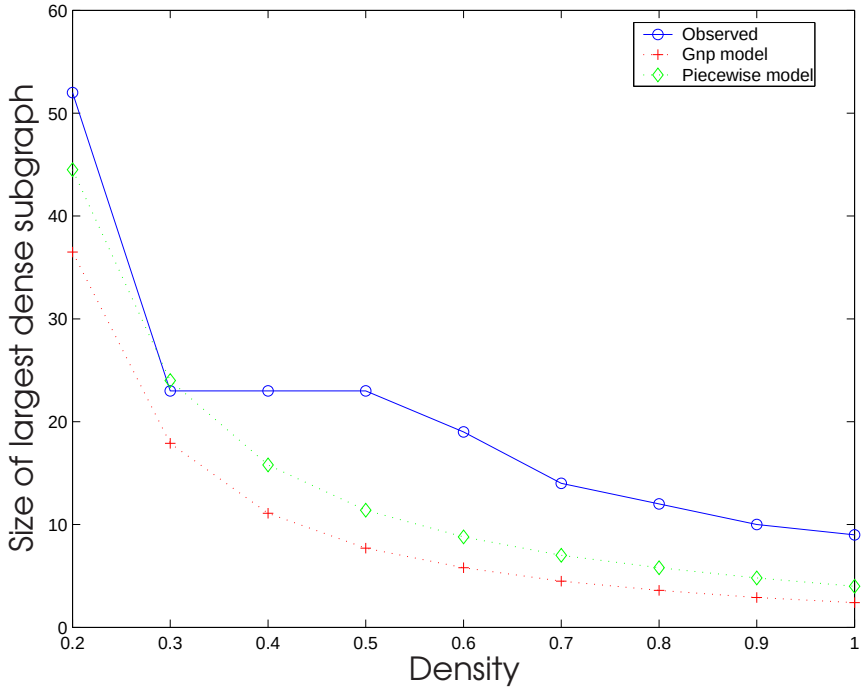
$\rho = 1.0$

Number of nodes vs. Size of largest dense subgraph
for PPI networks belonging to 9 Eukaryotic species

Behavior of Largest Dense Subgraph w.r.t Density



S. cerevisiae



H. sapiens

Density threshold vs. Size of largest dense subgraph
for Yeast and Human PPI networks