Hypergraph Motifs

2025-08-07

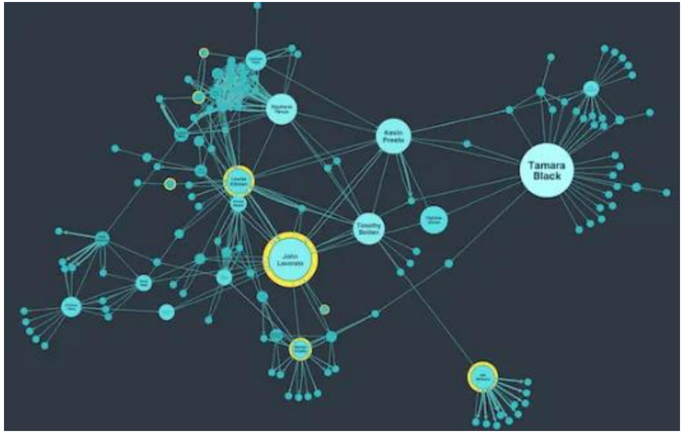
InHyeok Jeong
School of Computer Science and Engineering
Chung-Ang University
DMAIS Lab

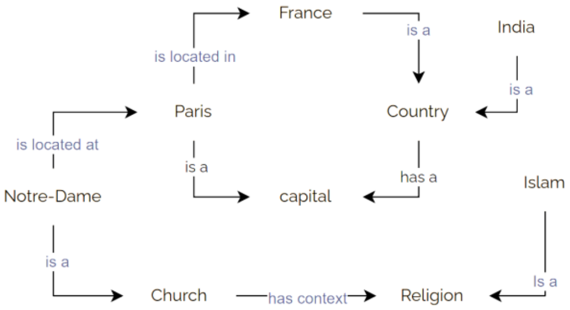
Content

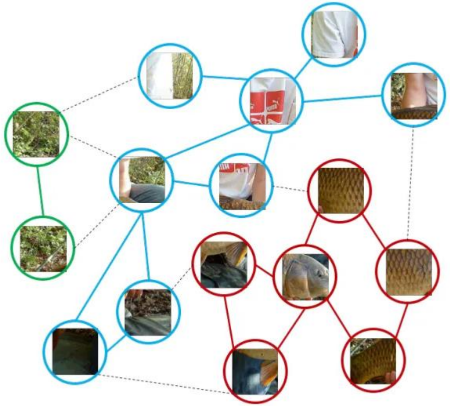
- Introduction
  - What graphs can represent
  - Analyzing complex networks
- Network Motifs
- Hypergraph Motif
  - H-motifs
  - TH-motifs
- Conclusion

What graphs can represent

- Graphs can represent everything
  - $e.g.$, Social networks, Text as a KG, Image





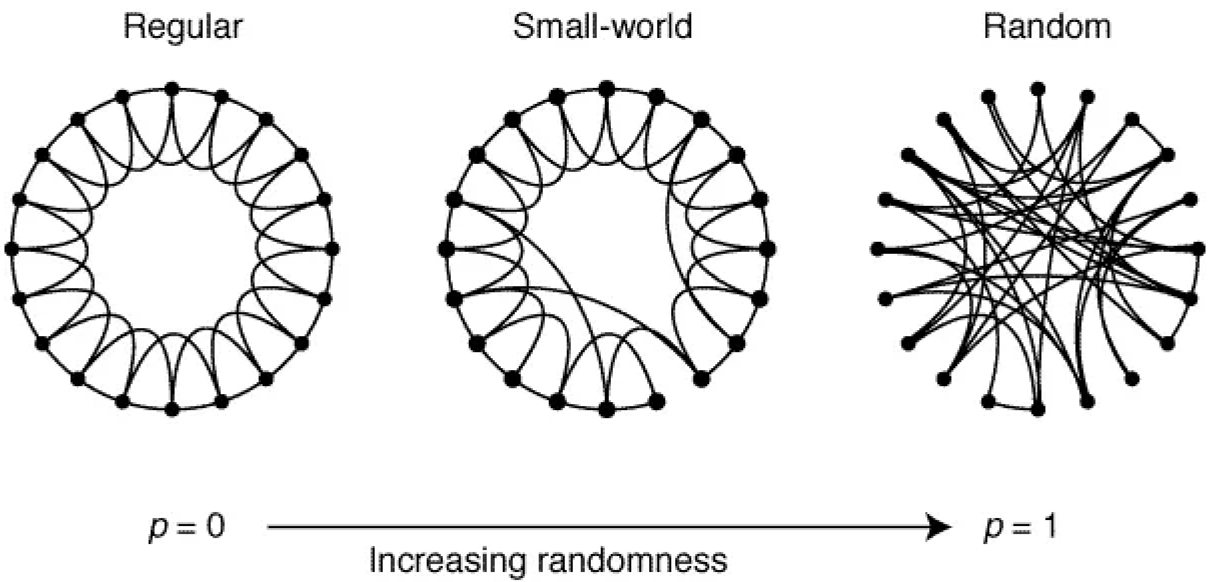


(b) Graph structure.

Han, Kai, et al. "Vision gnn: An image is worth graph of nodes." *Advances in neural information processing systems* 35 (2022): 8291-8303.
Wold, Sondre, Lilja Øvrelid, and Erik Velldal. "Text-to-KG alignment: Comparing current methods on classification tasks." *arXiv preprint arXiv:2306.02871* (2023).

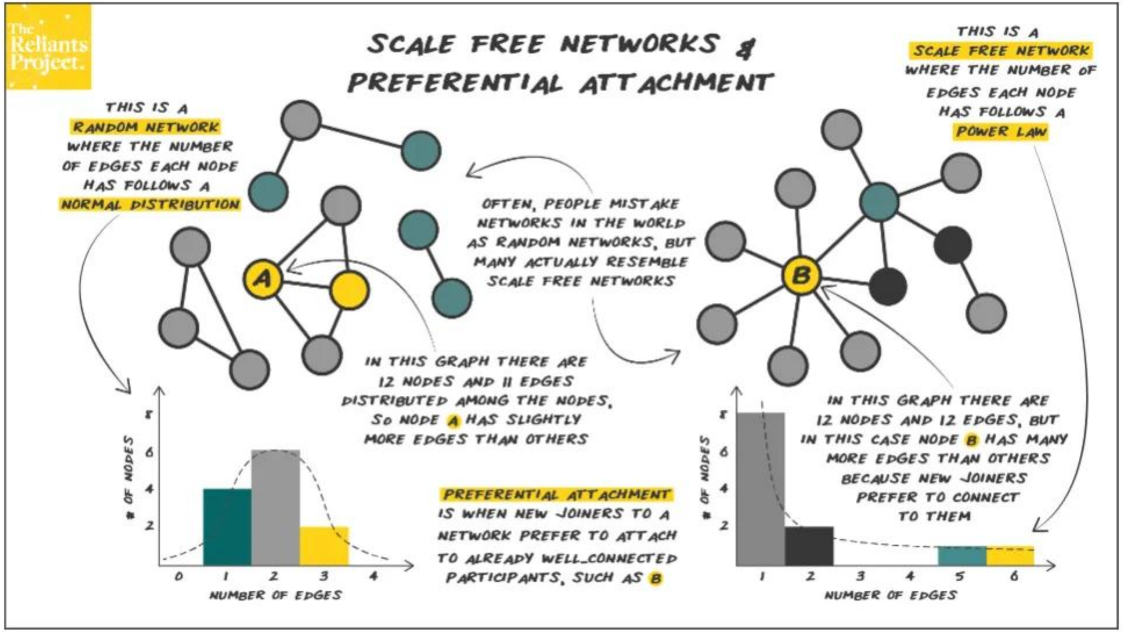
Analyzing complex networks

- ‘Overall’ complex networks
  - Small world networks



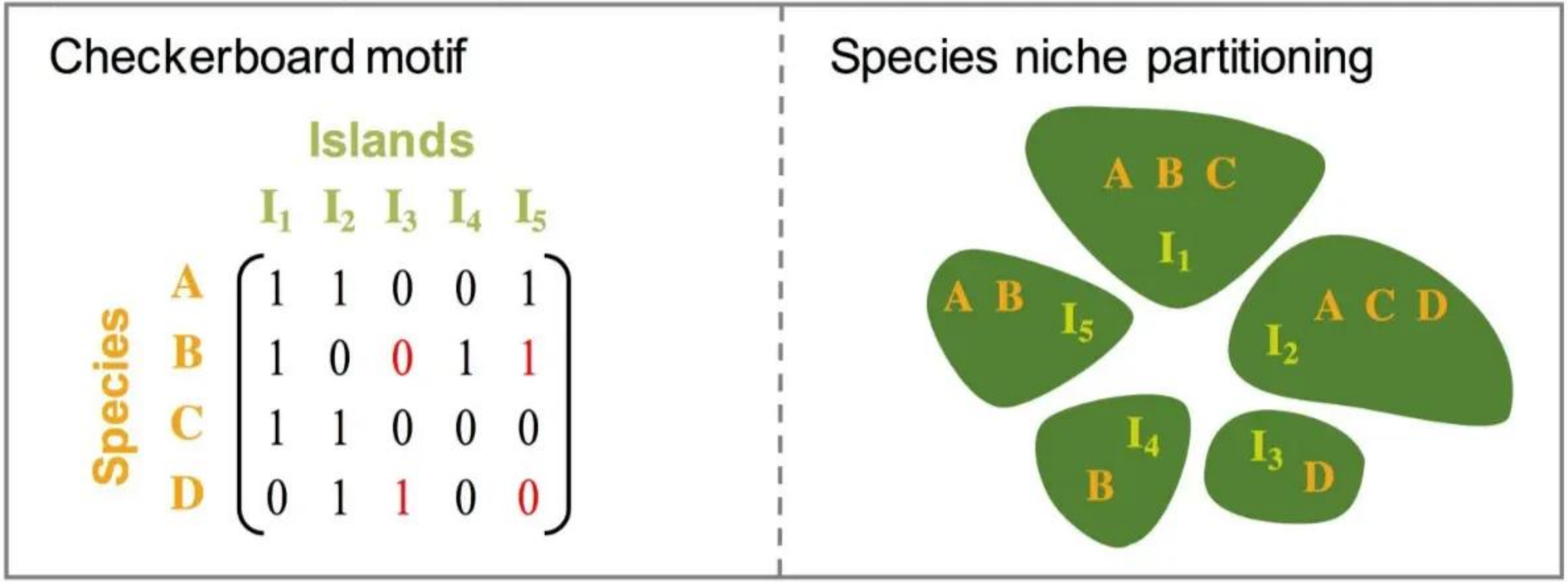
Analyzing complex networks

- ‘Overall’ complex networks
  - Scale-free networks



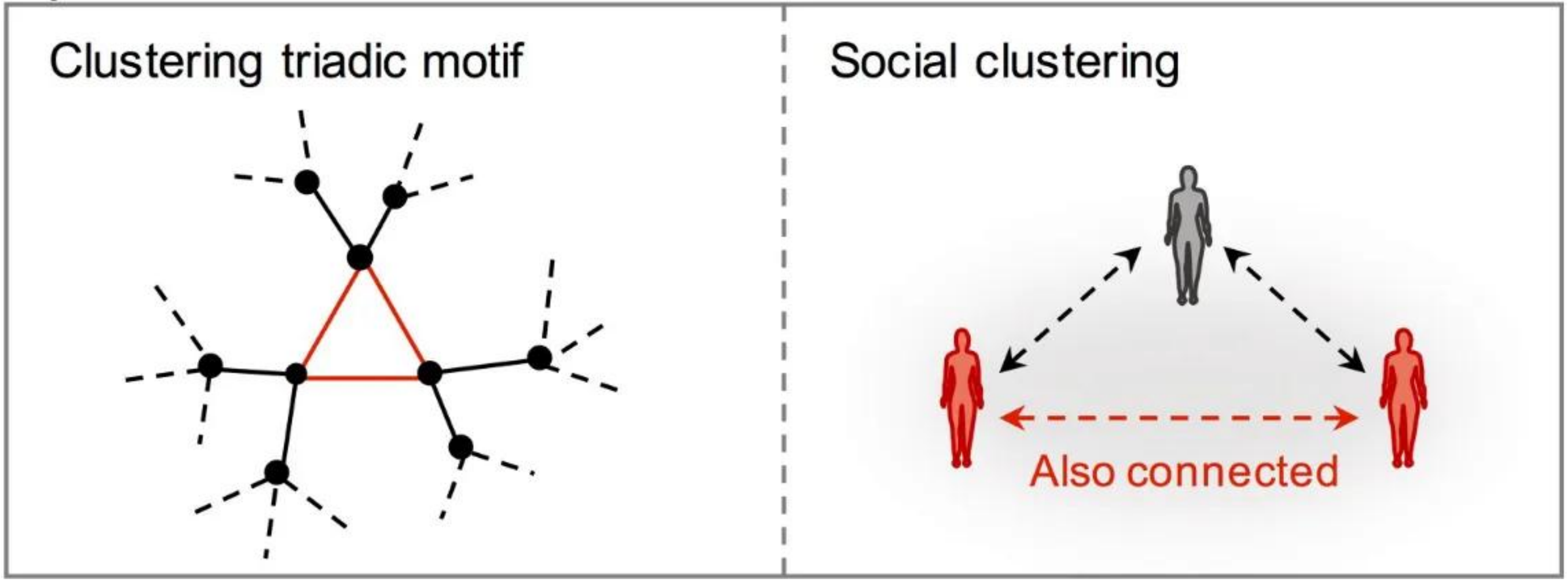
Analyzing complex networks

- ☐ **'Specific' complex networks**
  - Case Study : Ecological networks



Analyzing complex networks

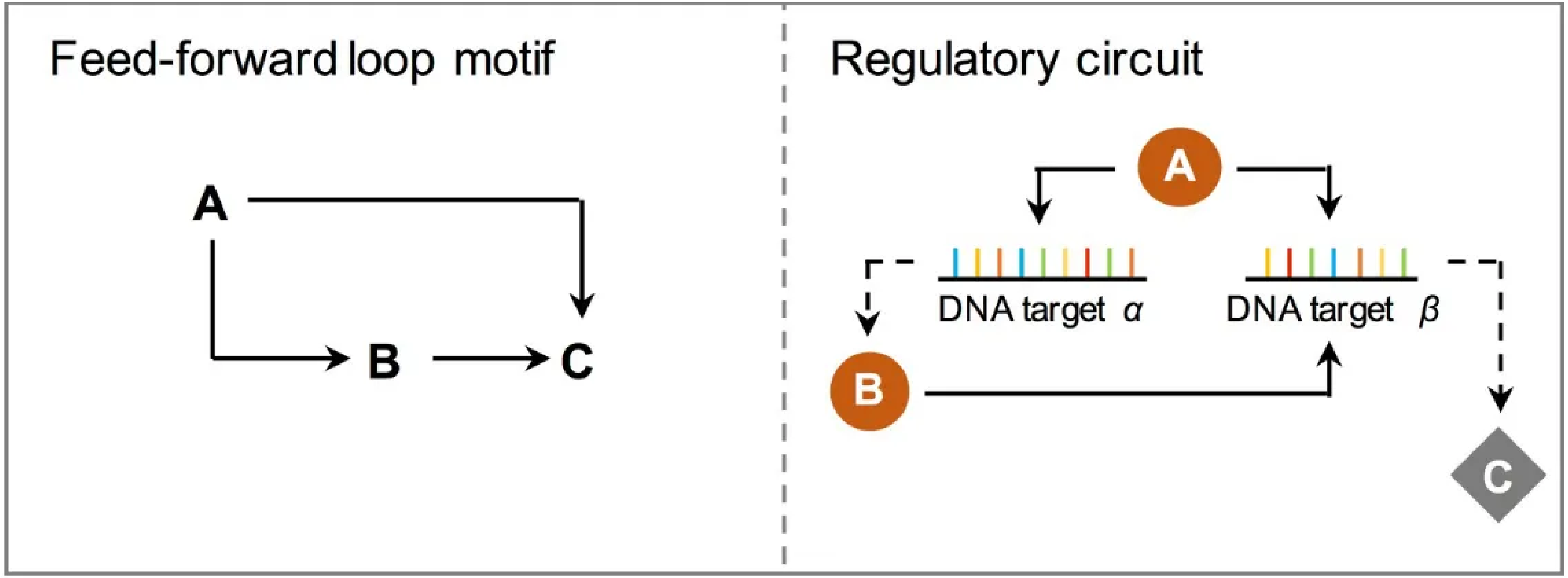
- ‘Specific’ complex networks
  - Case Study : Social networks



Stone, Lewi, Daniel Simberloff, and Yael Artzy-Randrup. "Network motifs and their origins." *PLoS computational biology* 15.4 (2019): e1006749.

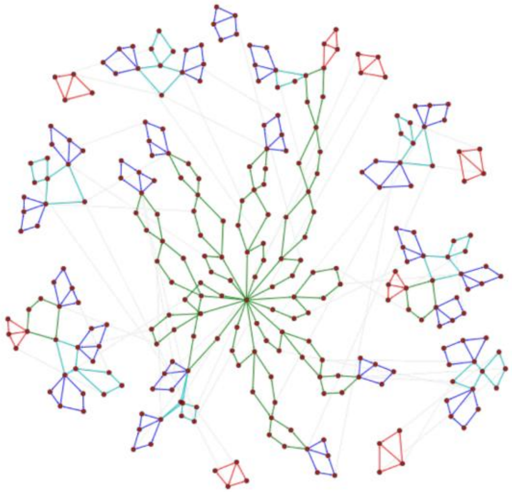
Analyzing complex networks

- ‘Specific’ complex networks
  - Case Study : Gene transcription networks



Analyzing complex networks

- ‘Specific’ complex network
  - The local structure pattern appears repeatedly on a specific complex network



Subgraph decomposition of an electronic circuit

→ Apply it to network theory!

Network Motifs

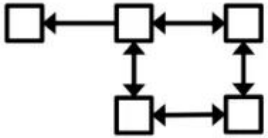
- **Recurring, significant patterns of interconnections**
  - Recurring : Found many times, i.e., with high frequency
  - Significant : More frequent than expected, i.e., in randomly generated networks
  - Pattern : Small Induced subgraph

Network Motifs

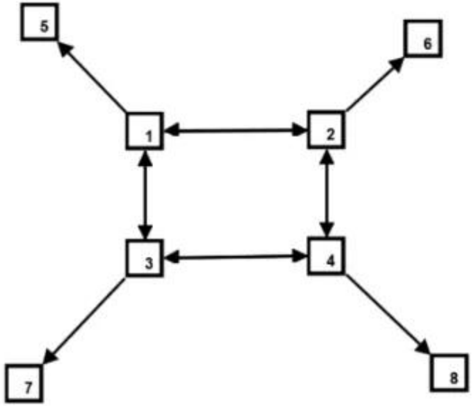
☐ **Recurring**

- Found many times, i.e., with high frequency

Motif of interest:



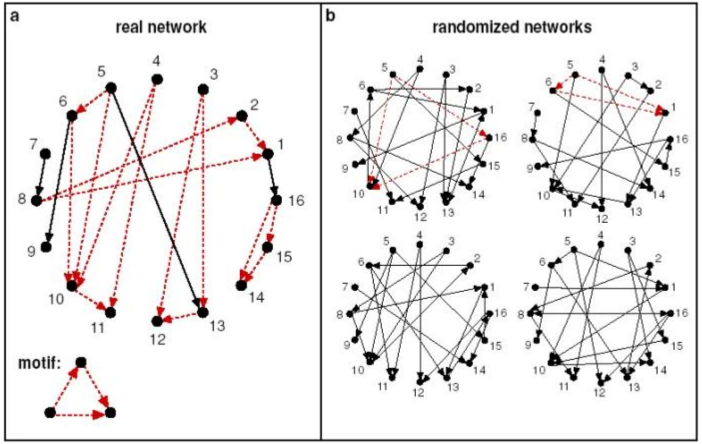
- Allow **overlapping of motifs**
- Network on the right has 4 occurrences of the motif:
  - {1,2,3,4,5}
  - {1,2,3,4,6}
  - {1,2,3,4,7}
  - {1,2,3,4,8}



Network Motifs

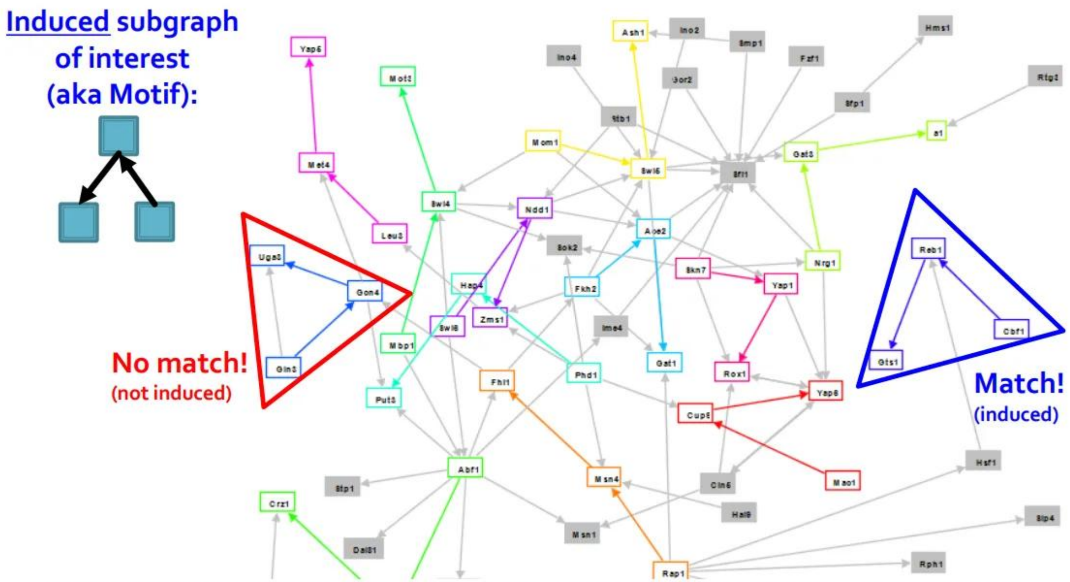
- **Significant**
  - More frequent than expected, i.e., in randomly generated networks

    - **Key idea: Subgraphs that occur in a real network much more often than in a random network have functional significance**



Network Motifs

- **Pattern**
  - Small Induced subgraph

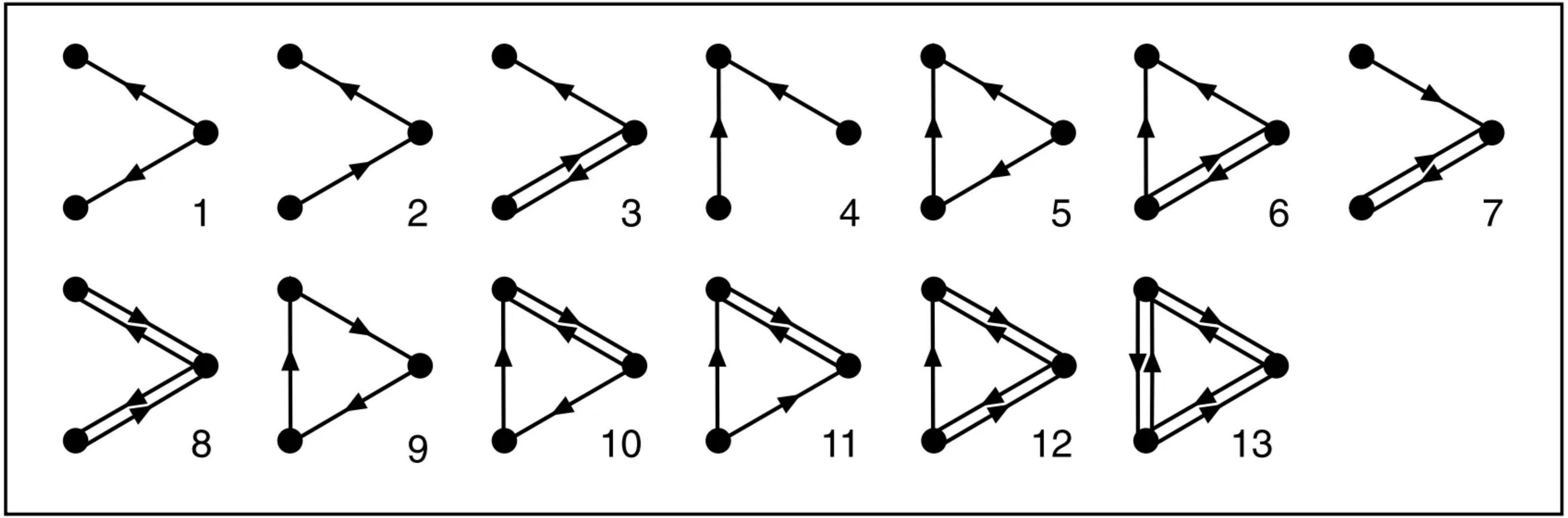


Induced subgraph of graph G is a graph, formed from a subset X of the vertices of graph G and all of the edges connecting pairs of vertices in subset X.

Network Motifs

☐ Generalization to network motifs

- Scalable to multiple nodes, but trade-off between expressiveness and complexity arises
- Network motifs using more than 3 nodes can represent a combination of ones using 3 nodes

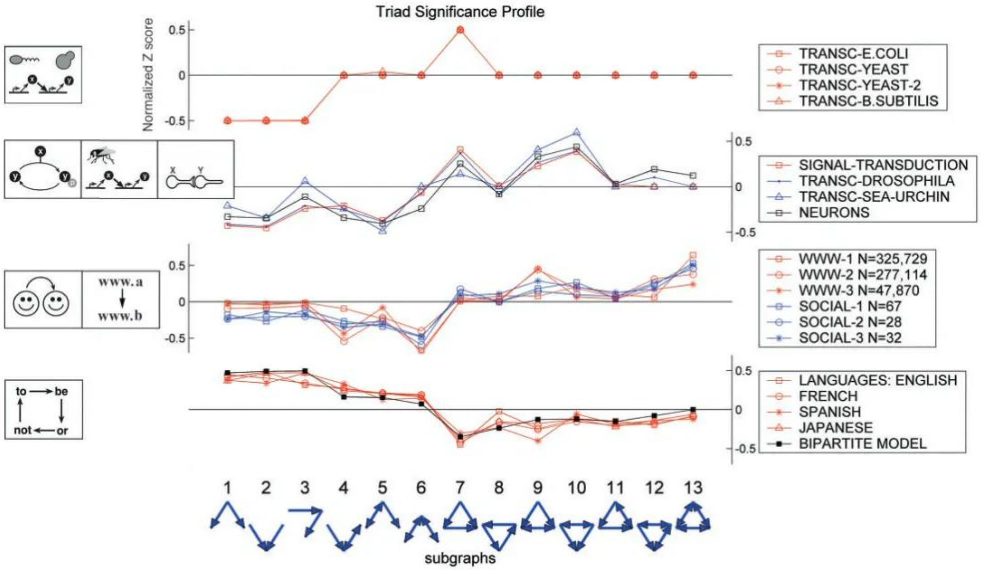


Milo, Ron, et al. "Network motifs: simple building blocks of complex networks." *Science* 298.5594 (2002): 824-827.

Why are network motifs important?

☐ A statistical perspective

- Help us understand how networks work
- Help us predict operation and reaction of the network in a given situation

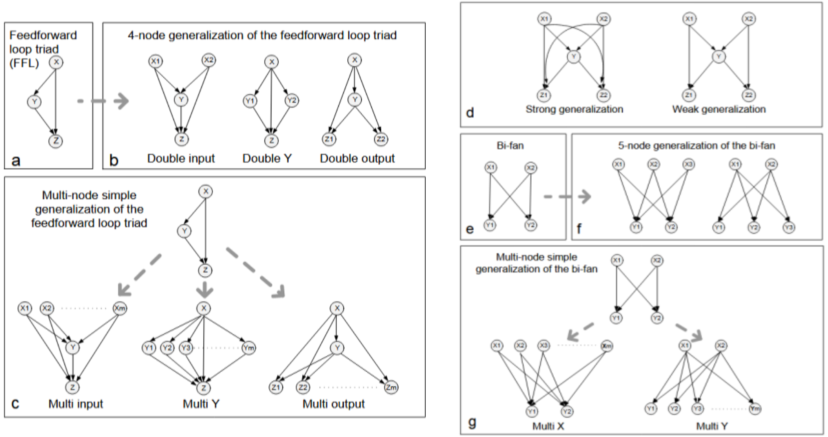


Network	Nodes	Edges	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

Why are network motifs important?

☐ A topological perspective

- Network motifs combine to form larger structures
- Networks that share the same motif can exhibit different generalizations of that motif



Generalization	Subgraph size	Transcriptional Networks		Neurons	Electronic chips
		E. coli	yeast	C. elegans	S15850
basic bi-fan	4 (2X,2Y)	+ (N=209)	+ (N=1812)	+ (N=126)	+ (N=1040)
multi output	5 (2X,3Y)	+ (N=264)	+ (N=14857)	+ (N=152)	+ (N=1990)
	6 (2X,4Y)	+ (C=0.015)	+ (C=3.5)	+(C=0.17)	+ (C=0.28)
multi input	5 (3X,2Y)	+ (N=20)	+ (N=81)	+ (N=25)	+ (N=226)
	6 (4X,2Y)	- (N=0)	+ (N=14)	+(C=0.015)	+ (C=0.001)
equal multi input-outputs	6 (3X,3Y)	+ (N=6)	+ (N=21)	- (N=0)	+ (N=301)

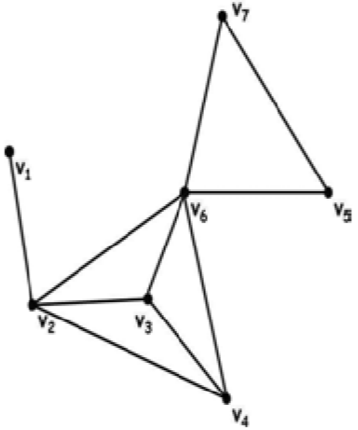
TABLE I: **Bi-fan generalizations in different networks.** (aX,bY) represents the multiplicity of each of the roles in the generalization (Fig. 2g). '+': Statistically significant generalizations, '-': non-significant generalizations. Number of appearances (N), or concentration ($x10^{-3}$) (C) [27] are listed.

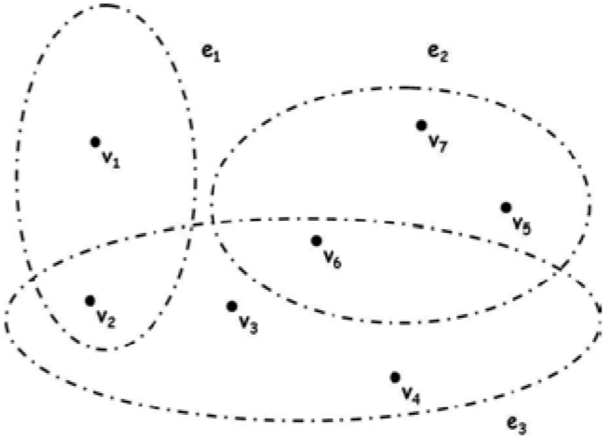
Kashtan, Nadav, et al. "Topological generalizations of network motifs." *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 70.3 (2004): 031909.

Limitation of Network Motifs

- **Interactions in many complex systems are groupwise rather than pairwise**
  - → This inherent limitation of graphs is addressed by hypergraphs

	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	1
v_3	0	0	1
v_4	0	0	1
v_5	0	1	0
v_6	0	1	1
v_7	0	1	0
v_2	1	0	1

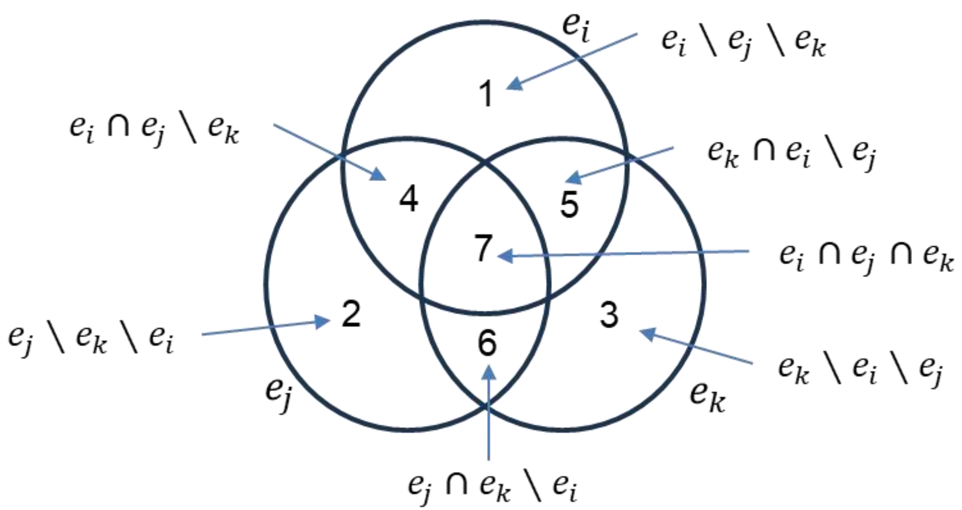




Zhou, Dengyong, Jiayuan Huang, and Bernhard Schölkopf. "Learning with hypergraphs: Clustering, classification, and embedding." *Advances in neural information processing systems* 19 (2006).

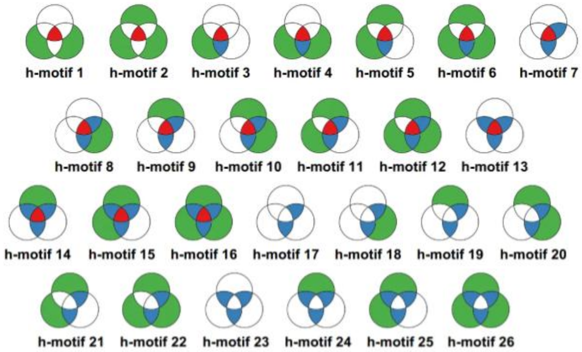
H-motifs

- **Representation**
  - A binary vector of size 7 whose elements represent the emptiness of each set
  - Describe the connectivity patterns of three connected hyperedges



H-motifs

- Representation
  - No symmetric patterns
    - Each connectivity pattern should be described by exactly one h-motif
  - No disconnected hyperedges
    - All k hyperedges should be connected
  - No duplicated hyperedges
    - Hyperedges containing the same set of nodes are not allowed



$$2^{2^k-1} - |P_1^{(k)} \cup P_2^{(k)} \cup P_3^{(k)}|$$

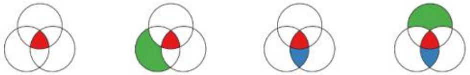


Figure 4: The h-motifs whose instances contain duplicated hyperedges.

→ *Remain 26 motifs!*

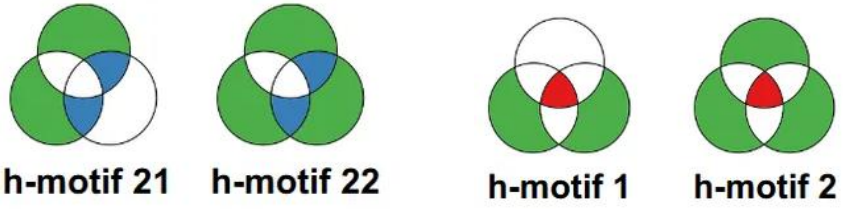
H-motifs

Properties

- **Exhaustive**
  - Capture connectivity patterns of *all possible* three connected hyperedges
- **Unique**
  - Connectivity pattern of any three connected hyperedges is captured by *at most one* h-motif
- **Size Independent**
  - Capture connectivity patterns *independently of the sizes of hyperedges*

H-motifs

- **Two types**
  - Open h-motifs
    - Instances contain two non-adjacent ($i.e.$, disjoint) hyperedges
  - Closed h-motifs
    - All three hyperedges in instances are adjacent to ($i.e.$, overlapped with) each other

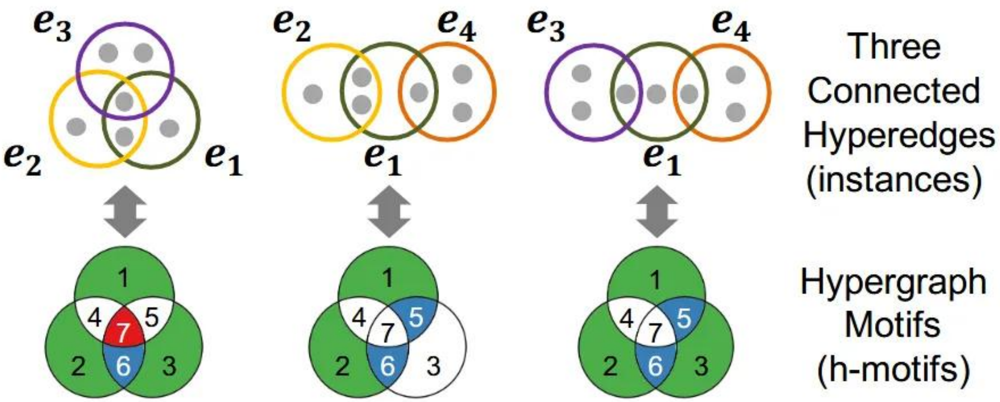


Open h-motif

Closed h-motif

H-motifs

- **Instances**
  - A set of three connected hyperedges is an *instance* of h-motif t
    - If their connectivity pattern corresponds to h-motif t
  - The count of each h-motif's instances is used to characterize the local structure



(d) H-motifs and instances

Lee, Geon, Jihoon Ko, and Kijung Shin. "Hypergraph Motifs: Concepts, Algorithms, and Discoveries." *Proceedings of the VLDB Endowment* 13.11.

Significance of h-motifs

- Compare the count of instances against the count of them in randomized hypergraphs

$$\Delta_t := \frac{M[t] - M_{rand}[t]}{M[t] + M_{rand}[t] + \epsilon}$$

Characteristic Profile (CP)

- ☐ **Normalize and concatenate the significances of all h-motifs in a hypergraph**
- ☐ **Summarize the local structural pattern of the hypergraph**

$$CP_t := \frac{\Delta_t}{\sqrt{\sum_{t=1}^{26} \Delta_t^2}}$$

Exact H-motif Counting

MoCHy-E

- Count the instances of each h-motif exactly
- Guarantee that each instance is counted exactly once

Algorithm 2: MoCHy-E: Exact H-motif Counting

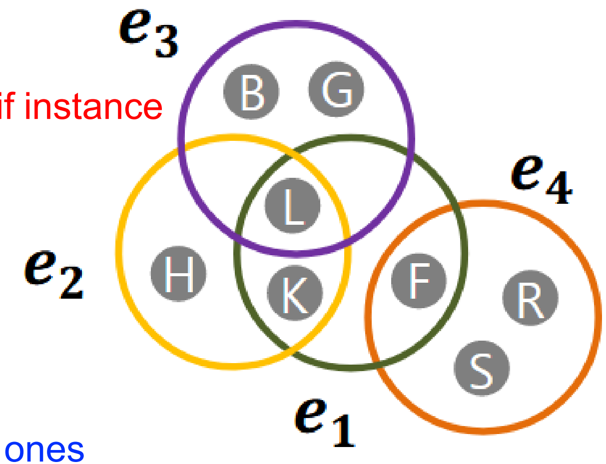
Input : (1) input hypergraph: $G = (V, E)$
(2) projected graph: $\bar{G} = (E, \wedge, \omega)$

Output: exact count of each h-motif t's instances: $M[t]$

```
M ← map whose default value is 0
for each hyperedge e_i ∈ E do
    for each unordered hyperedge pair {e_j, e_k} ∈ (N_{e_i} choose 2) do
        if e_j ∩ e_k = ∅ or i < min(j, k) then
            M[h({e_i, e_j, e_k})] += 1
return M
```

Determine a h-motif instance

Remove duplicated ones



Lee, Geon, Jihoon Ko, and Kijung Shin. "Hypergraph Motifs: Concepts, Algorithms, and Discoveries." *Proceedings of the VLDB Endowment* 13.11.

Approximate H-motif Counting

☐ MoCHy-A: Hyperedge Sampling

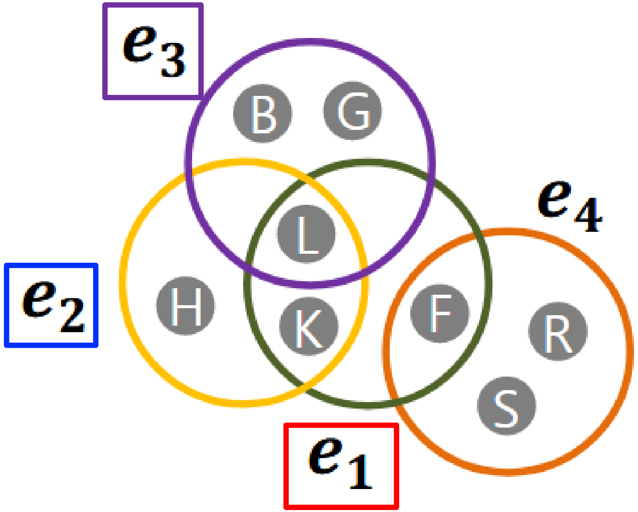
- Repeatedly samples s hyperedges from the hyperedge set E uniformly at random with replacement

Algorithm 4: MoCHy-A: Approximate H-motif Counting Based on Hyperedge Sampling

Input : (1) input hypergraph: $G = (V, E)$
(2) projected graph: $\bar{G} = (E, \wedge, \omega)$
(3) number of samples: s

Output: estimated count of each h-motif t's instances: $\bar{M}[t]$

$\bar{M}[t] \leftarrow$ map whose default value is 0
for $n \leftarrow 1 \ldots s$ **do**
    $e_i \leftarrow$ sample a uniformly random hyperedge
    for *each hyperedge* $e_j \in N_{e_i}$ **do**
        for *each hyperedge* $e_k \in (N_{e_i} \cup N_{e_j} \setminus \{e_i, e_j\})$ **do**
            if $e_k \notin N_{e_i}$ *or* $j < k$ **then**
                $\bar{M}[h(\{e_i, e_j, e_k\})] \mathrel{+}= 1$
for *each h-motif* t **do**
    $\bar{M}[t] \leftarrow \bar{M}[t] \cdot \frac{|E|}{3s}$
return $\bar{M}$



Approximate H-motif Counting

☐ MoCHy-A+: Hyperwedge Sampling

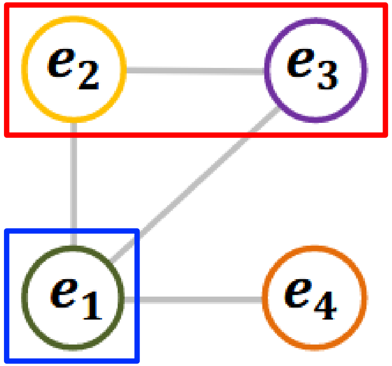
- Select r hyperwedges uniformly at random with replacement

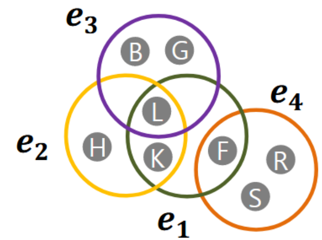
Algorithm 5: MoCHy-A$^+$: Approximate H-motif Counting Based on Hyperwedge Sampling

Input : (1) input hypergraph: $G = (V, E)$
(2) projected graph: $\bar{G} = (E, \wedge, \omega)$
(3) number of samples: r

Output: estimated count of each h-motif t's instances: $\hat{M}[t]$

$\hat{M} \leftarrow$ map whose default value is 0
for $n \leftarrow 1 \ldots r$ **do**
  $\wedge_{ij} \leftarrow$ a uniformly random hyperwedge
  for *each hyperedge* $e_k \in (N_{e_i} \cup N_{e_j} \setminus \{e_i, e_j\})$ **do**
    $\hat{M}[h(\{e_i, e_j, e_k\})] \mathrel{+}= 1$
for *each h-motif* t **do**
  if $17 \leq t \leq 22$ **then** ▷ open h-motifs
    $\hat{M}[t] \leftarrow \hat{M}[t] \cdot \frac{|\wedge|}{2r}$
  else ▷ closed h-motifs
    $\hat{M}[t] \leftarrow \hat{M}[t] \cdot \frac{|\wedge|}{3r}$
return $\hat{M}$





Experiments

Datasets

- Use the following eleven real-world hypergraphs from five different domains

Table 2: Statistics of 11 real hypergraphs from 5 domains.

Dataset	$\|V\|$	$\|E\|$	$\|\bar{e}\|$*	$\|\wedge\|$	# H-motifs
coauth-DBLP	1,924,991	2,466,792	25	125M	26.3B $\pm$ 18M
coauth-geology	1,256,385	1,203,895	25	37.6M	6B $\pm$ 4.8M
coauth-history	1,014,734	895,439	25	1.7M	83.2M
contact-primary	242	12,704	5	2.2M	617M
contact-high	327	7,818	5	593K	69.7M
email-Enron	143	1,512	18	87.8K	9.6M
email-EU	998	25,027	25	8.3M	7B
tags-ubuntu	3,029	147,222	5	564M	4.3T $\pm$ 1.5B
tags-math	1,629	170,476	5	913M	9.2T $\pm$ 3.2B
threads-ubuntu	125,602	166,999	14	21.6M	11.4B
threads-math	176,445	595,749	21	647M	2.2T $\pm$ 883M

$*$ The maximum size of a hyperedge.

Lee, Geon, Jihoon Ko, and Kijung Shin. "Hypergraph Motifs: Concepts, Algorithms, and Discoveries." *Proceedings of the VLDB Endowment* 13.11.

Experiments

☐ Comparison with Random

- Distinguish between real-world hypergraphs and randomized hypergraphs

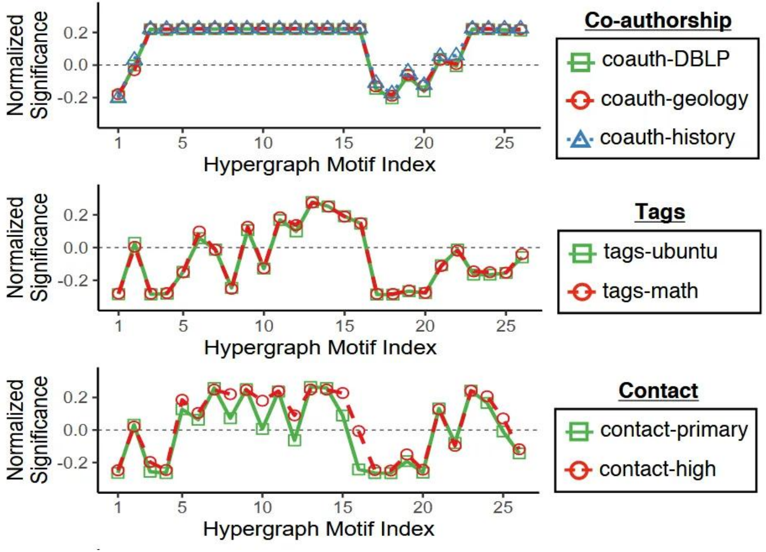
h-motif	coauth-DBLP				contact-primary				email-EU				tags-math				threads-math			
	count (rank)		RD	RC	count (rank)		RD	RC	count (rank)		RD	RC	count (rank)		RD	RC	count (rank)		RD	RC
	real	random			real	random			real	random			real	random			real	random		
1	9.6E07 (7)	1.3E09 (4)	3	-0.86	4.8E04 (16)	2.8E07 (5)	11	-1.00	7.5E06 (13)	1.7E08 (7)	6	-0.91	9.0E08 (13)	2.2E11 (6)	7	-0.99	6.4E08 (7)	2.4E11 (4)	3	-0.99
2	7.0E09 (2)	7.2E09 (2)	0	-0.01	1.1E08 (3)	8.6E07 (3)	0	0.12	6.3E08 (2)	8.2E08 (3)	1	-0.13	1.6E12 (2)	1.6E12 (2)	0	0.02	1.1E12 (2)	7.7E11 (2)	0	0.16
3	2.2E06 (17)	6.1E03 (14)	3	0.99	2.8E03 (21)	1.7E05 (16)	5	-0.97	1.6E06 (21)	7.8E05 (17)	4	0.34	3.0E06 (20)	1.1E09 (15)	5	-0.99	1.7E05 (20)	1.7E08 (14)	6	-1.00
4	9.6E06 (11)	1.1E05 (12)	1	0.98	8.4E02 (24)	9.2E05 (12)	12	-1.00	4.3E06 (16)	1.5E07 (12)	4	-0.55	1.5E08 (17)	1.6E10 (12)	5	-0.98	3.1E06 (13)	1.2E09 (11)	2	-0.99
5	1.5E08 (6)	1.2E05 (11)	5	1.00	4.6E06 (5)	1.6E06 (11)	6	0.49	7.5E07 (7)	1.1E07 (13)	6	0.74	7.4E09 (8)	2.5E10 (8)	0	-0.54	4.1E08 (8)	1.7E09 (10)	2	-0.61
6	9.9E08 (3)	1.8E06 (9)	6	1.00	1.3E07 (4)	8.2E06 (7)	3	0.24	3.9E08 (4)	1.9E08 (6)	2	0.34	6.8E11 (3)	3.3E11 (4)	1	0.35	1.4E10 (4)	1.1E10 (8)	4	0.11
7	1.9E05 (23)	0.0E00 (20)	3	1.00	1.6E04 (17)	2.0E02 (24)	7	0.98	7.5E04 (24)	1.2E02 (25)	1	1.00	8.3E05 (25)	9.1E05 (25)	0	-0.05	8.8E03 (24)	1.7E04 (24)	0	-0.32
8	3.9E05 (22)	0.0E00 (20)	2	1.00	4.6E03 (20)	2.6E03 (22)	2	0.27	4.2E06 (17)	2.5E04 (21)	4	0.99	2.0E06 (23)	3.4E07 (22)	1	-0.89	2.2E04 (23)	3.5E05 (21)	2	-0.88
9	2.4E06 (16)	0.0E00 (20)	4	1.00	1.7E05 (12)	4.6E03 (20)	8	0.95	1.8E06 (20)	1.1E04 (22)	2	0.99	1.4E08 (18)	5.4E07 (21)	3	0.45	5.1E05 (17)	4.5E05 (20)	3	0.06
10	7.6E06 (13)	7.5E00 (18)	5	1.00	5.7E04 (15)	5.5E04 (17)	2	0.03	2.8E07 (10)	1.7E06 (14)	4	0.88	7.1E08 (14)	1.9E09 (14)	0	-0.45	2.3E06 (15)	9.4E06 (17)	2	-0.61
11	8.6E06 (12)	0.9E00 (19)	7	1.00	4.1E05 (11)	2.4E04 (18)	7	0.89	9.0E06 (11)	1.9E05 (19)	8	0.96	3.5E09 (10)	7.4E08 (16)	6	0.65	2.8E06 (14)	3.1E06 (18)	4	-0.05
12	6.4E07 (8)	1.9E02 (16)	8	1.00	1.7E05 (13)	2.7E05 (14)	1	-0.24	8.2E07 (6)	2.4E07 (10)	4	0.55	6.9E10 (6)	2.4E10 (10)	4	0.49	8.2E07 (10)	6.2E07 (15)	5	0.14
13	1.6E04 (26)	0.0E00 (20)	6	1.00	5.5E03 (19)	1.6E00 (26)	7	1.00	2.7E04 (26)	0.4E00 (26)	0	1.00	1.1E06 (24)	1.7E04 (26)	2	0.97	1.5E02 (26)	8.6E00 (26)	0	0.89
14	1.4E05 (24)	0.0E00 (20)	4	1.00	6.0E03 (18)	7.1E01 (25)	7	0.98	7.2E05 (22)	3.7E02 (24)	2	1.00	2.8E07 (19)	1.8E06 (24)	5	0.88	3.9E03 (25)	9.3E02 (25)	0	0.61
15	6.5E05 (19)	0.0E00 (20)	1	1.00	1.7E03 (22)	8.6E02 (23)	1	0.34	3.6E06 (19)	5.0E04 (20)	1	0.97	2.9E08 (15)	5.7E07 (20)	5	0.67	2.7E04 (22)	2.0E04 (23)	1	0.16
16	2.0E06 (18)	0.0E00 (20)	2	1.00	1.4E02 (25)	3.2E03 (21)	4	-0.92	6.7E06 (14)	1.7E06 (15)	1	0.60	1.9E09 (11)	5.8E08 (18)	7	0.53	2.4E05 (18)	1.3E05 (22)	4	0.29
17	4.2E05 (21)	2.0E06 (8)	13	-0.65	1.0E03 (23)	6.3E05 (13)	10	-1.00	3.8E04 (25)	8.7E05 (16)	9	-0.92	5.1E05 (26)	5.0E08 (19)	7	-1.00	2.3E05 (19)	9.2E08 (12)	7	-1.00
18	2.6E06 (15)	6.4E07 (7)	8	-0.92	1.2E02 (26)	7.0E06 (8)	18	-1.00	6.0E06 (15)	4.0E07 (8)	7	-0.74	2.5E06 (22)	1.6E10 (13)	9	-1.00	8.3E05 (16)	1.3E10 (7)	9	-1.00
19	3.6E07 (9)	6.7E07 (6)	3	-0.30	2.0E06 (6)	1.2E07 (6)	0	-0.72	8.7E06 (12)	2.9E07 (9)	3	-0.54	9.4E08 (12)	2.4E10 (9)	3	-0.93	3.5E08 (9)	1.8E10 (6)	3	-0.96
20	3.4E08 (5)	2.2E09 (3)	2	-0.73	6.0E05 (10)	1.3E08 (2)	8	-0.99	2.2E08 (5)	1.2E09 (2)	3	-0.69	9.2E09 (7)	7.2E11 (3)	4	-0.97	1.9E09 (5)	2.4E11 (3)	2	-0.98
21	7.9E08 (4)	5.6E08 (5)	1	0.17	1.7E08 (2)	5.7E07 (4)	2	0.50	5.3E08 (3)	2.3E08 (4)	1	0.39	1.2E11 (5)	2.8E11 (5)	0	-0.40	2.8E10 (3)	8.6E10 (5)	2	-0.51
22	1.7E10 (1)	1.8E10 (1)	0	-0.03	3.1E08 (1)	5.8E08 (1)	0	-0.30	4.9E09 (1)	8.5E09 (1)	0	-0.27	6.6E12 (1)	7.6E12 (1)	0	-0.07	1.1E12 (1)	1.2E12 (1)	0	-0.02
23	2.4E04 (25)	1.5E01 (17)	8	1.00	1.2E05 (14)	5.4E03 (19)	5	0.91	8.8E04 (23)	4.0E03 (23)	0	0.91	2.6E06 (21)	7.9E06 (23)	2	-0.51	1.4E05 (21)	7.8E05 (19)	2	-0.70
24	4.4E05 (20)	1.4E03 (15)	5	0.99	7.7E05 (9)	1.8E05 (15)	6	0.63	4.2E06 (18)	5.4E05 (18)	0	0.77	2.2E08 (16)	7.2E08 (17)	1	-0.53	7.5E06 (12)	3.1E07 (16)	4	-0.61
25	3.8E06 (14)	4.6E04 (13)	1	0.98	1.7E06 (8)	1.8E06 (10)	2	-0.03	3.2E07 (9)	2.0E07 (11)	2	0.23	6.0E09 (9)	2.0E10 (11)	2	-0.54	8.0E07 (11)	4.2E08 (13)	2	-0.68
26	2.3E07 (10)	4.9E05 (10)	0	0.96	1.8E06 (7)	6.14E06 (9)	2	-0.54	7.5E07 (8)	2.1E08 (5)	3	-0.48	1.3E11 (4)	1.8E11 (7)	3	-0.14	1.2E09 (6)	1.9E09 (9)	3	-0.21

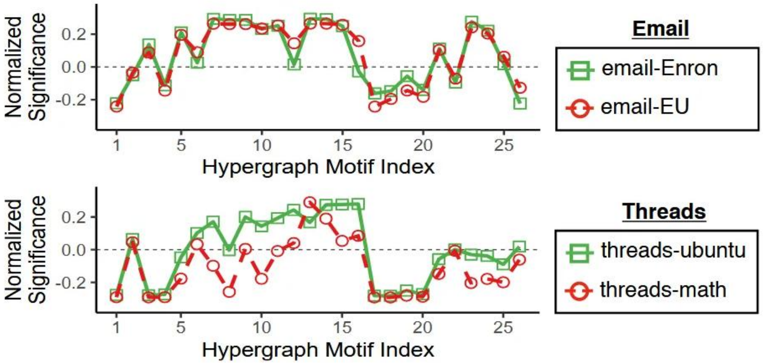
Lee, Geon, Jihoon Ko, and Kijung Shin. "Hypergraph Motifs: Concepts, Algorithms, and Discoveries." *Proceedings of the VLDB Endowment* 13.11.

Experiments

Comparison across Domains

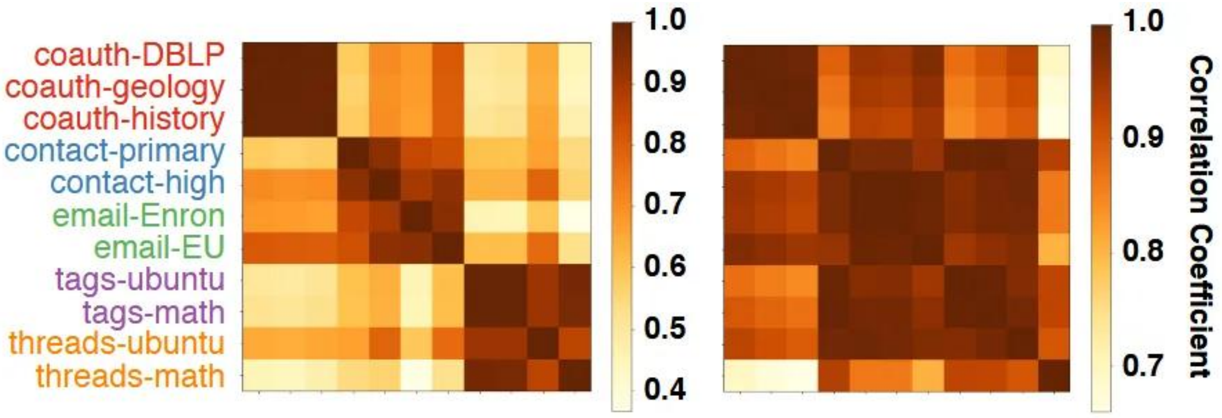
- Hypergraphs from the same domains have similar CPs





Experiments

- **CPs compared with CPs based on network motifs**
  - Distinguish the domains of the real-world hypergraphs better than the CPs based on network motifs



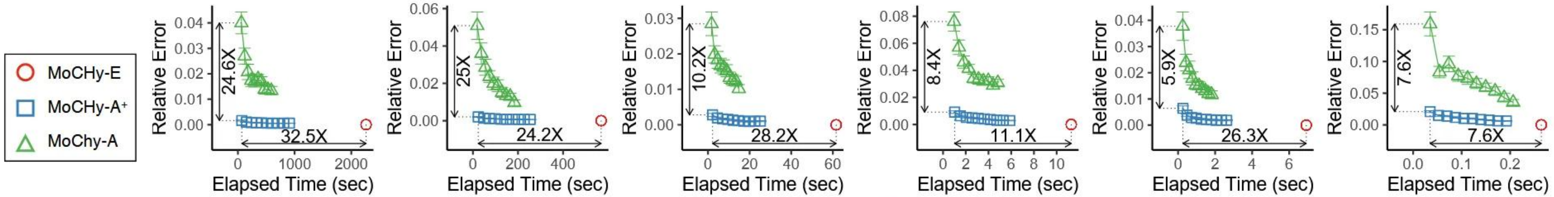
(a) Similarity matrix based on hypergraph motifs (b) Similarity matrix based on network motifs

Experiments

☐ Speed and Accuracy

- MoCHy-A+ gives the best trade-off between speed accuracy

$$\frac{\sum_{t=1}^{26}|M[t]-\bar{M}[t]|}{\sum_{t=1}^{26}M[t]} \text{ and } \frac{\sum_{t=1}^{26}|M[t]-\hat{M}[t]|}{\sum_{t=1}^{26}M[t]}$$



(a) threads-ubuntu (b) email-Eu (c) contact-primary (d) coauth-history (e) contact-high (f) email-Enron

Experiments

☐ Relation between h-motifs and global structural properties

- Correlation between global structural properties and characteristic profiles (CPs)

h-motif	Size: Node	Size: Hyperedge	Average Degree: Node	Average Degree: Hyperedge	Clustering Coeff.: Node	Clustering Coeff.: Hyperedge	Effective Diamter	# of H-motifs
1	+0.787	+0.648	−0.497	+0.003	+0.751	+0.115	+0.626	−0.530
2	−0.169	−0.110	−0.023	−0.645	−0.027	+0.479	+0.099	+0.101
3	+0.710	+0.587	−0.492	+0.065	+0.718	+0.067	+0.633	−0.501
4	+0.901	+0.783	−0.480	−0.071	+0.891	+0.122	+0.806	−0.410
5	+0.469	+0.315	−0.506	−0.126	+0.498	+0.316	+0.406	−0.730
6	+0.877	+0.762	−0.311	−0.161	+0.872	+0.151	+0.801	−0.249
7	+0.229	+0.071	−0.436	−0.188	+0.313	+0.394	+0.215	−0.744
8	+0.388	+0.237	−0.561	−0.154	+0.447	+0.359	+0.361	−0.772
9	+0.138	−0.023	−0.266	−0.081	+0.229	+0.293	+0.124	−0.606
10	+0.444	+0.288	−0.584	−0.178	+0.537	+0.363	+0.447	−0.733
11	+0.123	−0.041	+0.023	+0.098	+0.200	+0.077	+0.087	−0.329
12	+0.611	+0.540	−0.246	−0.096	+0.676	+0.050	+0.643	−0.061
13	−0.558	−0.485	+0.523	+0.552	−0.650	−0.499	−0.620	+0.413
14	−0.490	−0.567	+0.232	+0.156	−0.351	−0.010	−0.417	−0.064
15	+0.166	+0.042	−0.148	+0.080	+0.299	+0.045	+0.224	−0.224
16	+0.481	+0.464	−0.184	+0.028	+0.531	−0.110	+0.532	+0.106
17	+0.754	+0.608	−0.501	−0.127	+0.828	+0.231	+0.741	−0.481
18	+0.442	+0.298	−0.444	+0.031	+0.551	+0.149	+0.473	−0.532
19	+0.623	+0.471	−0.507	−0.089	+0.677	+0.251	+0.584	−0.596
20	+0.628	+0.473	−0.483	−0.001	+0.724	+0.136	+0.633	−0.514
21	+0.089	−0.047	−0.315	−0.136	+0.141	+0.346	+0.074	−0.676
22	+0.554	+0.522	−0.157	−0.299	+0.724	+0.127	+0.763	+0.154
23	+0.332	+0.181	−0.434	−0.130	+0.365	+0.334	+0.273	−0.722
24	+0.428	+0.275	−0.492	−0.147	+0.459	+0.341	+0.368	−0.737
25	+0.747	+0.593	−0.508	−0.151	+0.758	+0.269	+0.666	−0.610
26	+0.883	+0.812	−0.330	−0.203	+0.877	+0.118	+0.830	−0.129

Supplementary document. Available online: https://github.com/geonlee0325/MoCHy, 2020

Experiments

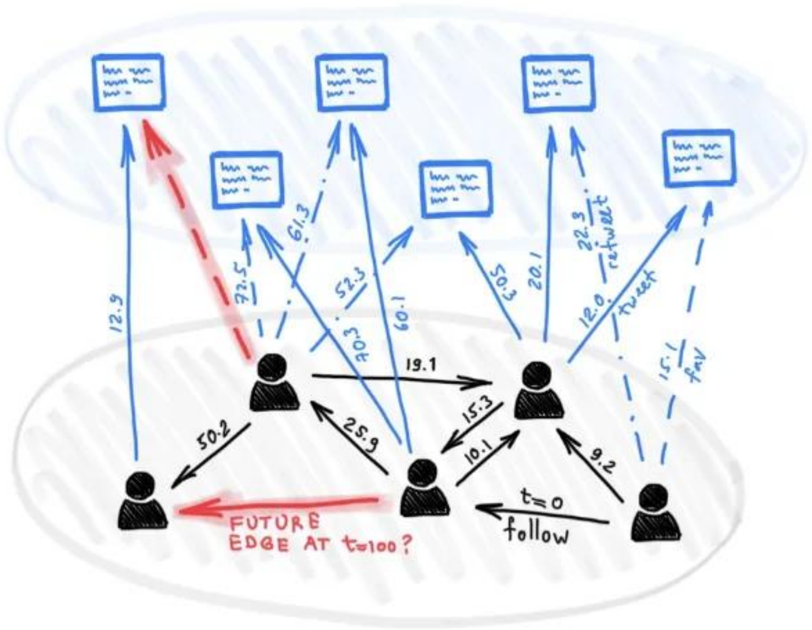
- Model relating global properties of hypergraphs with CPs of k sampled h-motifs
  - $y = w_0 + w_1 \cdot CP_{s_1} + w_2 \cdot CP_{s_2} + \cdots + w_k \cdot CP_{s_k}$
  - Using more CPs (larger k) improves the accuracy more
  - Can induce global properties using h-motifs

			Adjusted R^2	Selected h-motifs
$k=3$	**Size**	Node	0.94857	3, 4, and 20
		Hyperedge	0.92933	3, 17, and 20
	Average Degree	Node	0.87320	7, 9, and 15
		Hyperedge	0.87907	2, 20, and 26
	Clustering Coefficient	Node	0.98085	2, 19, and 22
		Hyperedge	0.89210	2, 18, and 22
	Efficient Diameter		0.95428	1, 19, and 26
	# of H-motifs		0.96930	7, 9, and 14
$k=5$	**Size**	Node	0.99607	4, 10, 20, 22, and 25
		Hyperedge	0.98710	4, 10, 20, 21, and 26
	Average Degree	Node	0.99352	7, 9, 14, 15, and 19
		Hyperedge	0.97775	2, 3, 13, 14, and 18
	Clustering Coefficient	Node	0.99710	3, 8, 10, 17, and 22
		Hyperedge	0.97359	1, 9, 13, 16, and 19
	Efficient Diameter		0.99417	1, 17, 22, 23, and 24
	# of H-motifs		0.99800	5, 7, 9, 11, and 14
$k=7$	**Size**	Node	0.99995	4, 7, 15, 20, 21, 24, and 26
		Hyperedge	0.99990	5, 7, 9, 20, 24, 25, and 26
	Average Degree	Node	0.99993	6, 7, 8, 9, 12, 13, and 14
		Hyperedge	0.99979	2, 4, 7, 11, 20, 21, and 23
	Clustering Coefficient	Node	0.99999	3, 4, 11, 16, 18, 22, and 24
		Hyperedge	0.99976	1, 2, 3, 17, 18, 21, and 24
	Efficient Diameter		0.99997	3, 8, 14, 18, 19, 22, and 25
	# of H-motifs		0.99997	7, 8, 9, 10, 11, 14, and 22

Supplementary document. Available online: https://github.com/geonlee0325/MoCHy, 2020

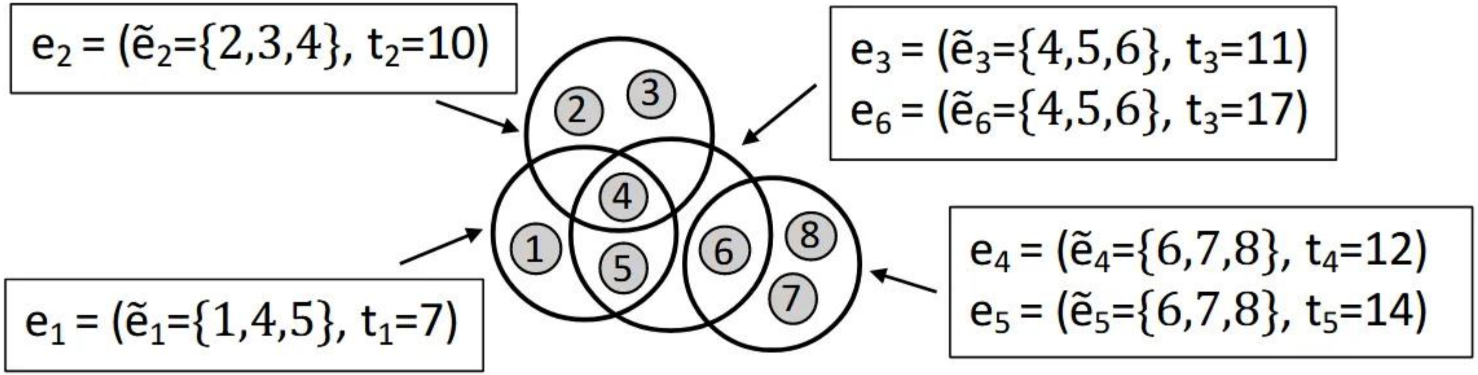
Real-world graphs can change over time

☐ **Add/Delete nodes, edges**



Temporal Hypergraphs

- $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$
  - $\mathcal{E} = \{(e_1, t_1), (e_2, t_2), \dots\}$
  - $X \in \mathbb{R}^{|\mathcal{V}| \times f}$
    - A matrix that encodes node attribute information

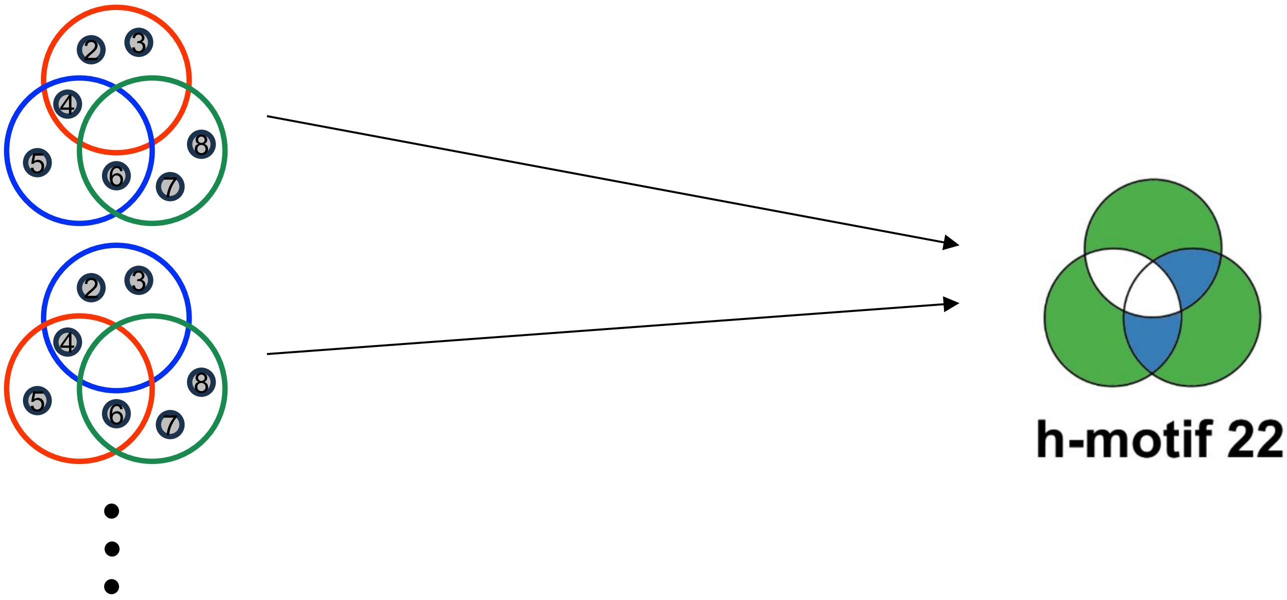


(a) An example temporal hypergraph

Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Limitation of h-motifs

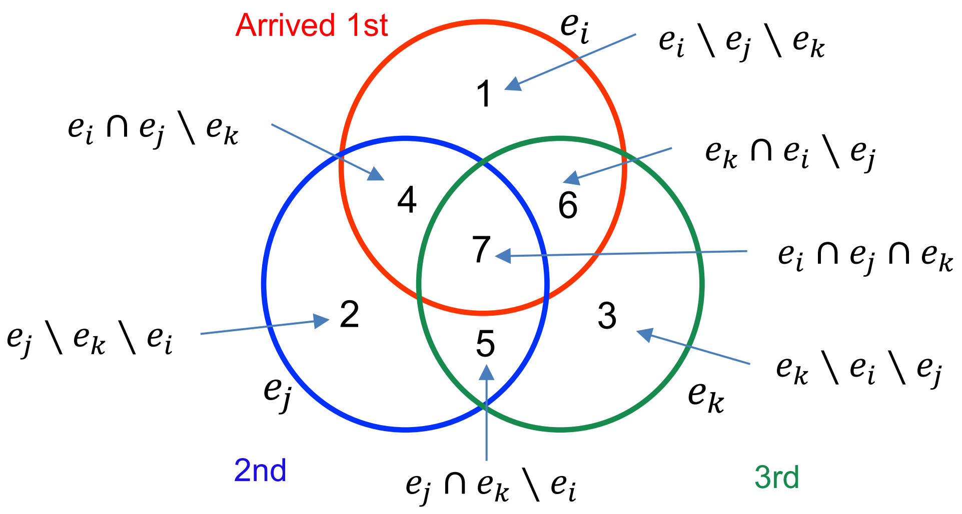
- Temporal dynamics are completely ignored



TH-motifs

Representation

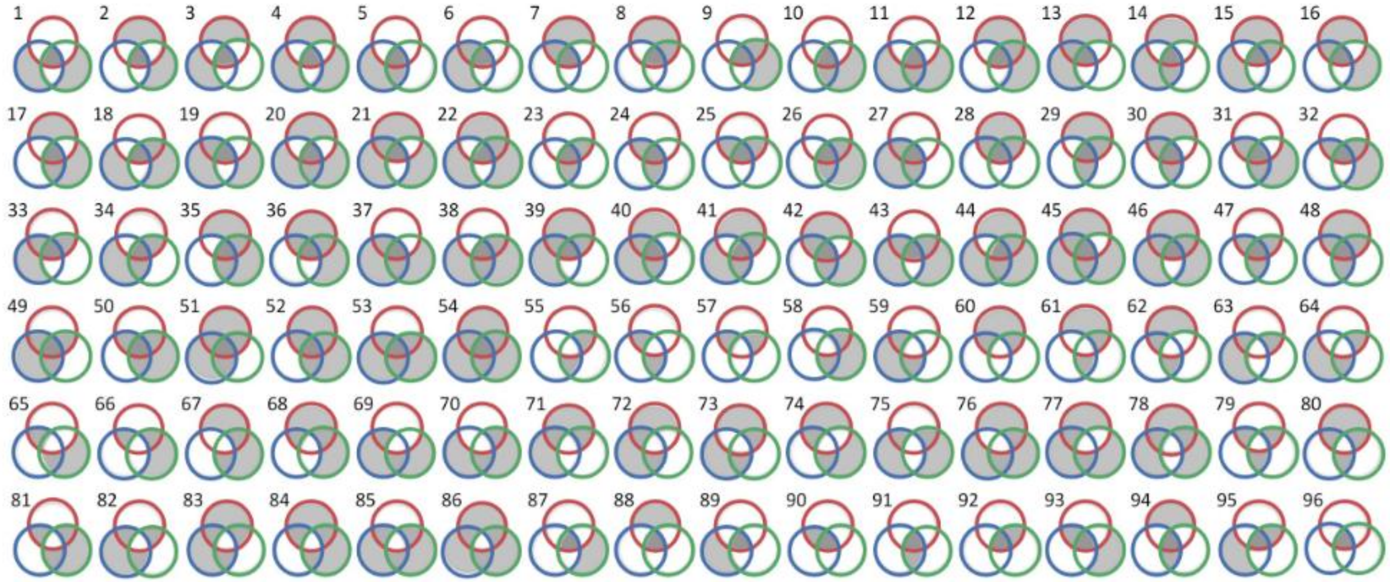
- $< e_i = (\widetilde{e_i}, t_i), e_j = (\widetilde{e_j}, t_j), e_k = (\widetilde{e_k}, t_k) >$
- $t_i < t_j < t_k$
- $t_k - t_i \leq \delta$ ($i.e.$, they arrive within a predefined time interval δ)



TH-motifs

☐ Representation

- Consider 96 cases, after excluding those describing disconnected hyperedges



Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

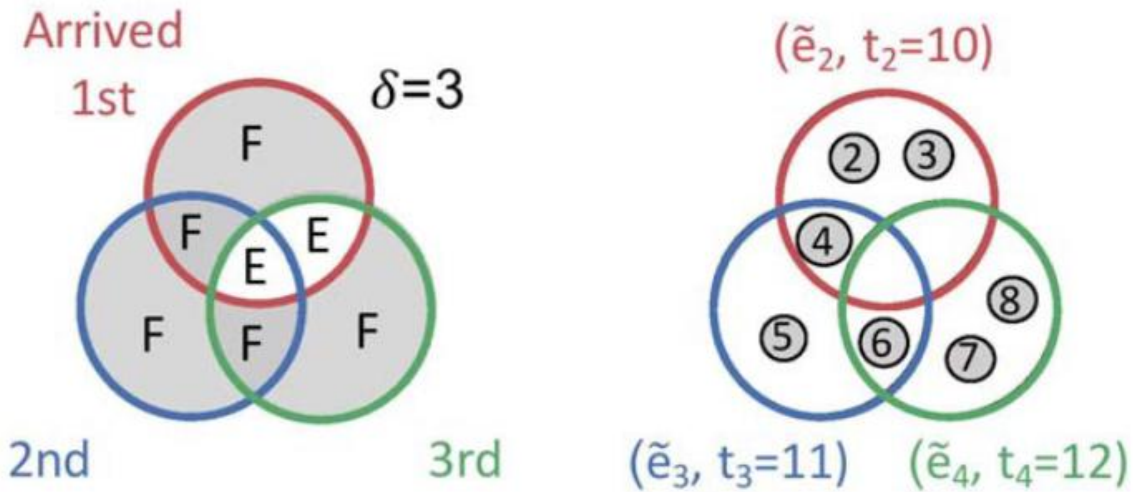
TH-motifs

- Three types
  - Triple-inducing
    - Hyperedges in its instance $< e_i, e_j, e_k >$ are distinct ($i.e., \widetilde{e_i} \neq \widetilde{e_j}, \widetilde{e_j} \neq \widetilde{e_k}, and\ \widetilde{e_k} \neq \widetilde{e_i}$)
  - Pair-inducing
    - Two are duplicated while the remaining one is different
  - Single-inducing
    - All three hyperedges are duplicated

TH-motifs

Instances

- A sequence $\langle e_i, e_j, e_k \rangle$ of three temporal hyperedges is an instance of TH-motif t



Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Algorithms

☐ THyMe

- Directly enumerate each instance of TH-motifs
- Consider the temporal hyperedges that occur in the δ-sized temporal window
- Counting the instances in temporal hypergraph can be more computationally challenging

Algorithm 1: THYME: Preliminary Algorithm

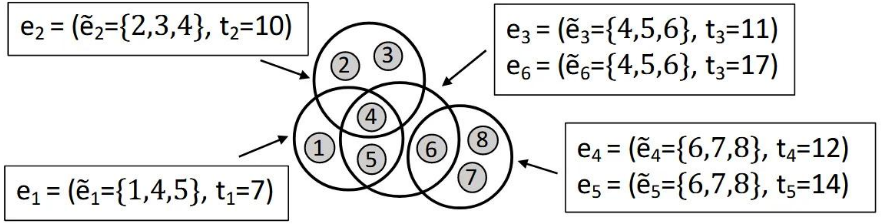
Input : (1) temporal hypergraph: $T = (V, \mathcal{E})$
(2) time interval δ

Output: # of each temporal h-motif t's instances: $M[t]$

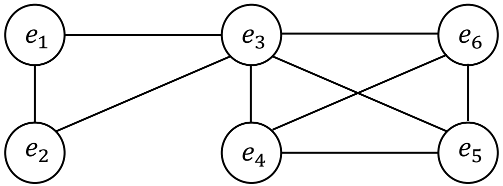
$M \leftarrow$ map initialized to 0
$P = (V_P = \varnothing, E_P = \varnothing)$
$w_s \leftarrow 1$
for each temporal hyperedge $e_i = (\tilde{e}_i, t_i) \in \mathcal{E}$ **do**
  `insert` (e_i)
  while $t_{w_s} + \delta < t_i$ **do**
    `remove` (e_{w_s})
    $w_s \leftarrow w_s + 1$
  $S \leftarrow$set of 3 connected temporal hyperedges including e_i
  for each instance $\langle e_j, e_k, e_i \rangle \in S$ **do**
    $M[h(\tilde{e}_j, \tilde{e}_k, \tilde{e}_i)]$ += 1
return M
Procedure `insert` $(e_i = (\tilde{e}_i, t_i))$
  $V_P \leftarrow V_P \cup \{e_i\}$
  $N_{e_i} \leftarrow \{e : e \in V_P \setminus \{e_i\} \text{ and } \tilde{e}_i \cap \tilde{e} \neq \varnothing\}$
  $E_P \leftarrow E_P \cup \{(e_i, e) : e \in N_{e_i}\}$
Procedure `remove` $(e_i = (\tilde{e}_i, t_i))$
  $V_P \leftarrow V_P \setminus \{e_i\}$
  $N_{e_i} \leftarrow \{e : e \in V_P \text{ and } \tilde{e}_i \cap \tilde{e} \neq \varnothing\}$
  $E_P \leftarrow E_P \setminus \{(e_i, e) : e \in N_{e_i}\}$

Update projected graph

Count instances



(a) An example temporal hypergraph



Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Algorithms

- **THyMe+**
  - $Q = (V_Q, E_Q, t_Q)$
    - V_Q
      - A static hyperedge
    - E_Q
      - A pair of static hyperedges that share any nodes
    - t_Q
      - A set of timestamps of temporal hyperedges inducing a particular static hyperedge

Algorithms

☐ THyMe+

Algorithm 2: THYME+: Proposed Algorithm

Input : (1) temporal hypergraph: $T = (V, \mathcal{E})$
(2) time interval δ

Output: # of each temporal h-motif t's instances: $M[t]$

$M \leftarrow$ map initialized to 0
$Q = (V_Q = \varnothing, E_Q = \varnothing, t_Q = \varnothing)$
$w_s \leftarrow 1$
for each temporal hyperedge $e_i = (\tilde{e}_i, t_i) \in \mathcal{E}$ **do**
  `insert` (e_i)
  while $t_{w_s} + \delta < t_i$ **do**
    `remove` (e_{w_s})
    $w_s \leftarrow w_s + 1$
  $S \leftarrow$ set of 3 connected static hyperedges including $\tilde{e}_i$
  for each instance $\{\tilde{e}_i, \tilde{e}_j, \tilde{e}_k\} \in S$ **do**
    `comb3` $(\tilde{e}_i, \tilde{e}_j, \tilde{e}_k)$
  for each pair $(\tilde{e}_i, \tilde{e}_j) \in N_{\tilde{e}_i}$ **do**
    `comb2` $(\tilde{e}_i, \tilde{e}_j)$
  `comb1` $(\tilde{e}_i)$
return M

Procedure `insert` $(e_i = (\tilde{e}_i, t_i))$
  if $\tilde{e}_i \notin V_Q$ **then**
    $V_Q \leftarrow V_Q \cup \{\tilde{e}_i\}$
    $N_{\tilde{e}_i} \leftarrow \{\tilde{e} : \tilde{e} \in V_Q \setminus \{\tilde{e}_i\} \text{ and } \tilde{e}_i \cap \tilde{e} \neq \varnothing\}$
    $E_Q \leftarrow E_Q \cup \{(\tilde{e}_i, \tilde{e}) : \tilde{e} \in N_{\tilde{e}_i}\}$
  $t_Q(\tilde{e}_i) \leftarrow t(\tilde{e}_i) \cup \{t_i\}$

Procedure `remove` $(e_i = (\tilde{e}_i, t_i))$
  $t(\tilde{e}_i) \leftarrow t(\tilde{e}_i) \setminus \{t_i\}$
  if $t_Q(\tilde{e}_i) = \varnothing$ **then**
    $V_Q \leftarrow V_Q \setminus \{\tilde{e}_i\}$
    $N_{\tilde{e}_i} \leftarrow \{\tilde{e} : \tilde{e} \in V_Q \text{ and } e_i \cap e \neq \varnothing\}$
    $E_Q \leftarrow E_Q \setminus \{(\tilde{e}_i, \tilde{e}) : \tilde{e} \in N_{\tilde{e}_i}\}$

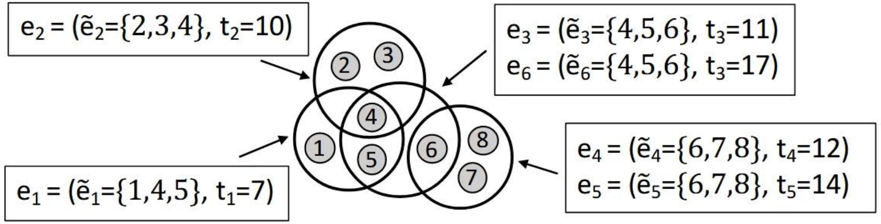
Procedure `comb3` $(\tilde{e}_i, \tilde{e}_j, \tilde{e}_k)$
  $M[h(\tilde{e}_j, \tilde{e}_k, \tilde{e}_i)] \mathrel{+}= \sum_{t \in t_Q(\tilde{e}_j), t' \in t_Q(\tilde{e}_k)} \mathbb{1}[t < t']$
  $M[h(\tilde{e}_k, \tilde{e}_j, \tilde{e}_i)] \mathrel{+}= \sum_{t \in t_Q(\tilde{e}_j), t' \in t_Q(\tilde{e}_k)} \mathbb{1}[t' < t]$

Procedure `comb2` $(\tilde{e}_i, \tilde{e}_j)$
  $M[h(\tilde{e}_i, \tilde{e}_j, \tilde{e}_i)] \mathrel{+}= \sum_{t \in t_Q(\tilde{e}_i) \setminus \{t_i\}, t' \in t_Q(\tilde{e}_j)} \mathbb{1}[t < t']$
  $M[h(\tilde{e}_j, \tilde{e}_i, \tilde{e}_i)] \mathrel{+}= \sum_{t \in t_Q(\tilde{e}_i) \setminus \{t_i\}, t' \in t_Q(\tilde{e}_j)} \mathbb{1}[t' < t]$
  $M[h(\tilde{e}_j, \tilde{e}_j, \tilde{e}_i)] \mathrel{+}= \binom{|t_Q(\tilde{e}_j)|}{2}$

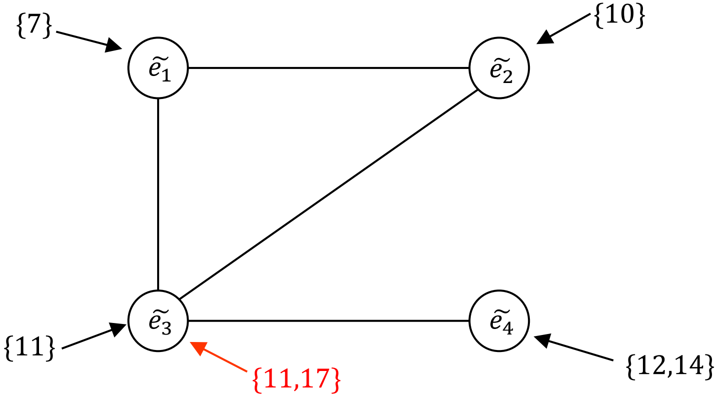
Procedure `comb1` $(\tilde{e}_i)$
  $M[h(\tilde{e}_i, \tilde{e}_i, \tilde{e}_i)] \mathrel{+}= \binom{|t_Q(\tilde{e}_i) - \{t_i\}|}{2}$

Update projected graph

Count instances



(a) An example temporal hypergraph



Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Experiments

☐ Datasets

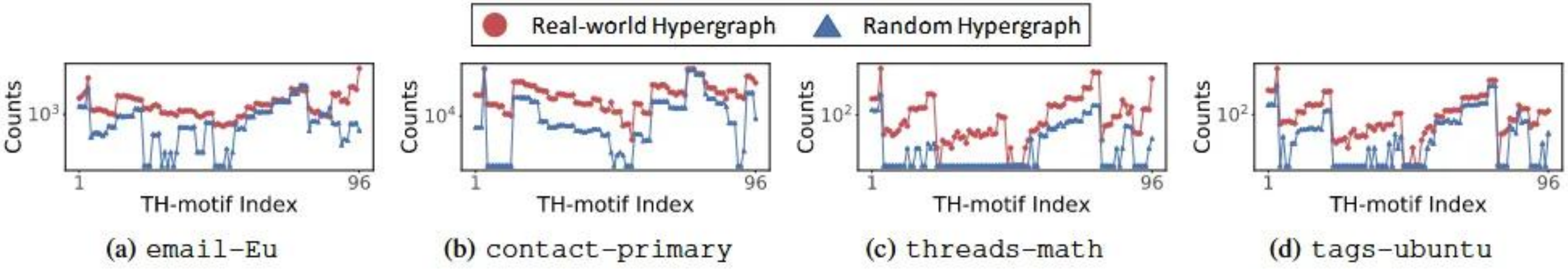
- Use the following eleven real-world hypergraphs from five different domains

Dataset	$\lvert V \rvert$	$\lvert \mathcal{E} \rvert$	$\lvert E_{\mathcal{E}} \rvert$	$\max_{e \in \mathcal{E}} \lvert e \rvert$
email-Enron	143	10,885	1,514	37
email-Eu	986	235,263	25,148	40
contact-primary	242	106,879	12,704	5
contact-high	327	172,035	7,818	5
threads-ubuntu	90,054	192,947	166,999	14
threads-math	153,806	719,792	595,749	21
tags-ubuntu	3,021	271,233	147,222	5
tags-math	1,627	822,059	170,476	5
coauth-DBLP	1,836,596	3,700,681	2,467,389	280
coauth-Geology	1,091,979	1,591,166	1,204,704	284
coauth-History	503,868	1,813,147	896,062	925

Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Experiments

- **Real hypergraphs are not 'random'**
  - The distributions are clearly different

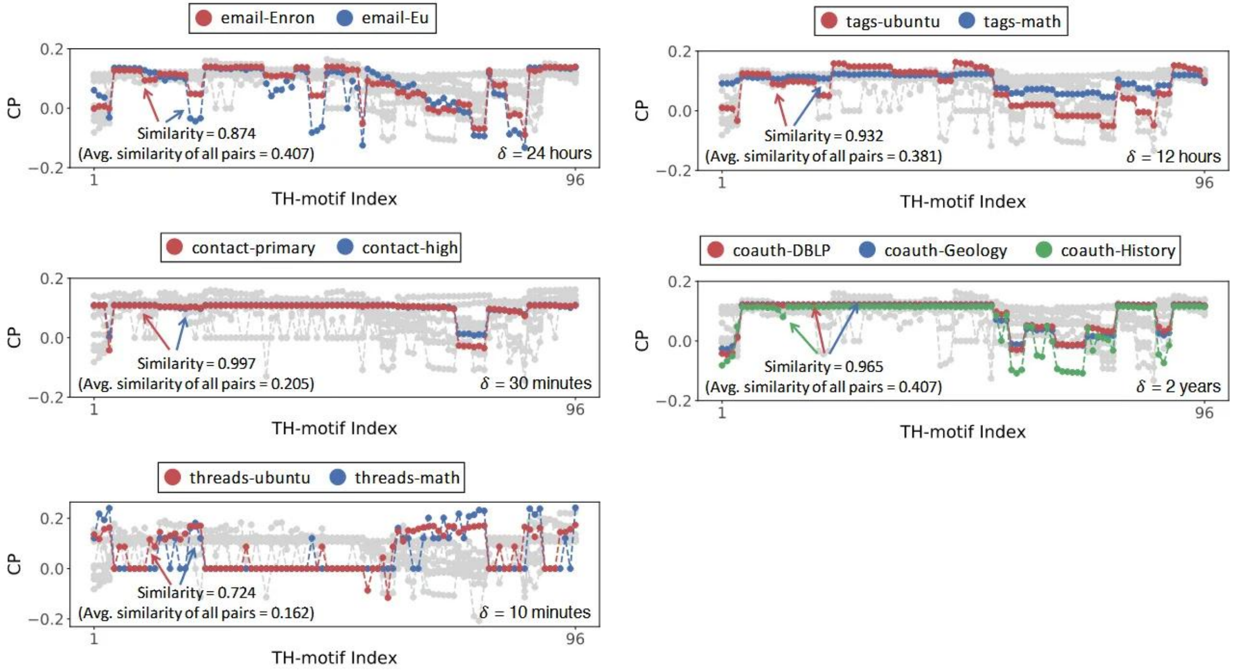


Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Experiments

☐ TH-motifs distinguish domains

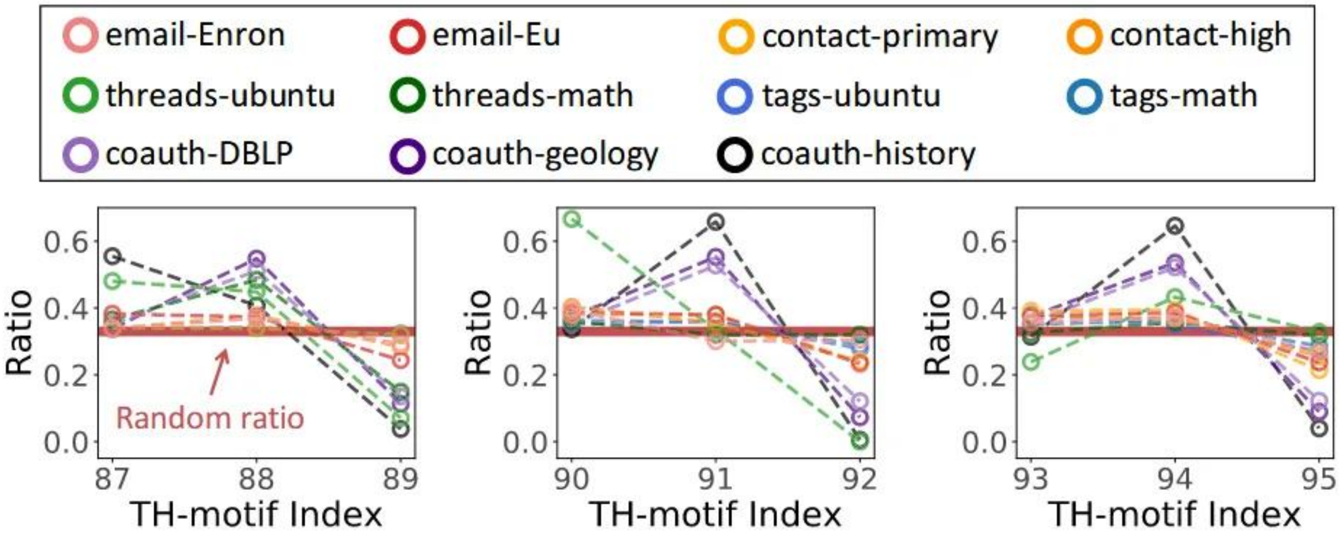
- CPs of the temporal hypergraphs from the same domain are similar



Experiments

☐ Orders of hyperedges matter

- Duplicated temporal hyperedges tend to occur in a short time and thus affect the count distributions of TH-motifs

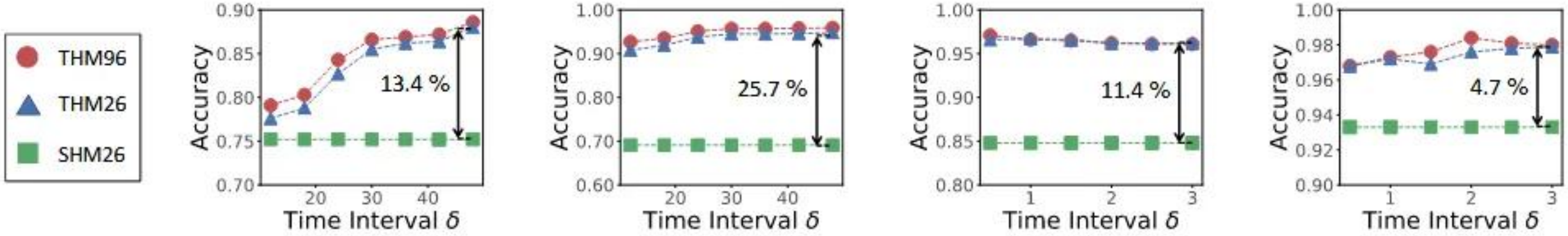


Lee, Geon, and Kijung Shin. "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting." *2021 IEEE international conference on data mining (ICDM)*. IEEE, 2021.

Experiments

☐ TH-motifs help predict future hyperedges

- TH-motifs incorporate temporal information in addition to them, and thus they are more informative



(a) email-Enron (b) email-Eu (c) contact-primary (d) contact-high

Fig. 7: TH-motifs provide informative features of temporal hyperedges. THM96 and THM26, which use the counts of TH-motifs' instance as features, are more accurate than SHM26, which uses the counts of static h-motifs' instances, in predicting future temporal hyperedges. Results in small datasets where the instances of static h-motifs can be exactly counted are reported.

Conclusion

Analyzing complex networks

- ‘Overall’ complex networks
- ‘Specific’ complex networks

Network Motifs

- Recurring, significant patterns of interconnections

Hypergraph Motifs

- H-motifs
- TH-motifs