From Closing Triangles to Closing Higher-Order Motifs
Ryan A. Rossi¹ | Anup Rao¹, Sungchul Kim¹, Eunyee Koh¹, Nesreen K. Ahmed²

¹ Adobe Research
² Intel Labs
Higher-Order Motif Closures

- Proposed general notion of a motif closure that goes beyond simple triangle closures
- Introduce higher-order ranking and link prediction methods based on closing higher-order network motifs
- Demonstrate that these new motif closures often outperform triangle-based methods

**Definition 1 (Motif Closure).** A node pair \((i, j)\) is said to close a motif \(H\) iff adding an edge \((i, j)\) to \(E\) closes an instance \(F \in I_{G'}(H)\) of motif \(H\) where \(G' = (V, E \cup \{(i, j)\})\) and \(I_{G'}(H)\) is the set of unique instances of motif \(H\) in \(G'\).
Higher-Order Motif Closure Frequency

**Definition 2 (Higher-Order Motif Closure Frequency).** Let $G' = (V, E')$ where $E' = E \cup \{(i, j)\}$ and let $I_{G'}(H)$ be the set of unique instances of motif $H$ in $G'$. Then the frequency of closing a higher-order motif $H$ between node $i$ and $j$ is:

$$W_{ij} = \sum_{F \in I_{G'}(H)} \mathbb{1}(\{i, j\} \in E'(F))$$

where $W_{ij}$ is equal to the number of unique instances of $H$ that contain nodes $\{i, j\} \subset V(G')$ as an edge.
Algorithm 1 Higher-Order Motif Closures

Input: a graph $G = (V, E)$, node pair $(i, j)$, and network motif/graphlet $H$

Output: the frequency $W_{ij}$ of motif closures of $H$ for nodes $i$ and $j$

1. Set $E' \leftarrow E \cup \{(i, j)\}$ and $G' = (V, E')$
2. Use fast algorithm [1, 4] to compute $W_{ij} = \#$ of occurrences of motif $H$ between node $i$ and $j$ in $G'$
Experiments

The experiments investigate the following key questions:

- **Q1.** Do other motif closures perform better than triangle closure & its variants for some graphs?
- **Q2.** Does the “best” motif closure depend highly on the underlying network and its structural properties or is there one motif closure that always outperforms the others?

Experimental Setup

- Hold-out 10% of the observed node pairs uniformly at random
- Randomly sample the same number of negative node pairs
- Use the methods to obtain a ranking of the node pairs
- Repeat this 10 times and average the result
## Results

Mean average precision (MAP) results for ranking methods based on closing higher-order motifs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4-path</td>
<td>0.829</td>
<td>0.687</td>
<td>0.607</td>
<td>0.594</td>
<td>0.649</td>
<td>0.778</td>
<td>0.865</td>
<td>0.729</td>
<td>0.893</td>
<td>0.873</td>
<td>0.914</td>
<td>0.788</td>
</tr>
<tr>
<td>4-star</td>
<td>0.880</td>
<td>0.787</td>
<td>0.595</td>
<td>0.696</td>
<td>0.922</td>
<td>0.814</td>
<td>0.895</td>
<td>0.861</td>
<td>0.840</td>
<td>0.813</td>
<td>0.889</td>
<td>0.688</td>
</tr>
<tr>
<td>4-cycle</td>
<td>0.881</td>
<td>0.958</td>
<td>0.651</td>
<td>0.926</td>
<td>0.827</td>
<td>0.885</td>
<td>0.908</td>
<td>0.935</td>
<td>0.927</td>
<td>0.957</td>
<td>0.930</td>
<td>0.900</td>
</tr>
<tr>
<td>4-tailed-triangle</td>
<td>0.804</td>
<td>0.612</td>
<td>0.570</td>
<td>0.752</td>
<td>0.773</td>
<td>0.663</td>
<td>0.773</td>
<td>0.681</td>
<td>0.689</td>
<td>0.779</td>
<td>0.600</td>
<td>0.496</td>
</tr>
<tr>
<td>4-chordal-cycle</td>
<td>0.801</td>
<td>0.837</td>
<td>0.598</td>
<td>0.842</td>
<td>0.312</td>
<td>0.966</td>
<td>0.840</td>
<td>0.854</td>
<td>0.977</td>
<td>0.996</td>
<td>0.986</td>
<td>0.947</td>
</tr>
<tr>
<td>4-clique</td>
<td>0.804</td>
<td>0.838</td>
<td>0.595</td>
<td>0.843</td>
<td>0.293</td>
<td>0.963</td>
<td>0.842</td>
<td>0.847</td>
<td>0.972</td>
<td>0.997</td>
<td>0.986</td>
<td>0.965</td>
</tr>
<tr>
<td>CN</td>
<td>0.705</td>
<td>0.872</td>
<td>0.613</td>
<td>0.839</td>
<td>0.422</td>
<td>0.814</td>
<td>0.833</td>
<td>0.897</td>
<td>0.839</td>
<td>0.960</td>
<td>0.949</td>
<td>0.852</td>
</tr>
<tr>
<td>Jaccard Sim.</td>
<td>0.705</td>
<td>0.873</td>
<td>0.618</td>
<td>0.841</td>
<td>0.537</td>
<td>0.933</td>
<td>0.853</td>
<td>0.918</td>
<td>0.955</td>
<td>0.997</td>
<td>0.973</td>
<td>0.918</td>
</tr>
<tr>
<td>Adamic/Adar</td>
<td>0.705</td>
<td>0.883</td>
<td>0.621</td>
<td>0.842</td>
<td>0.549</td>
<td>0.940</td>
<td>0.856</td>
<td>0.920</td>
<td>0.959</td>
<td>0.997</td>
<td>0.976</td>
<td>0.919</td>
</tr>
</tbody>
</table>

**Result 1.** Higher-order motif closures can outperform triangle closure (common neighbors) and other methods based on it.
**Result 2.** The best motif closure depends highly on the structural characteristics of the graph and its domain (biological vs. social network) as shown in the above Table.
Results

- Average runtime in milliseconds to compute all \{3, 4\}-node motif closures for each node pair.
- The runtime includes the baselines since they require 3-node motifs.

**Result 3.** For any 4-node motif \(H\), counting the number of motif closures \(W_{ij}\) that would arise if an edge between \(i\) and \(j\) was added to \(G\) is fast taking less than a millisecond on average across all graphs.
Summary of Contributions

- Proposed General Notion of "Motif Closure"
  - Moves beyond simple triangle closures

- Motif closures are often more predictive than triangle closures

- Important Findings & Implications of Results
  - Need to consider other motif closures (besides triangle closures)
  - Best motif closure depends highly on the network structure and processes governing it
  - Existing supervised methods can benefit new motif closures (by leveraging full spectrum, most dense to least)
Thanks for listening!