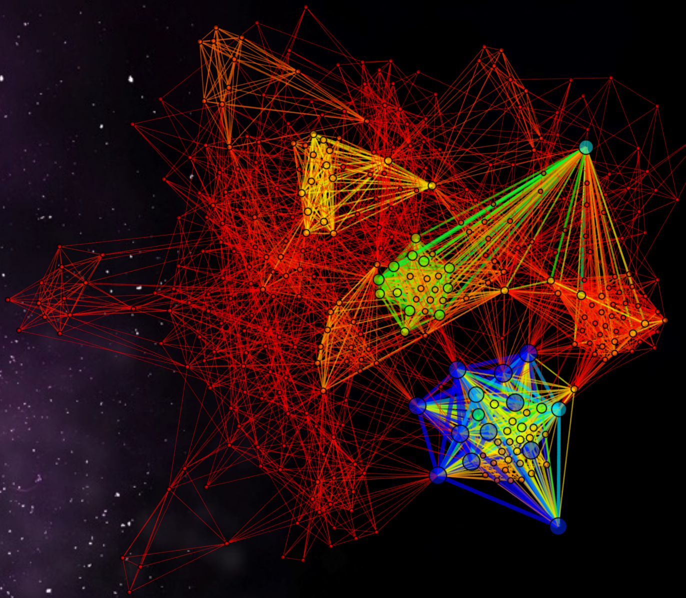
From Closing Triangles to Closing Higher-Order Motifs

Ryan A. Rossi[1] | Anup Rao[1], Sungchul Kim[1], Eunyee Koh[1], Nesreen K. Ahmed[2]

[1] Adobe Research

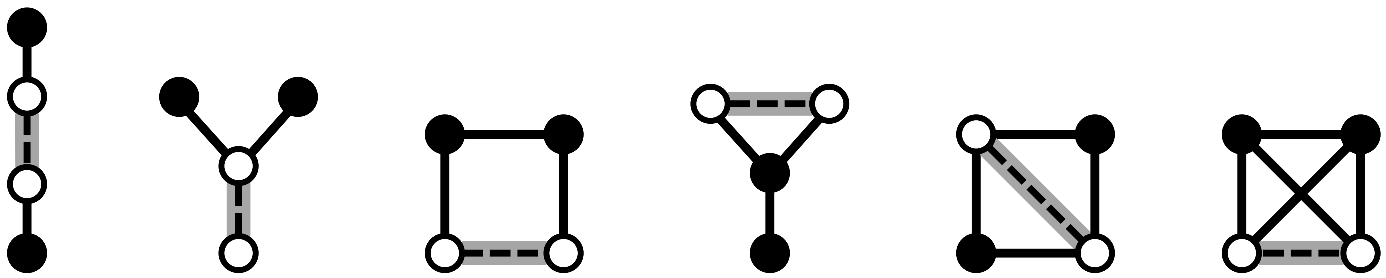
[2] Intel Labs



MAKE IT AN EXPERIENCE

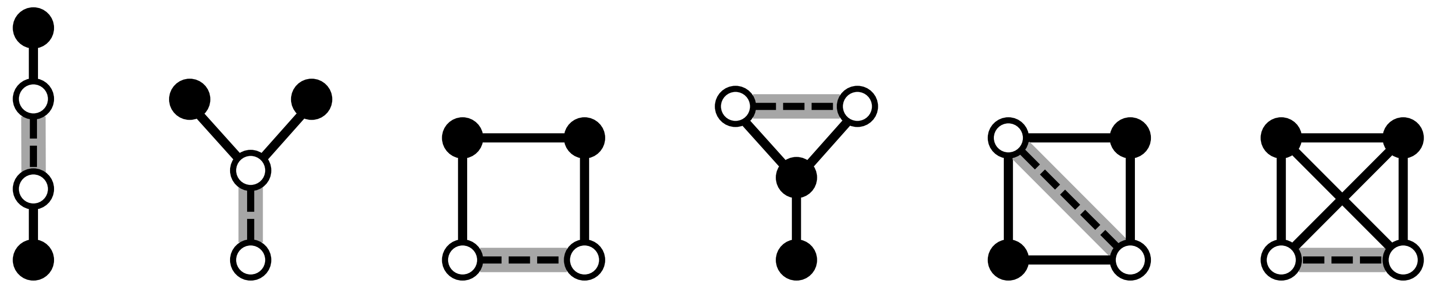
Higher-Order Motif Closures

- 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

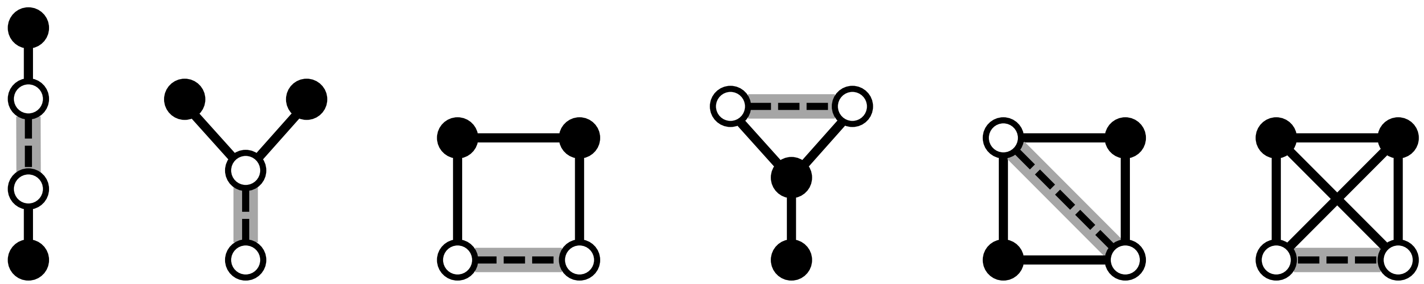


Definition 2 (Higher-Order Motif Closure Frequency). *Let $G' = (V, E')$ where $E' = E \cup \{(i, j)\}$ and let $I_{G'}(H)$ be the set of unique instances of motif H in G'. Then the frequency of closing a higher-order motif H between node i and j is:*

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

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

Higher-Order Motif Closure Frequency



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.

	bn-mouse	*bio-DM-LC*	*bio-CE-HT*	*bio-DM-HT*	*ia-reality*	*web-polblogs*	*biogrid-worm*	*biogrid-plant*	*biogrid-yeast*	*email-dnc-corec.*	*soc-advogato*	*econ-wm1*
4-path	0.829	0.687	0.607	0.594	0.649	0.778	0.865	0.729	0.893	0.873	0.914	0.788
4-star	0.880	0.787	0.595	0.696	**0.922**	0.814	0.895	0.861	0.840	0.813	0.889	0.688
4-cycle	**0.881**	**0.958**	**0.651**	**0.926**	0.827	0.885	**0.908**	**0.935**	0.927	0.957	0.930	0.900
4-tailed-triangle	0.804	0.612	0.570	0.752	0.773	0.663	0.773	0.681	0.689	0.779	0.600	0.496
4-chordal-cycle	0.801	0.837	0.598	0.842	0.312	**0.966**	0.840	0.854	**0.977**	0.996	**0.986**	0.947
4-clique	0.804	0.838	0.595	0.843	0.293	0.963	0.842	0.847	0.972	**0.997**	**0.986**	**0.965**
CN	0.705	0.872	0.613	0.839	0.422	0.814	0.833	0.897	0.839	0.960	0.949	0.852
Jaccard Sim.	0.705	0.873	0.618	0.841	0.537	0.933	0.853	0.918	0.955	**0.997**	0.973	0.918
Adamic/Adar	0.705	0.883	0.621	0.842	0.549	0.940	0.856	0.920	0.959	**0.997**	0.976	0.919

Result 1. Higher-order motif closures can outperform triangle closure (common neighbors) and other methods based on it

Results

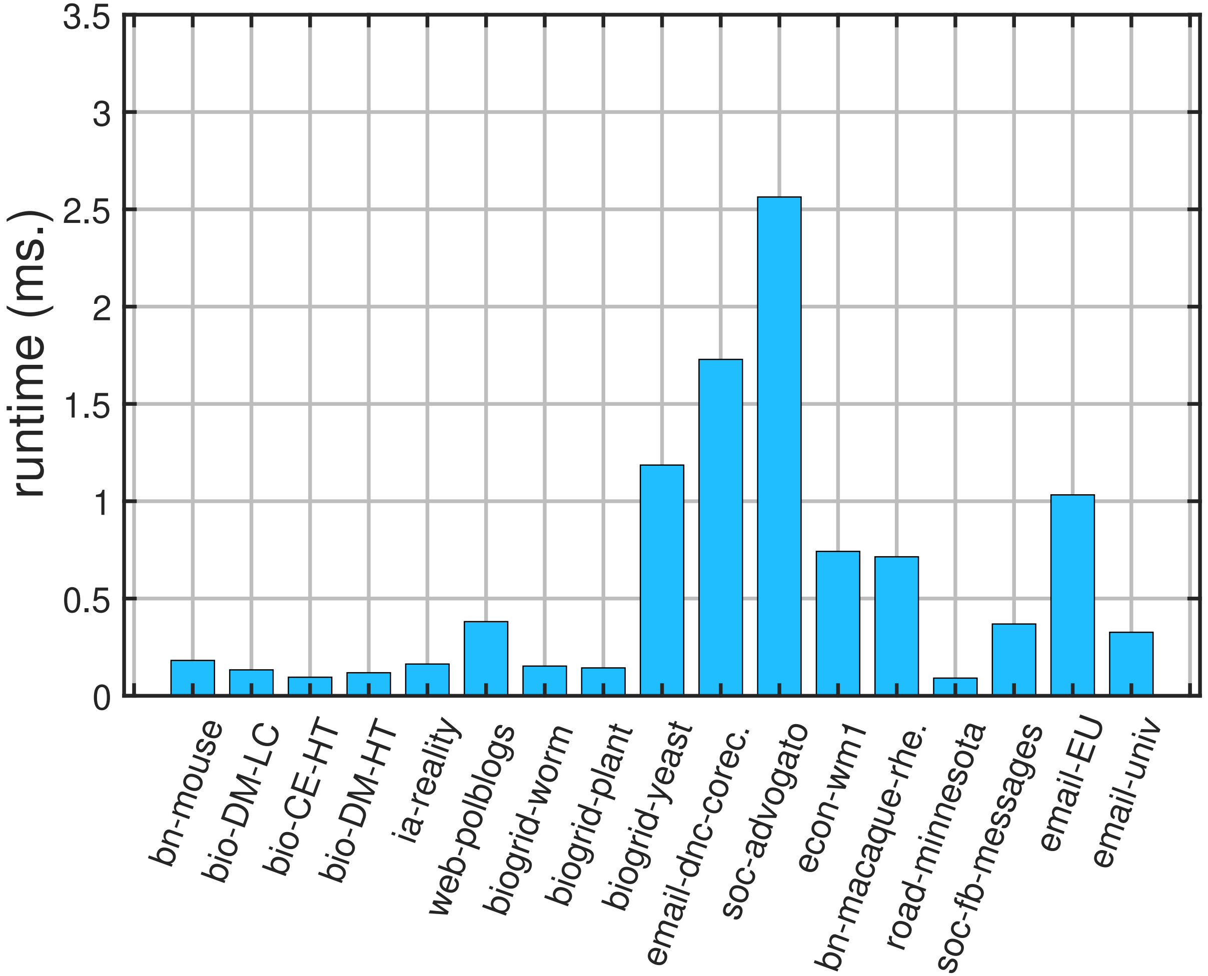
Mean average precision (MAP) results for ranking methods based on closing higher-order motifs.

	bn-mouse	*bio-DM-LC*	*bio-CE-HT*	*bio-DM-HT*	*ia-reality*	*web-polblogs*	*biogrid-worm*	*biogrid-plant*	*biogrid-yeast*	*email-dnc-corec.*	*soc-advogato*	*econ-wm1*
4-path	0.829	0.687	0.607	0.594	0.649	0.778	0.865	0.729	0.893	0.873	0.914	0.788
4-star	0.880	0.787	0.595	0.696	**0.922**	0.814	0.895	0.861	0.840	0.813	0.889	0.688
4-cycle	**0.881**	**0.958**	**0.651**	**0.926**	0.827	0.885	**0.908**	**0.935**	0.927	0.957	0.930	0.900
4-tailed-triangle	0.804	0.612	0.570	0.752	0.773	0.663	0.773	0.681	0.689	0.779	0.600	0.496
4-chordal-cycle	0.801	0.837	0.598	0.842	0.312	**0.966**	0.840	0.854	**0.977**	0.996	**0.986**	0.947
4-clique	0.804	0.838	0.595	0.843	0.293	0.963	0.842	0.847	0.972	**0.997**	**0.986**	**0.965**
CN	0.705	0.872	0.613	0.839	0.422	0.814	0.833	0.897	0.839	0.960	0.949	0.852
Jaccard Sim.	0.705	0.873	0.618	0.841	0.537	0.933	0.853	0.918	0.955	**0.997**	0.973	0.918
Adamic/Adar	0.705	0.883	0.621	0.842	0.549	0.940	0.856	0.920	0.959	**0.997**	0.976	0.919

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!

Appendix