

From Closing Triangles to Closing Higher-Order Motifs

Ryan A. Rossi
Adobe Research

Anup Rao
Adobe Research

Sungchul Kim
Adobe Research

Eunyee Koh
Adobe Research

Nesreen Ahmed
Intel Labs

ABSTRACT

This work introduces higher-order ranking and link prediction methods based on *closing higher-order network motifs*. In particular, we propose the general notion of a *motif closure* that goes beyond simple triangle closures and demonstrate that these new motif closures often outperform triangle-based methods. This result implies that one should consider other motif closures beyond simple triangles. We also find that the “best” motif closure depends highly on the underlying network and its structural properties. Furthermore, the methods are fast and efficient for *real-time* applications such as online visitor stitching, web search, and recommendation. The experimental results indicate the importance of these new motif closures. Finally, the new motif closures can serve as a basis for developing better (un)supervised ranking/link prediction methods.

1 HIGHER-ORDER MOTIF CLOSURES

We first formally define the notion of a higher-order motif closure:

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' .

A few examples of *higher-order motif closures* are shown in Figure 1. The edge (i, j) shown as a dotted line in Figure 1 closes each motif. For instance, the edge between node i and j in the rightmost motif in Figure 1 closes a 4-clique. We now formally introduce the frequency of higher-order motif closures for a node pair (i, j) as follows:

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:

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

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

We provide a simple routine in Alg. 1 for computing the weight W_{ij} representing the frequency of closing motif H between node i and j . The approach has two simple steps. First, given an arbitrary node pair (i, j) , a motif H of interest, and the current graph $G = (V, E)$, we simply add the node pair (i, j) as an edge by setting $E' \leftarrow E \cup \{(i, j)\}$ and $G' = (V, E')$ (Alg. 1 Line 1). This can

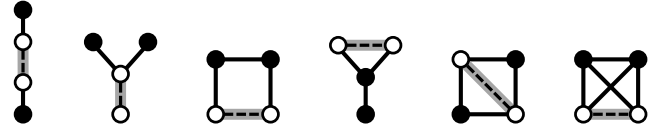


Figure 1: Motif Closures. The unshaded/white nodes are node i and j . Given a node pair $(i, j) \notin E$ (unshaded/white nodes) and any motif/induced subgraph H , the “edge” between i and j (dotted gray line) is said to close an instance F of H if the edge (i, j) were to actually exist in G .

be performed *implicitly* without any additional work. However, it is shown in Alg. 1 since after adding (i, j) to the edge set, we can use the fastest known algorithm [1] for counting the occurrences of motif (induced subgraph/graphlet) H between node i and j in G' . Second, we compute the number of instances of motif H that contain nodes i and j in G' (Alg. 1 Line 2). Given a set $\mathcal{Y} = \{y_1, y_2, \dots, y_j, \dots\}$ of nodes (items, ads, songs, friends) to be ranked, Alg. 1 computes $W_{ij} = f(x_i, y_j), \forall j = 1, \dots, |\mathcal{Y}|$.

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'
-

2 EXPERIMENTS

The experiments investigate the following key questions:

- Q1 Do other motif closures perform better than triangle closure and 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?

For comparison, we use triangle closure (common neighbors) and its variants since these have the same desired properties as the higher-order motif closure methods described in this paper. We hold-out 10% of the observed node pairs uniformly at random and randomly sample the same number of negative node pairs. We repeat this 10 times and average the results. We then use the methods to obtain a ranking of the node pairs in this set. Recall the proposed techniques do not require learning a sophisticated model nor do they require training data. As such, the notion of motif closure proposed in this work can be used in a real-time streaming fashion and has many other advantages to more sophisticated model-based approaches. Mean Average Precision (MAP) results are provided in Table 1.

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

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7024-0/20/04.

<https://doi.org/10.1145/3366424.3382688>

Table 1: Mean average precision (MAP) results for ranking (and prediction) methods based on closing higher-order motifs.

	<i>bn-mouse</i>	<i>bio-DM-LC</i>	<i>bio-CE-HT</i>	<i>bio-DM-HT</i>	<i>ia-reality</i>	<i>web-polblogs</i>	<i>biogrid-worm</i>	<i>biogrid-plant</i>	<i>biogrid-yeast</i>	<i>email-dnc-corec.</i>	<i>soc-advogato</i>	<i>econ-wm1</i>	<i>bn-macaque-rhe.</i>	<i>road-minnesota</i>	<i>soc-fb-messages</i>	<i>email-EU</i>	<i>email-univ</i>
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	0.942	0.326	0.844	0.854	0.707
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	0.961	0.388	0.807	0.972	0.695
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	0.773	0.950	0.870	0.902	0.847
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	0.530	0.834	0.722	0.937	0.582
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	0.750	0.939	0.935	0.782	0.969
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	0.759	0.939	0.960	0.798	0.982
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	0.342	0.945	0.790	0.890	0.941
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	0.764	0.944	0.841	0.933	0.949
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	0.777	0.945	0.848	0.935	0.953

In nearly all cases, the higher-order motif closures achieve better precision than techniques based on closing lower-order triangles. This result has a number of important implications. First, it implies that one should also consider other motif closures that go beyond simple triangles. Second, it brings new opportunities for research on different and more useful variants based on the new motif closures.

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 Table 1.*

This implies that no single motif closure will always perform best for all graphs and is consistent with the no-free-lunch theorem [5]. Furthermore, different structural properties are known to be important for different network domains, and these higher-order motifs capture the important structural properties since they cover the full spectrum of 4-node induced subgraphs. In Table 1, biological and brain networks achieve best performance using the ranking given by 4-cycle and 4-star closures. This also holds true for the interaction (*ia-reality*) and road network investigated. The 4-star and 4-cycle motif closures are more sparse compared to the 4-chordal-cycle (paw motif) and 4-clique motif closure. In the web graph, economic, and social networks, both the 4-chordal-cycle (diamond motif closure) and 4-clique motif closure achieves significantly better performance than the other motif closures. Notice that both these motif closures are composed of two or more triangles and thus can be seen as a stronger triadic closure motif.

Runtime Performance: We report the average runtime in milliseconds to compute all motif closures for each node pair in G . The methods were implemented in python and all experiments were performed on a laptop (MacBook Pro 2017, 3.1 GHz Intel Core i7, 16GB RAM). For most graphs, it takes *less than a millisecond* on average as shown in Figure 2 and therefore is fast for large-scale ranking problems. Note these results include the runtime to compute 3-node motif closures as well (and thus includes methods such as common neighbors), since the algorithm used to count them leverages 3-node motifs to derive the 4-node motifs efficiently [1].

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.*

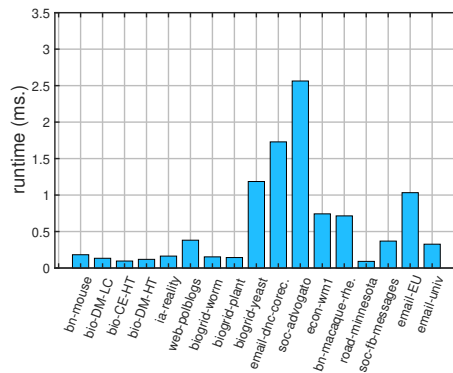


Figure 2: Average runtime in milliseconds to compute all {3, 4}-node motif closures for each node pair. The runtime includes the baselines since they require 3-node motifs.

3 CONCLUSION

In this paper, we proposed the general notion of a motif closure that goes beyond simple triangle closures. Indeed, triangle closure has been used as a basis for link prediction over the last decade. In this work, we demonstrated that other motif closures are often more predictive than their triangle-based counterparts. This result has three important implications. First, it implies that one should also consider motif closures that are different from triangles. Second, the “best” motif closure (*i.e.*, the motif closure that is most predictive of a link) depends highly on the underlying network structure and processes that govern it. Third, existing supervised learning methods [2, 3] can benefit from these new motif closures by leveraging the full range of motif closures (going from least to most dense as shown in Figure 1).

REFERENCES

- [1] Nesreen K. Ahmed, Jennifer Neville, Ryan A. Rossi, and Nick Duffield. 2015. Efficient Graphlet Counting for Large Networks. In *ICDM*. 10.
- [2] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. 2006. Link prediction using supervised learning. In *SDM Workshop*.
- [3] Ryan N Lichtenwalter, Jake T Lussier, and Nitesh V Chawla. 2010. New perspectives and methods in link prediction. In *KDD*. 243–252.
- [4] Ryan A. Rossi, Rong Zhou, and Nesreen K. Ahmed. 2018. Estimation of Graphlet Counts in Massive Networks. In *TNNLS*. 44–57.
- [5] David H Wolpert, William G Macready, et al. 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 1 (1997), 67–82.