**Motif-based Clustering of Directed Networks**

Thomas P. Reith, Laurie J. Heyer

Mathematics Department, Davidson College

---

**Network Motifs**
- Recurrent, statistically significant subgraphs found at higher frequencies in real networks than randomly generated networks
- Can be thought of as “simple building blocks of complex networks”

*Feedforward Loop*

*Multi-Input Motif*

---

**Clustering Criteria**
- Given a motif, we consider nodes that play the same role in different occurrences of the motif across the network to be more similar
- Nodes that never play the same role are less similar
- To quantify this similarity, we count the number of times each node of the network plays each specific role in the given motif

*Simple Example*

**Algorithm Steps**
1. Create an $m \times n$ matrix $R$, where $m$ is the number of nodes in the network, $n$ is the number of roles in the motif, and each entry $R_{ij}$ is the number of times that node $i$ plays role $j$.

If multiple role matrices are created based on different motifs and concatenated, the network can be clustered based on multiple motifs.

2. Apply agglomerative hierarchical clustering to the rows of $R$. Distance metrics and linkage criteria are user-defined.

*Node Roles*
- We define the role of a node as a unique combination of its in and out degrees
- Feedforward loop has three different roles:
  - “sender”
  - “middleman”
  - “receiver”
- Multi-Input Motif only has two roles:
  - “sender”
  - “receiver”

---

**Subgraph Isomorphism**
- How do we find all instances of a given motif in a network?
- Our method for directed graphs is extended from Ullmann’s method for undirected graphs
- Suppose there is an isomorphism between a graph (motif) $G_1$ and some subgraph of a larger graph $G_2$.
- We can encode this isomorphism in an $m \times n$ permutation matrix $M$, where $m = |V(G_1)|$ and $n = |V(G_2)|$.
- If vertex $i \in G_2$ maps to vertex $j \in G_1$, then we set $m_{ij} = 1$.

- Let $A$ and $B$ be the adjacency matrices for $G_1$ and $G_2$.
- $M$ encodes an isomorphism iff $A \equiv \text{MBM}^T$ component-wise.
- Create initial matrix $M_0$ by setting $m_{ij} = 1$ if it is possible for vertex $i \in G_2$ to map to vertex $j \in G_1$, i.e., if indeg($i$) ≤ indeg($j$) and outdeg($i$) ≤ outdeg($j$).
- Recursively cycle through all possible permutation matrices $M$ generated from $M_0$ and check if each encodes an isomorphism.
- Idea can be extended to graphs with colored edges
- Represent edge colors in adjacency matrices by 1 and -1
- Then $M$ encodes an isomorphism iff $A \equiv \text{MBM}^T$, where $=$ represents the Hadamard product for component-wise multiplication

---

**Clustering E. coli**
- Network contains 423 nodes, 578 edges
- Each node represents DNA operon; each edge is directed from operon encoding transcription factor to operon regulated by said transcription factor
- Edges can be activating, repressing, or both
- Majority of feedforward motifs in the network are of coherent type, where direct effect has same sign as indirect effect

---

**Potential Applications**
- Gene networks (E. coli)
- Neuronal networks (C. elegans)
- Social networks
- Industrial networks
- Criminal networks
- Computer networks

**Goal**: Identify the “role” of each node in the network

---

**References**

---

**Acknowledgements**
- Davidson Research Initiative
- Dr. Timothy Chartier
- Students and faculty in summer research group