EE430
Introduction to Systems Biology

Week 5 Course Notes

Instructor: Bilge Karaçalı, PhD
Topics

• Network motifs in transcription regulation
  – Statistical characterization of network motifs
  – Single input modules
  – Multi-output feed forward loops
  – Bi-fans and dense overlapping regulons
Statistically Significant Network Properties

- A property $\mathcal{P}$ observed in a real gene transcription network is **noteworthy** if the probability of observing it in a comparable random network is small
  - The actual gene transcription networks are not random networks
  - If the chance probability is small, there would be reason to believe that the property arose due to a functional benefit selected for during the evolutionary mechanism
  - A great part of systems biology is identifying such noteworthy biomolecular network properties and describing their functional characteristics
Optimal Gene Transcription Networks

- Gene transcription networks are not wired randomly
  - The connections produce a specific output: a functioning biomolecular mechanism
  - The properties of the connections are optimized for
    - efficiency
    - adaptability
    - response speed

- The design parameters are determined as a result of selection
  - Number of nodes: the number of genes in the gene transcription regulation
  - Number of edges: the number of regulatory interactions between the genes
  - The organization of the edges between the nodes

- Understanding and modeling gene transcription networks requires
  - Identifying the network structures that are repeatedly used
  - Evaluating their dynamic behavior and kinetic properties
Gene Transcription Network Motifs

• The modules observed in a conspicuously great number in gene transcription networks are called network motifs
  – Conspicuousness determined in terms of statistical significance
  – Number of observations much greater than the case in a comparable random network
    • Same number of nodes
    • Same number of edges
Erdos-Renyi Random Networks

• Suppose we have a gene transcription network with \( N \) nodes and \( E \) edges that possesses the network property \( \mathcal{P} \).

• Consider the set of all possible random networks also with \( N \) nodes and \( E \) edges.
  - Each of these networks either does or does not possess the same property \( \mathcal{P} \).

• Question: If we were to select one network from this set, what is the probability that it also possesses the property \( \mathcal{P} \)?
  - Two sub-questions to be answered:
    • How many random networks are there?
    • How many of those possess the property?
  - If this probability is large, then \( \mathcal{P} \) is a common property of such networks (with \( N \) nodes and \( E \) edges).
  - If the probability is small, then \( \mathcal{P} \) is a rare property, and there must be a reason for its presence in the gene transcription network in consideration.
Erdos-Renyi Random Networks

- Consider arbitrarily constructed networks with
  - $N$ nodes and
  - $E$ directed edges
    - Including self edges
- The probability $P(\mathcal{P})$ that such a network has a given property $\mathcal{P}$ is given by

$$
P(\mathcal{P}) = \frac{\text{(\# of networks of } \ N \ \text{nodes and } \ E \ \text{edges with } \mathcal{P})}{\text{(\# of networks of } \ N \ \text{nodes and } \ E \ \text{edges})}
$$

- Note that the number of networks with $N$ nodes and $E$ edges is $(N^2)!/(E!(N^2-E)!)$
  - In a network if $N$ nodes, there are a total of $N^2$ places where a directed edge can be placed
    - $N(N-1)/2$ different undirected edges
    - $N(N-1)$ different directed edges
    - $N$ possible self-edges
  - Over $N^2$ possibilities, the number of ways in which the $E$ edges can be selected is therefore $(N^2)!/(E!(N^2-E)!)$
- These random networks are called Erdos-Renyi networks
Example: Is Autoregulation a Network Motif?

- In the *Escherichia Coli* transcription network, there are
  - \( N = 424 \) nodes
  - \( E = 519 \) edges
  - \( K = 40 \) self-edges
- Let the property \( \mathcal{P} \) be that a (randomly selected comparable Erdos-Renyi) network has \( K \) self edges
- The total number of comparable Erdos-Renyi networks is \( \frac{N^2}{2}! / (E!(N^2-E)!) \)
- A particular collection of \( K \) self edges is observed in a total of \( \frac{N(N-1)}{2}! / (E!(N(N-1)-E')!) \) random networks where
  \[ E' = E - K \]
  - After reserving \( K \) edges to be self-edges, there remains \( E - K \) non-self edges
- Since there are \( N! / ((N-K)!K!) \) distinct collections of \( K \) self edges, the probability \( P(\mathcal{P}) \) that a randomly selected Erdos-Renyi network has \( K \) self edges is
  \[ P(\mathcal{P}) = \frac{N(N-1))! \cdot N! \cdot (E!(N^2-E)!) / (E!(N(N-1)-E')!) \cdot ((N-K)!K!) \cdot (N^2)!}{< 10^{-46}} \]
  ➔ It is extremely unlikely for 40 self edges to have come about by chance in such a gene transcription network!!!
Example: Is Autoregulation a Network Motif?

• Alternative analysis:
  – The probability that a randomly selected edge is a self-edge is $1/N$
    • The probability that the end-point of the edge is the same as its beginning
  – Thus, the average number of self edges in an Erdos-Renyi network with $N = 424$ nodes and $E = 519$ edges is
    $$E / N = 1.2241$$
  – Since the number $k$ of self edges in such Erdos-Renyi networks follows a Poisson distribution, the probability density $p(k)$ is
    $$p(k) = v^k \exp(-v)/k!$$
    with $v = 1.2241$
  – Then, the probability for random networks with 40 self edges is
    $$p(40) = 1.1717 \times 10^{-45}$$
  ➔ Accidental observation of 40 self edges in an Erdos-Renyi network is highly unlikely
Example: Is Autoregulation a Network Motif?

• Consider the slightly modified question:
  “What is the probability that a random Erdos-Renyi network with $N$ nodes and $E$ edges has at least $K$ self edges?”
  – A random network having $K$ self edges may be unlikely, but how about having $K + 1$ self edges?
  – Conversely, is the probability of having $K$ self edges low because $K$ is too high or because $K$ is too low?
  – The probability of having at least $K$ edges should still be low if indeed $K$ is too high to be expected from a random Erdos-Renyi network

• Let
  – $\mathcal{N}_{N,E}$ denote a random Erdos-Renyi network with $N$ nodes and $E$ edges
  – $o_{se}(\mathcal{N})$ denote the operator that determines the number of self edges in a network $\mathcal{N}$

• The inquired probability becomes
  \[
  \Pr\{o_{se}(\mathcal{N}_{N,E}) \geq K \} = \Pr\{o_{se}(\mathcal{N}_{N,E}) = K \} + \Pr\{o_{se}(\mathcal{N}_{N,E}) = K + 1 \} + \ldots + \Pr\{o_{se}(\mathcal{N}_{N,E}) = \min\{N,E\} \}.
  \]
Example: Is Autoregulation a Network Motif?

- Also known is
  \[ \text{Pr}\{o_{se}(N,E) = K \} = \frac{(N(N-1))! \cdot N! \cdot (E!(N^2-E)!)}{(E'!(N(N-1)-E')!)(N-K)!K! \cdot (N^2)!} \]

- Thus, for the self edges in *Escherichia Coli*,
  \[ \text{Pr}\{o_{se}(n_{424,519}) \geq 40 \} = 4.7089 \times 10^{-47} \]

\[
\text{Pr}\{o_{se}(n_{N,E}) = K \} = \sum_{k=K}^{\min\{N,E\}} \frac{N(N-1)}{E-K} \cdot \frac{\min\{N,E\}}{K} \cdot \frac{N^2}{E} \cdot \frac{\binom{N^2}{E}}{E!((N-K)!K!)}
\]

\[
\text{Pr}\{o_{se}(n_{N,E}) \geq K \} = \sum_{k=K}^{\min\{N,E\}} \text{Pr}\{o_{se}(n_{N,E}) = k \} = 1 - \sum_{k=0}^{K-1} \text{Pr}\{o_{se}(n_{N,E}) = k \}
\]
Example: Are Feed-Forward Loops Network Motifs?

• Consider the same Escherichia Coli transcription network with
  – $N = 424$ nodes
  – $E = 519$ edges

• Question: What is the statistical significance of this network possessing 42 feed-forward loops?

• Procedure:
  – Compute the average number of feed-forward loops observed in comparable Erdos-Renyi networks
  – Use the Poisson distribution assumption for the number of feed-forward loops in a random network
  – Find the $P$ value for the observed number of feed-forward loops
    \[
    P = \Pr\{O_{\text{ffl}}(N,E) \geq 42\}\]
Example: Are Feed-Forward Loops Network Motifs?

- What is the average number of feed-forward loops observed in $N$-node $E$-edge Erdos-Renyi networks?
  - The number of networks with the feed-forward loop $1 \rightarrow 2, 2 \rightarrow 3, \text{ and } 1 \rightarrow 3$ is $\mathbf{C}(N^2-3,E-3)$
  - In such a network, the number of distinct feed forward loops is $N(N-1)(N-2)$
  - Therefore, in $\mathbf{C}(N^2,E)$ Erdos-Renyi networks, there are $N(N-1)(N-2)\mathbf{C}(N^2-3,E-3)$ feed forward loops
  - Thus, the average number $\nu$ of feed-forward loops is
  \[
  \nu = \frac{N(N-1)(N-2)\mathbf{C}(N^2-3,E-3)}{\mathbf{C}(N^2,E)} = \frac{N(N-1)(N-2)E(E-1)(E-2)}{(N^2(N^2-1)(N^2-2))}
  \]
Example: Are Feed-Forward Loops Network Motifs?

- With $N = 424$ and $E = 519$,
  - $\nu = 1.81$
  - $P = 1.11 \times 10^{-16}$

The feed forward loop in the E. Coli gene transcription network is a network motif!!
Single Input Modules

• A single transcription factor regulating several other genes that have no other transcription factors
  – The global transcription factor is also usually autoregulatory

• Single input modules are network motifs
  – The average number of connections per node in an Erdos-Renyi network is $\lambda = \frac{E}{N}$
    • $\lambda$ is the mean connectivity
  – Using the Poisson distribution assumption, it turns out that a node with more than a few connections is extremely rare
    • $\lambda$ is usually close to 1
      – Gene transcription networks are conspicuously sparse
Single Input Modules

- Consider a single input module with four genes regulated by one transcription factor
- Kinetic analysis:
  - Premises:
    - $[X]$ is initially zero
    - At time $t = 0$, $X$ starts to be expressed via simple regulation
    - $\kappa_{Z1} < \kappa_{Z2} < \kappa_{Z3} < \kappa_{Z4}$
  - Dynamic evaluation:
    - At time $t = 0$, $[X]$ starts to rise following an exponential curve towards its steady-state
    - As $[X]$ crosses the thresholds $\kappa_{Z1}$, $\kappa_{Z2}$, $\kappa_{Z3}$, $\kappa_{Z4}$, it triggers the expressions of the corresponding genes
    - When $X$ is turned off, the regulated genes are turned off in the reverse order
      ➔ A first-in-last-out (FILO) order

\[ X \]
\[ Z_1 \]
\[ Z_2 \]
\[ Z_3 \]
\[ Z_4 \]
Single Input Modules

$\beta_{X/\alpha_X}$

$\beta_{Z_1/\alpha_{Z_1}}$

$\beta_{Z_2/\alpha_{Z_2}}$

$\beta_{Z_3/\alpha_{Z_3}}$

$\beta_{Z_4/\alpha_{Z_4}}$
Single Input Modules

• Single input modules allow temporal control of gene expression
  – The times at which genes start expression can be fine-tuned by adjusting their corresponding $\kappa$ thresholds
• The start order of expression is reversed in the cessation of expression
  – First-in-last-out
  – This creates suboptimal situations of expressing genes required in the earlier phases all the way to the end of the process
Multi-Output Feed-Forward Loops

• The feed-forward loop can be generalized in three ways:
  – Multiple global regulators
  – Multiple middle regulators
  – Multiple global regulatees

• The multiplicity of the last node produces the multi-output feed-forward loops
  – The multi-output feed-forward loop is a network motif in gene transcription networks
Multi-Output Feed-Forward Loops

- Kinetic analysis:
  - Premises:
    - Both \([X]\) and \([Y]\) are initially zero
    - The expression of \(X\) starts with simple regulation at time \(t = 0\)
    - \(\kappa_{XZ_1} < \kappa_{XZ_2} < \kappa_{XZ_3}\)
    - \(\kappa_{YZ_3} < \kappa_{YZ_2} < \kappa_{YZ_1}\)
  - Dynamic evaluation:
    - At time \(t = 0\), \([X]\) starts to rise, triggering the expressions of \(Z_1, Z_2,\) and \(Z_3,\) as well as \(Y\)
    - The expression order of \(Z_1, Z_2,\) and \(Z_3\) follows that of \(\kappa_{XZ_1}, \kappa_{XZ_2},\) and \(\kappa_{XZ_3}\)
    - After a while when \(X\) is turned off, the expressions of \(Z_1, Z_2,\) and \(Z_3\) are controlled by \(Y\)
    - The order in which \(Z_1, Z_2,\) and \(Z_3\) are turned off follows that of \(\kappa_{YZ_3}, \kappa_{YZ_2}, \kappa_{YZ_1}\)
Multi-Output Feed-Forward Loops
Multi-Output Feed-Forward Loops

• The multi-output feed-forward loops also create temporal order of gene expression
• The additional control node allows maintaining the order in which the genes are sequentially turned on in the turn off as well
  – First-in-first-out (FIFO) ordering
Bi-Fans and Dense Overlapping Regulons

• The final motif in gene transcription networks consists of a set of genes jointly regulating a second set of genes
  – The underlying structure is the bi-fan where a pair of genes jointly regulate another pair of genes
  – Generalization of this structure produces a family of motifs with overlapping regulatory connections, termed dense overlapping regulons
    • A regulon is the set of genes regulated by the same transcription factor
Dense Overlapping Regulons

• By definition, dense overlapping regulons consist of two layers of nodes:
  – The regulating layer
  – The regulated layer

• Each node in the regulated layer realizes a combinatorial logic function of the associated nodes in the input layer
  – AND
  – OR
  – NAND
  – NOR
  – …
Dense Overlapping Regulons

\[ Z_1 = (X_1 \text{ AND } X_2) \text{ AND } X_3 \]
\[ Z_2 = \text{NOT}((X_1 \text{ OR } X_2) \text{ OR } (\text{NOT } X_3)) \]
\[ Z_3 = (\text{NOT } X_1) \text{ OR } ((\text{NOT } X_2) \text{ AND } X_3) \]
\[ Z_4 = \text{NOT}(X_1 \text{ AND } (X_2 \text{ OR } (\text{NOT } X_3))) \]
Dense Overlapping Regulons

- The input function of the regulated genes are multi-dimensional input functions that integrate the inputs from the regulating genes
  - Not all regulating genes necessarily regulate each and every gene in the regulated layer
  - The shapes of input functions as well as the regulatory connections are subject to selection
- Dense overlapping regulons operate as decision making devices, integrating inputs from many sources in order to determine a course of action
  - Large dense overlapping regulons controlling the expressions of tens to hundreds of genes are common in gene transcription networks
  - Genes in dense overlapping regulons share the same function such as
    - stress response
    - nutrient metabolism
    - ...
Global Structure of Gene Transcription Networks

• Along with dense overlapping regulons, the intricate combination of autoregulatory control, feed-forward loops, and single input modules constitute gene transcription networks
  – The backbone of the network is formed by several dense overlapping regulons that operate as decision making mechanisms
  – The remaining three motifs carry out the decisions made through the dense overlapping regulons with
    • Robustness against noisy input signals
    • Temporal process coordination and synchronization
    • Rapid response amid production and decay constraints
• Virtually all genes in all organisms participate in one of these network motifs
• Thus, the combined evaluation of these network motifs should suffice to provide a working assessment of the functioning of any gene transcription network
Summary

• Gene transcription networks consist of four basic network motifs
  – Autoregulation: speed-up the response time and permanent decision making
  – Feed-forward loops: sign-sensitive filtering, speed-up the response time, and pulse generation
  – Single input modules: FILO and FIFO temporal programs
  – Dense overlapping regulons: logic computations

• Not only that these network elements are observed much more frequently in gene transcription networks, they account for virtually all regulatory interactions
  – Significance of occurrence frequency evaluated against that in comparable random networks

• The kinetic evaluation and computer simulations allow understanding how these network elements operate in isolation as well as parts of a larger, connected system