
Presenter: Mitra Ansariola
Advisor: Dr. Molly Megraw
Collaborative project with: Dr. David Koslicki
Network Motif Discovery

- Count the number of times subgraph M is observed in network N.

- Generate a collection of background networks RN.

- Is Subgraph M over-represented in the original network N as compared to background networks?
Why Network Motif Discovery is Important for Biology

- A network motif is an over-represented small subgraph in a network of interest.
- Network motif discovery helps to generate testable hypotheses about potential functional elements in a living cell.

Appropriate background network generation is key to a biologically meaningful network motif outcome.
Background Network Sampling Challenges

- How to generate background networks?
  - Performing edge permutations in such a way that preserves the in/out degrees of nodes in the original network

- Constraint on edge permutations: No multiple edges

- Challenges:
  - No closed form formula is available to compute the number of all possible permutations
  - No systematic way of producing such permutations is available
  - For a realistic graph with hundreds of nodes and edges, number of possibilities is enormous (trillions of possibilities)
IndeCut: Estimates the degree of uniformity/independence for a sample

All possible edge permutations for in-degree=2,1,1 and out-degree=2,1,1

Sample space

Sample Average

Sample graph

E: Expected Average of Sample space

Degree of uniformity/independence = \( \|E - A\|_c \)

Maximum Entropy Matrix is a proxy to E

Approximate cut norm \( \|Z - A\|_c \) with \( \|Z - A\|_{\infty \to 1} \)

Cut norm bounds

Cut norm bounds closer to zero \( \Rightarrow \) The more uniform/independent the sample

(A) Uniform, independent

(B) Non-uniform, not independent

(C) Uniform, not independent
Results of Using IndeCut

- IndeCut is the very first practical method for determining the degree of uniformity/independence of network motif discovery algorithm sampling.

- We used IndeCut to evaluate sampling performance of four published network motif discovery methods on real-world biological networks.
  - Results show many algorithms have difficulty in performing uniform/independent sampling on large biological networks containing master regulators.

Please visit our poster P03 on FRIDAY, DECEMBER 8, 7:30PM – 8:30PM