Network Structure

Is there structure in this network, or is it “random”?

http://cs.brown.edu/courses/csci2950-c/
Network Motifs

Subnetworks with more occurrences than expected by chance.

• How to find?
• How to assess statistical significance?

Shen-Orr et al. 2002

Network Motifs

Subnetworks with more occurrences than expected by chance.

• How to find?
  – Exhaustive: Count all $k$-node subgraphs.
  – Heuristic methods: sampling, greedy, etc.
  – Approximate counting via randomized algorithms.
Counting Subnetworks

\[ G = (V,E). \ |V| = n. \ |E| = m. \]

- Network-centric approach
  - Count/enumerate all subgraphs with \( \leq k \) vertices.
  - Impractical for large \( n, m, k \)

- Query-based approach
  - Enumerate query graphs \( Q \).
  - For each \( Q \), count occurrences. (Subgraph isomorphism)
  - \( Q \) could be a non-induced subgraph.

Counting non-induced subgraphs

Suppose want to count \textbf{paths} in \( G = (V,E) \).

Idea: use color-coding to count colorful paths
  - Dynamic programming solution (Whiteboard)

Can extend dynamic program to count trees and bounded treewidth graphs.
Relation between Forward and Viterbi

**VITERBI**

Initialization:
\[ V_0(0) = 1 \]
\[ V_k(0) = 0, \text{ for all } k > 0 \]

Iteration:
\[ V_j(i) = e_j(x_i) \max_k V_k(i-1) a_{kj} \]

Termination:
\[ P(x, \pi^*) = \max_k V_k(N) \]

**FORWARD**

Initialization:
\[ f_j(0) = 1 \]
\[ f_i(0) = 0, \text{ for all } k > 0 \]

Iteration:
\[ f_i(i) = e_i(x_i) \sum_k f_k(i-1) a_{ki} \]

Termination:
\[ P(x) = \sum_k f_k(N) a_{k0} \]

Importance of Network Motifs

- Building block of networks.
- Indicate modular structure of biological networks.
- Appearance of some motifs might be explained by particular dynamics (e.g. feedforward and feedback loops)

Healthy skepticism about all these claims, particularly because data is incomplete.
Sources

Network Inference

Given data (e.g. protein phosphorylation state), what is the network?

Mukherjee S, Speed T P PNAS 2008;105:14313-14318

©2008 by National Academy of Sciences
Posterior mode for protein data.

Prior sensitivity for protein data.