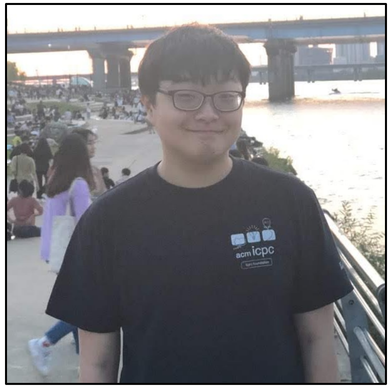
Hypergraph Motifs:
Concepts, Algorithms, and Discoveries



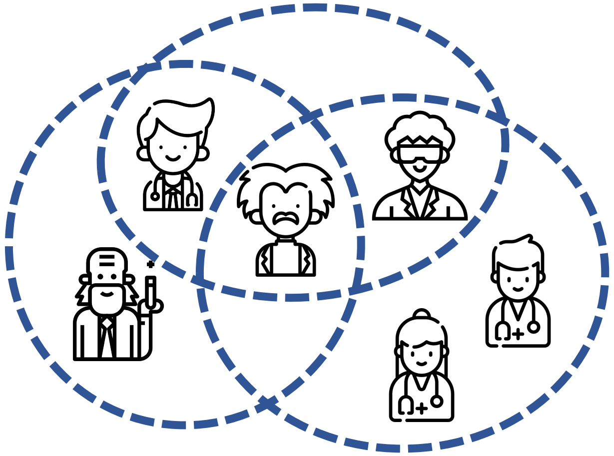
Geon Lee

Jihoon Ko

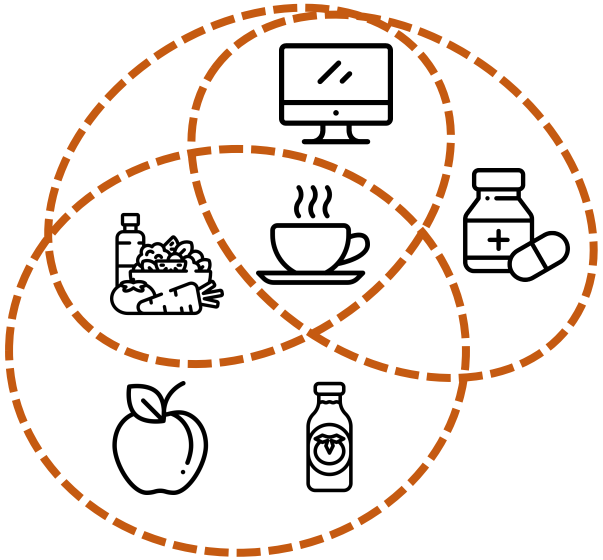
Kijung Shin

Hypergraphs are Everywhere

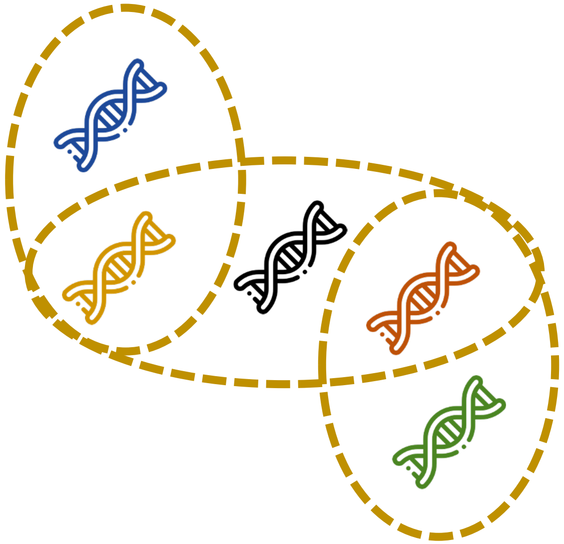
- Hypergraphs consist of nodes and hyperedges.
- Each hyperedge is a subset of any number of nodes.



Collaborations of Researchers



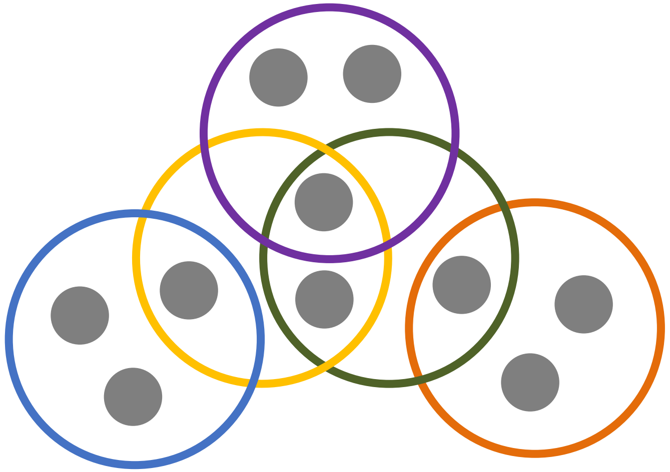
Co-purchases of Items



Joint Interactions of Proteins

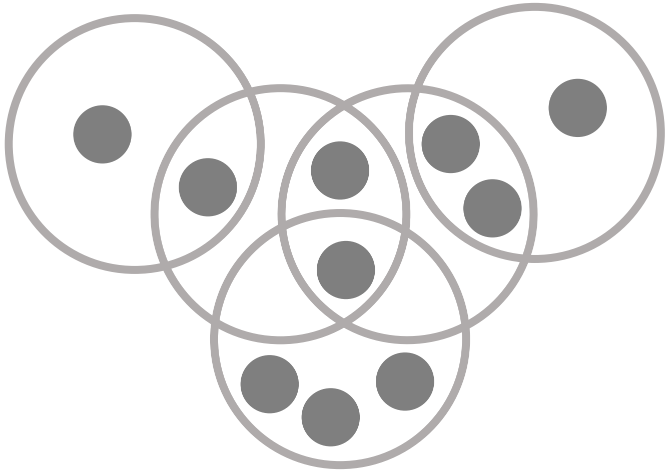
Our Questions

Q1 What are structural design principles of real-world hypergraphs?



Real-world Hypergraph

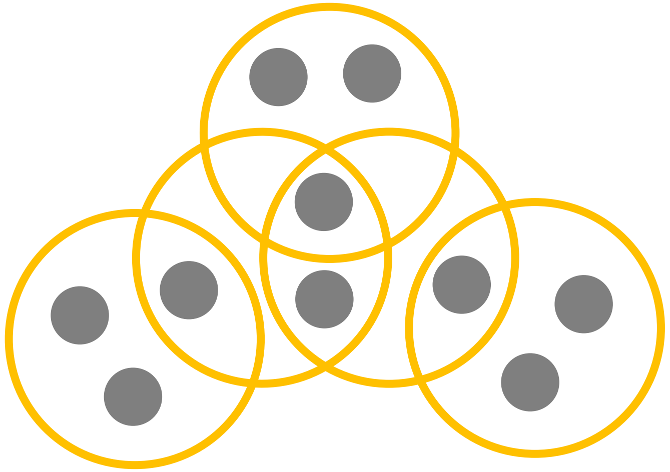
VS



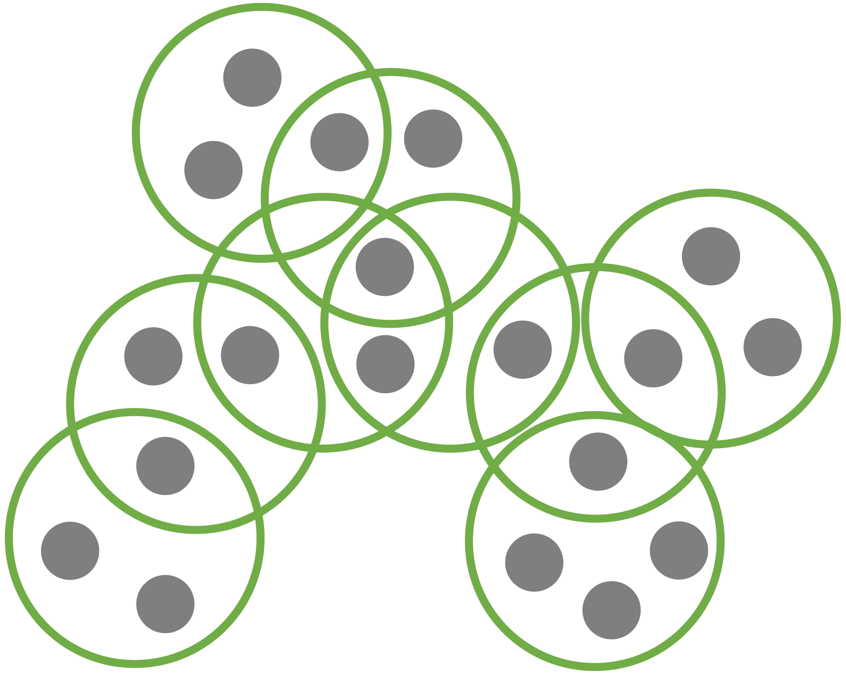
Randomized Hypergraph

Our Questions (cont.)

Q2 How can we compare local structures of hypergraphs of different sizes?



VS

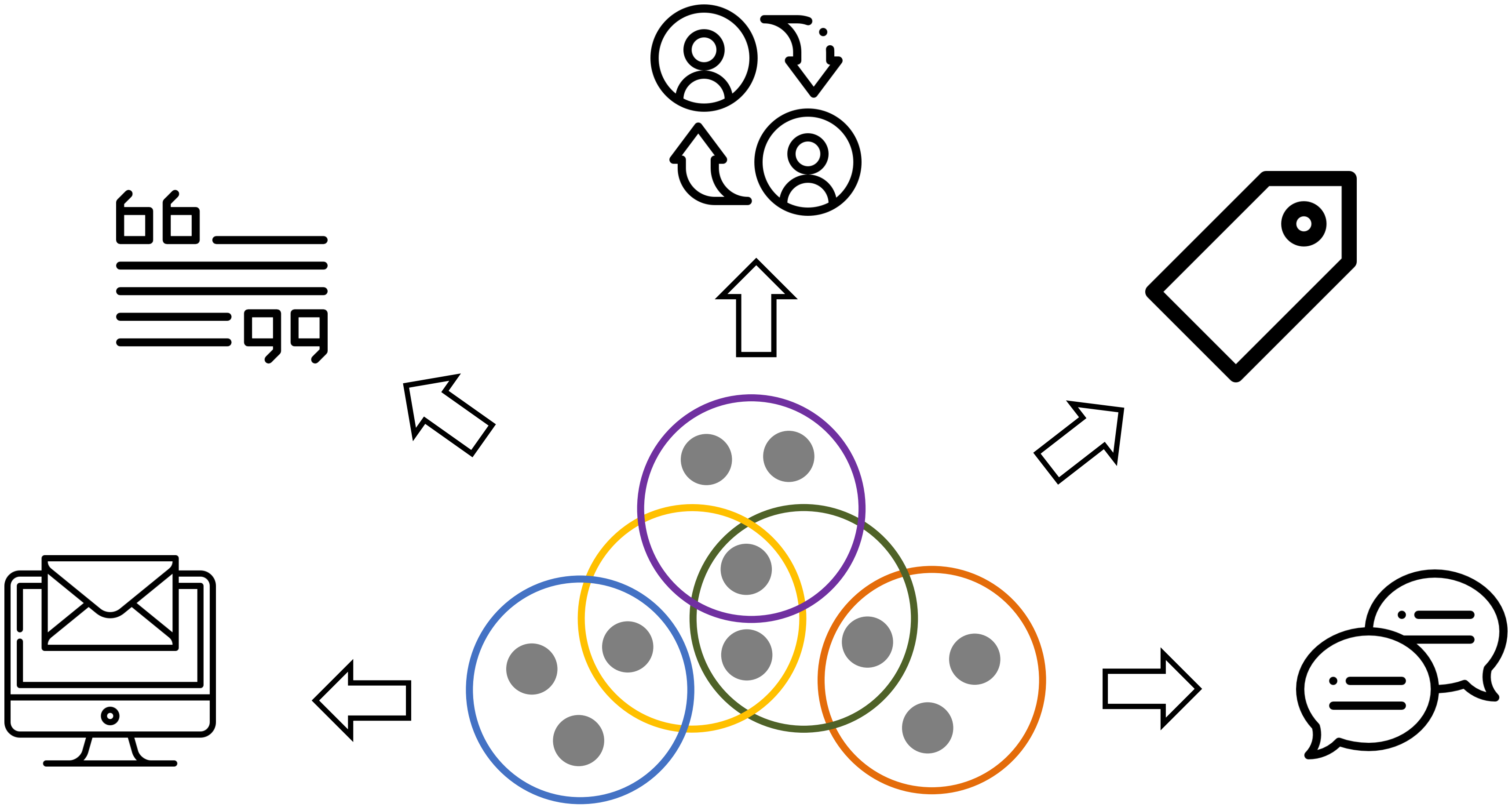


Small Hypergraph

Large Hypergraph

Our Questions (cont.)

Q3 How can we identify domains which hypergraphs are from?

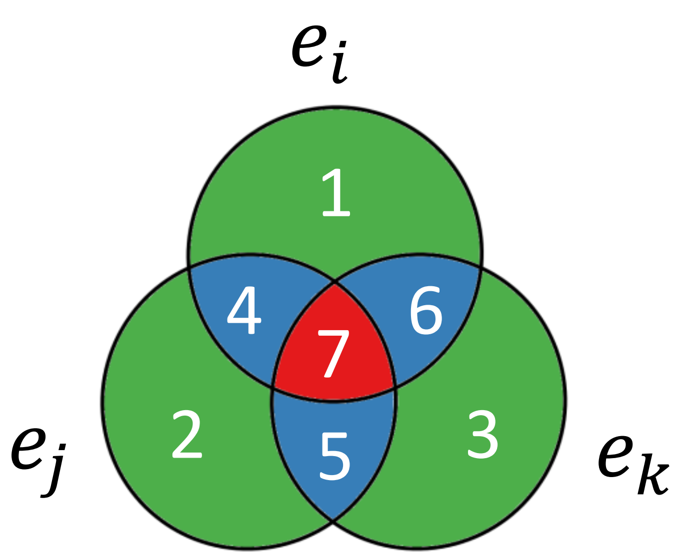


Roadmap

- Hypergraph Motif
- Proposed Method: MoCHy
- Experimental Results
- Conclusions

Hypergraph Motifs: Definition

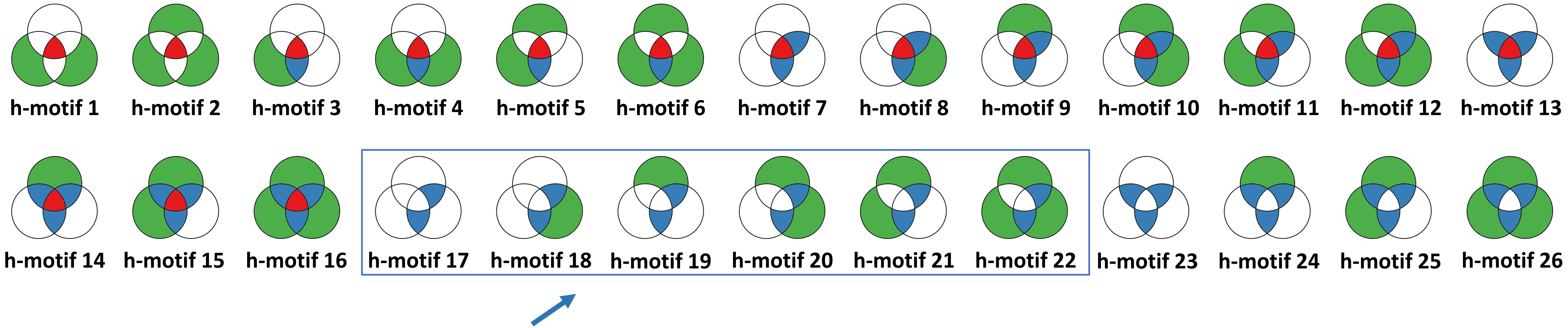
- Hypergraph motifs (h-motifs) describe connectivity patterns of three connected hyperedges.
- H-motifs describe the connectivity pattern of hyperedges e_i, e_j, and e_k by the emptiness of seven subsets.



(1) $e_i \backslash e_j \backslash e_k$
(2) $e_j \backslash e_k \backslash e_i$
(3) $e_k \backslash e_i \backslash e_j$
(4) $e_i \cap e_j \backslash e_k$
(5) $e_j \cap e_k \backslash e_i$
(6) $e_k \cap e_i \backslash e_j$
(7) $e_i \cap e_j \cap e_k$

Hypergraph Motifs: Definition (cont.)

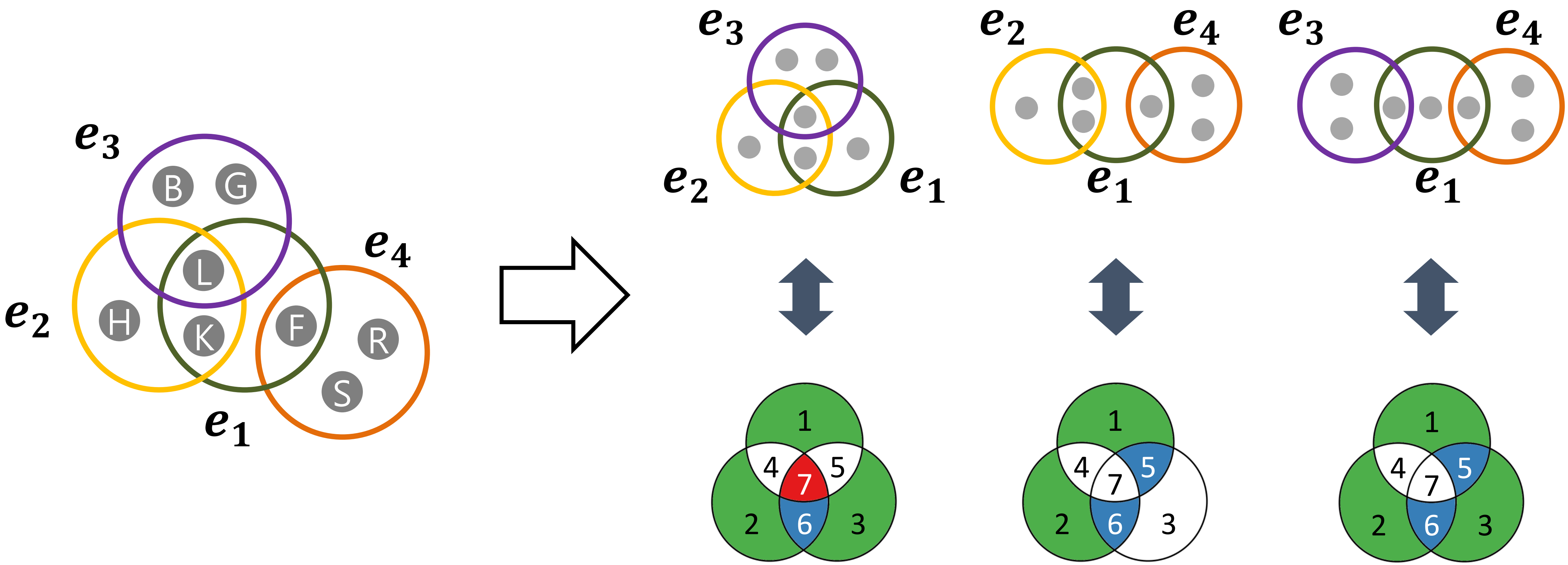
- While there can exist 2^7 h-motifs, 26 h-motifs remain once we exclude:
  1. symmetric ones
  2. those with duplicated hyperedges
  3. those cannot be obtained from connected hyperedges.



Open h-motifs contain two non-adjacent hyperedges. Others are **closed h-motifs**.

Hypergraph Motifs: Example

- Hypergraph motifs (h-motifs) describe connectivity patterns of three connected hyperedges.



Hypergraph Motifs: Properties

Exhaustive

H-motifs 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

H-motifs capture connectivity patterns **independently of the sizes of hyperedges**.

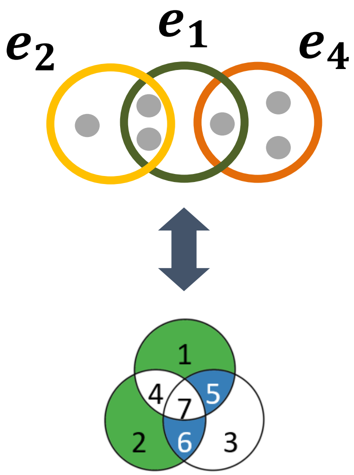
Hypergraph Motifs: Properties

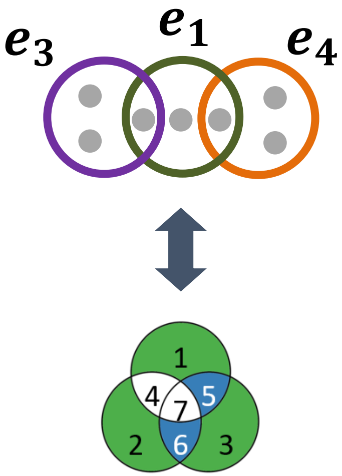
Question:

Why are **non-pairwise relations** considered?

Answer:

Non-pairwise relations play a key role in capturing the local structural patterns of real-world hypergraphs.





$\{e_1, e_2, e_4\}$ and $\{e_1, e_3, e_4\}$ have same pairwise relations, while their connectivity patterns are distinguished by h-motifs.

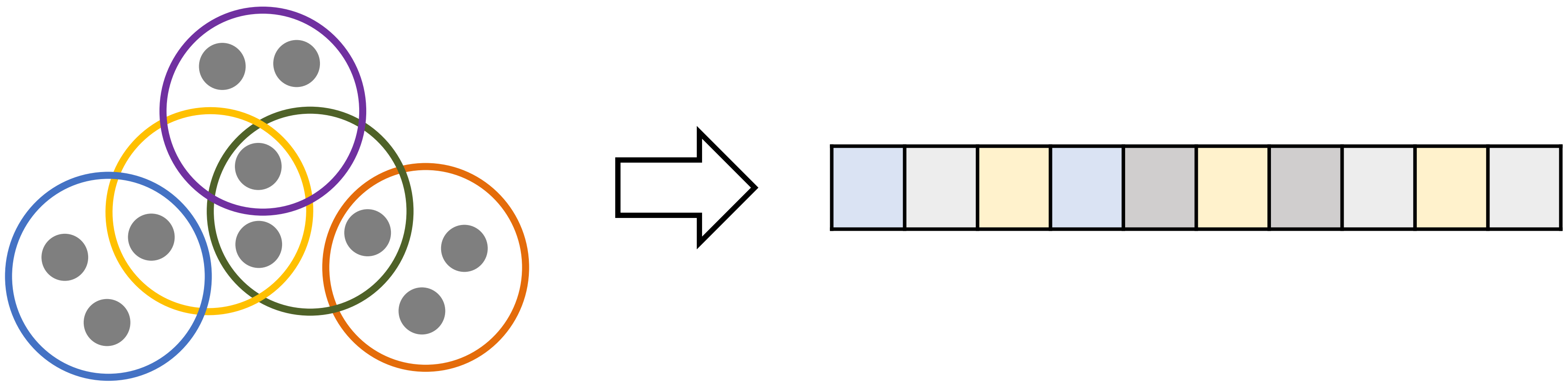
Characteristic Profiles

Question:

How can we **summarize** structural properties of hypergraphs?

Answer:

We compute a compact vector of **normalized significance of every h-motif**.



Characteristic Profiles (cont.)

- Significance of h-motifs

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

# of instances of h-motif t in the given hypergraph

# of instances of h-motif t in randomized hypergraphs

- Characteristic Profiles (CPs)

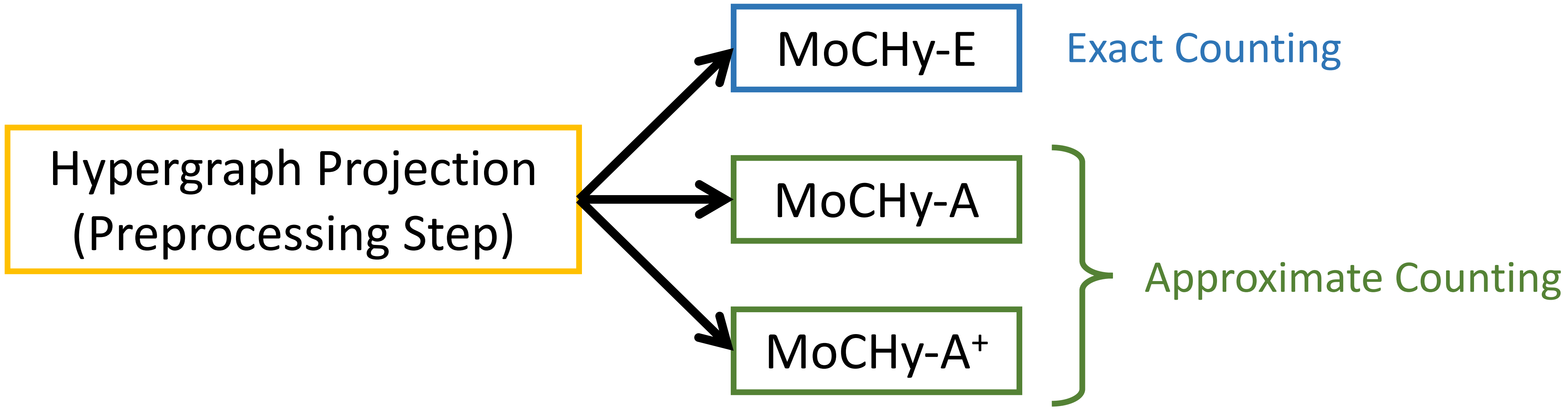
$$CP_t \coloneqq \frac{\Delta_t}{\sqrt{\sum_{t=1}^{26} \Delta_t^2}}$$

Roadmap

- Hypergraph Motif
- Proposed Method: MoCHy
- Experimental Results
- Conclusions

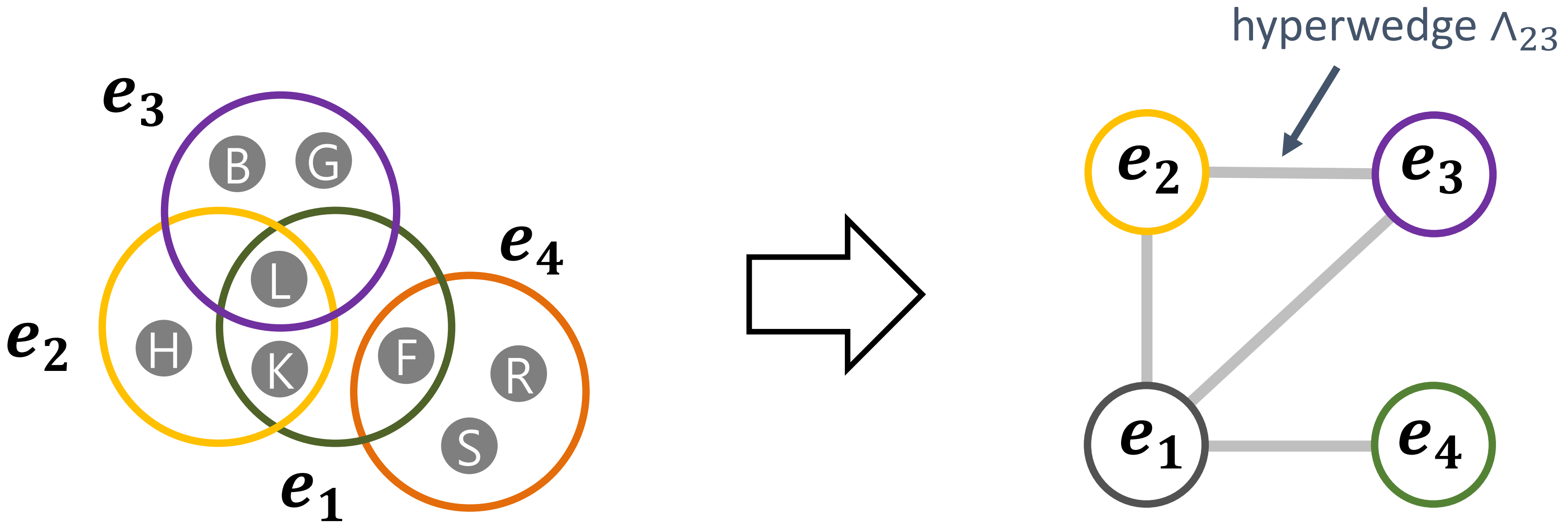
MoCHy: **Mo**tif **C**ounting in **Hy**pergraphs

- Given a hypergraph, how can we count the instances of each h-motif?
- We present MoCHy (**Mo**tif **C**ounting in **Hy**pergraphs), a family of parallel algorithms for counting the instances of h-motifs.



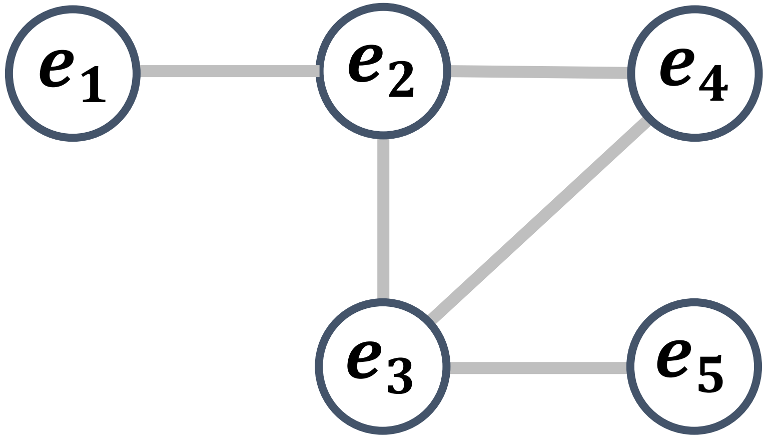
MoCHy: Hypergraph Projection

- Every version of MoCHy builds the projected graph $\bar{G} = (E, \wedge, \omega)$ of the input hypergraph $G = (V, E)$.
- To find the neighbors of each hyperedge e_i, find hyperedge e_j that contains any node $v \in e_i$.



MoCHy-E: Exact Counting

- **Enumerate** all three connected hyperedges in the hypergraph.
- **Increment** the count of h-motif corresponding to each instance.



$\{e_1, e_2, e_3\}$	$M[h(\{e_1, e_2, e_3\})] \uparrow$
$\{e_1, e_2, e_4\}$	$M[h(\{e_1, e_2, e_4\})] \uparrow$
$\{e_2, e_3, e_4\}$	$M[h(\{e_2, e_3, e_4\})] \uparrow$
$\{e_2, e_3, e_5\}$	$M[h(\{e_2, e_3, e_5\})] \uparrow$
$\{e_3, e_4, e_5\}$	$M[h(\{e_3, e_4, e_5\})] \uparrow$

Definition

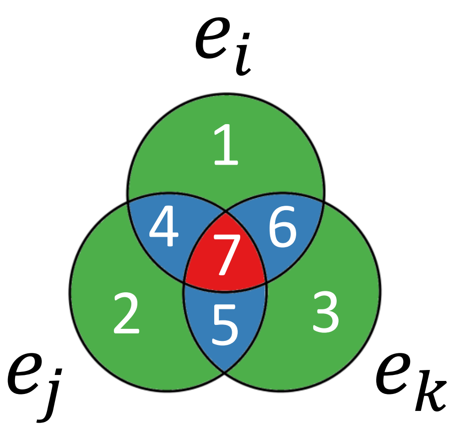
$\boldsymbol{h}(\{\boldsymbol{e_i}, \boldsymbol{e_j}, \boldsymbol{e_k}\})$: h-motif corresponding to an instance $\{e_i, e_j, e_k\}$

$\boldsymbol{M}[\boldsymbol{t}]$: count of h-motif t's instances

MoCHy-E: Exact Counting

Question:

How to compute $h(\{e_i, e_j, e_k\})$?



Answer:

Check the emptiness of seven sets by computing the cardinalities of each set ($O(\min(|e_i|, |e_j|, |e_k|))$).

(1) $|e_i \backslash e_j \backslash e_k| = |e_i| - |e_i \cap e_j| - |e_k \cap e_i| + |e_i \cap e_j \cap e_j|$

(2) $|e_j \backslash e_k \backslash e_i| = |e_j| - |e_i \cap e_j| - |e_j \cap e_k| + |e_i \cap e_j \cap e_j|$

(3) $|e_k \backslash e_i \backslash e_j| = |e_k| - |e_k \cap e_i| - |e_j \cap e_k| + |e_i \cap e_j \cap e_j|$

(4) $|e_i \cap e_j \backslash e_k| = |e_i \cap e_j| - |e_i \cap e_j \cap e_j|$

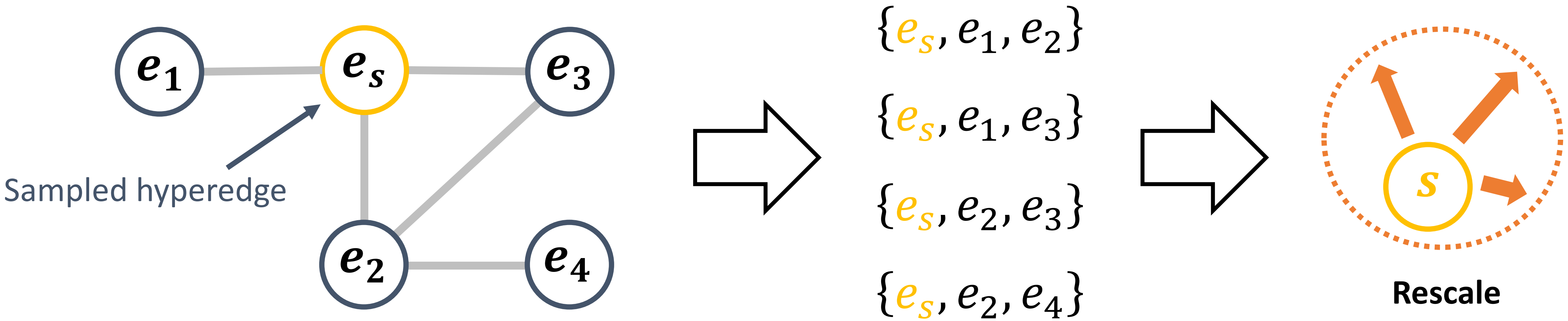
(5) $|e_j \cap e_k \backslash e_i| = |e_j \cap e_k| - |e_i \cap e_j \cap e_j|$

(6) $|e_k \cap e_i \backslash e_j| = |e_k \cap e_i| - |e_i \cap e_j \cap e_j|$

(7) $|e_i \backslash e_j \backslash e_k|$

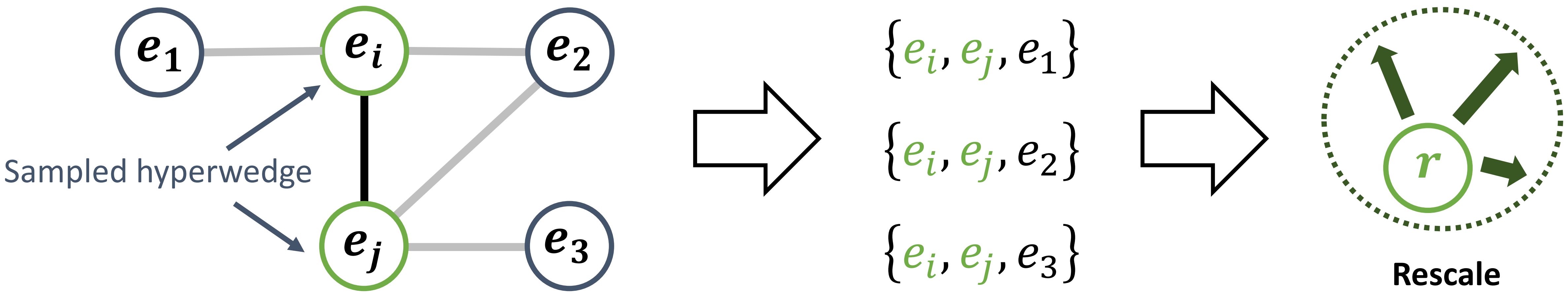
MoCHy-A: Hyperedge Sampling

- Sample s hyperedges from the hyperedge set E uniformly at random.
- For each sampled hyperedge e_s, count the number of instances of each h-motif t that contains e_s.
- Rescale the total approximate counts based on the sample size.



MoCHy-A$^+$: Hyperwedge Sampling

- Sample r hyperwedges from the hyperwedge set $\wedge$ uniformly at random.
- For each sampled hyperwedge $\wedge_{ij} = \{e_i, e_j\}$, count the number of instances of each h-motif t that contains $\wedge_{ij}$.
- Rescale the total approximate counts based on the sample size.



Roadmap

- Hypergraph Motif
- Proposed Method: MoCHy
- Experimental Results
- Conclusions

Experimental Settings

- 11 real-world hypergraphs from 5 different domains

Co-authorship | E-mail | Contact | Tags | Threads

- All versions of MoCHy implemented using C++ and OpenMP.

&

for parallelization

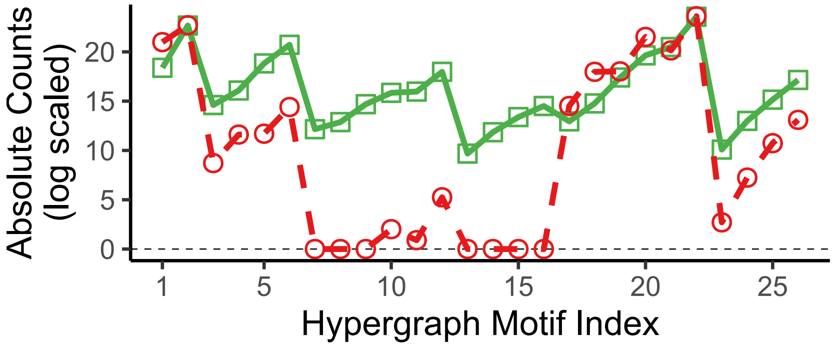
EXP1. Comparison with Random

- Real-world and random hypergraphs have distinct distributions of h-motif instances.

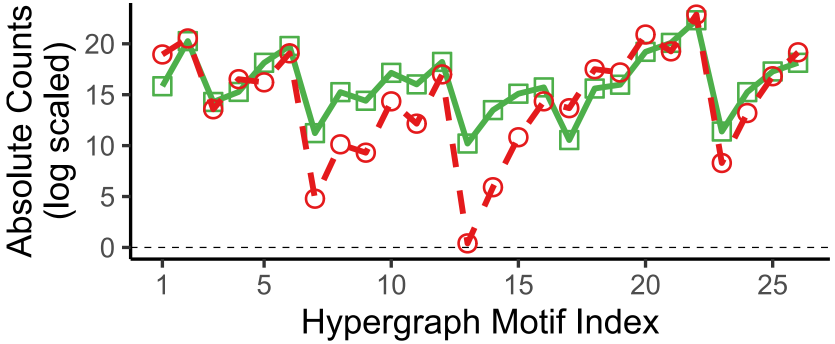
h-motif	coauth-DBLP				contact-primary				email-EU				tags-math				threads-math			
	count (rank)		RD	RC	count (rank)		RD	RC	count (rank)		RD	RC	count (rank)		RD	RC	count (rank)		RD	RC
	real	random			real	random			real	random			real	random			real	random		
1	9.6E07 (7)	1.3E09 (4)	3	-0.86	4.8E04 (16)	2.8E07 (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

EXP1. Comparison with Random

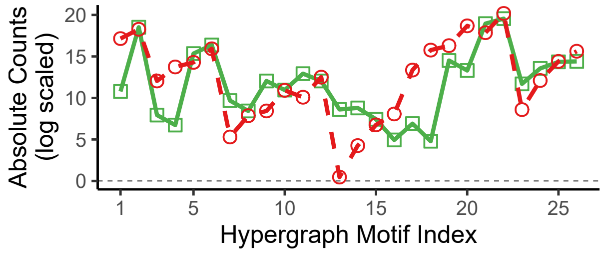
- Real-world and random hypergraphs have distinct distributions of h-motif instances.



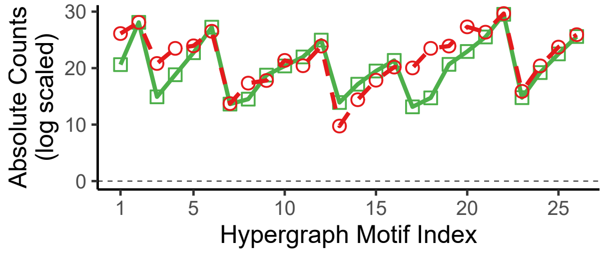
coauth-DBLP



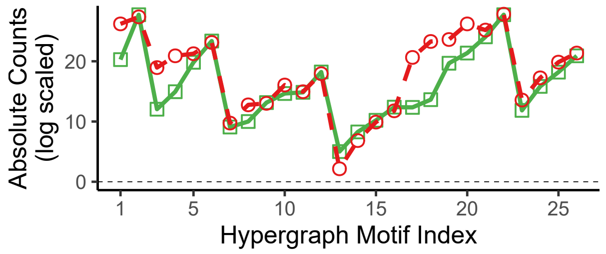
email-Eu



contact-primary



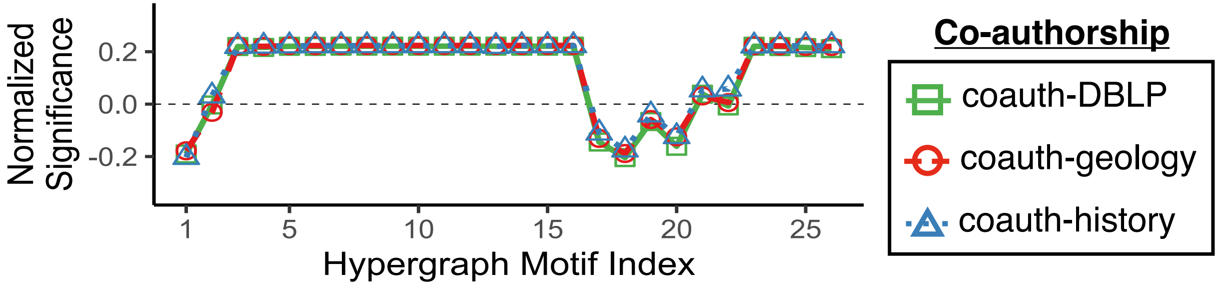
tags-math

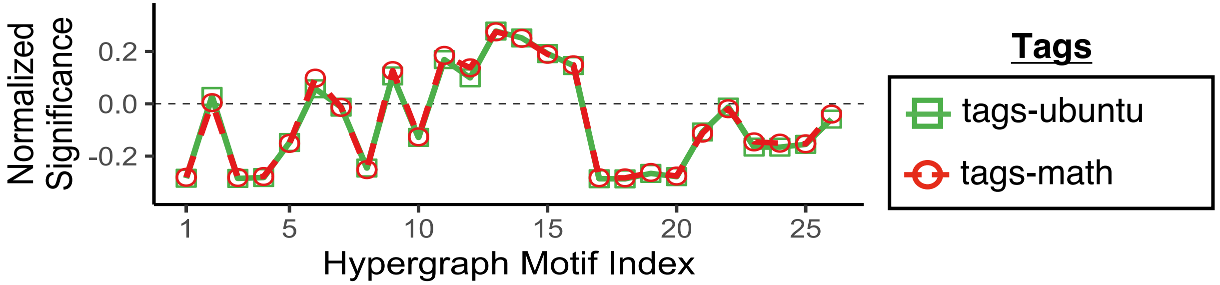


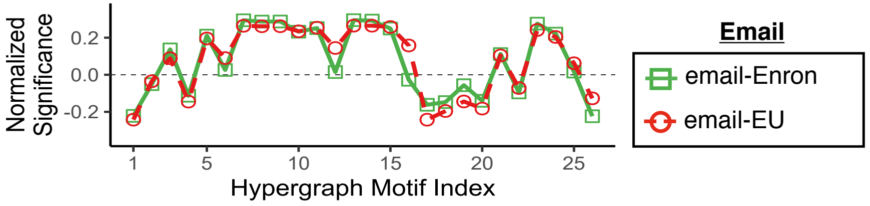
threads-math

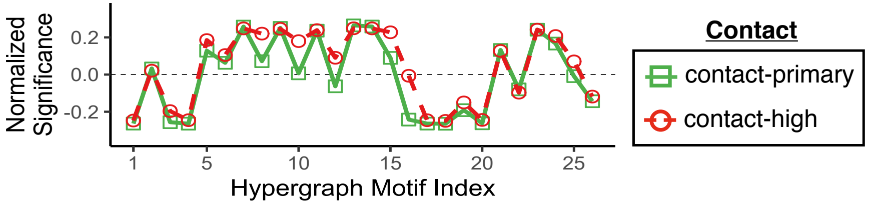
EXP2. Comparison across Domains

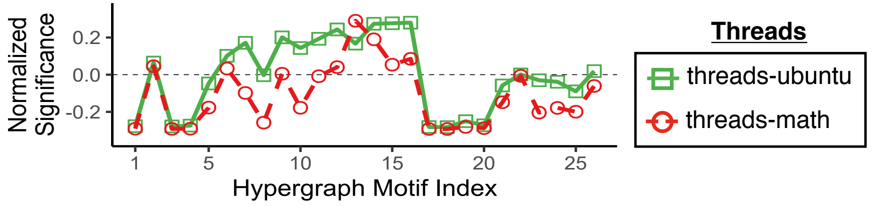
- The CPs are similar within domains but different across domains.





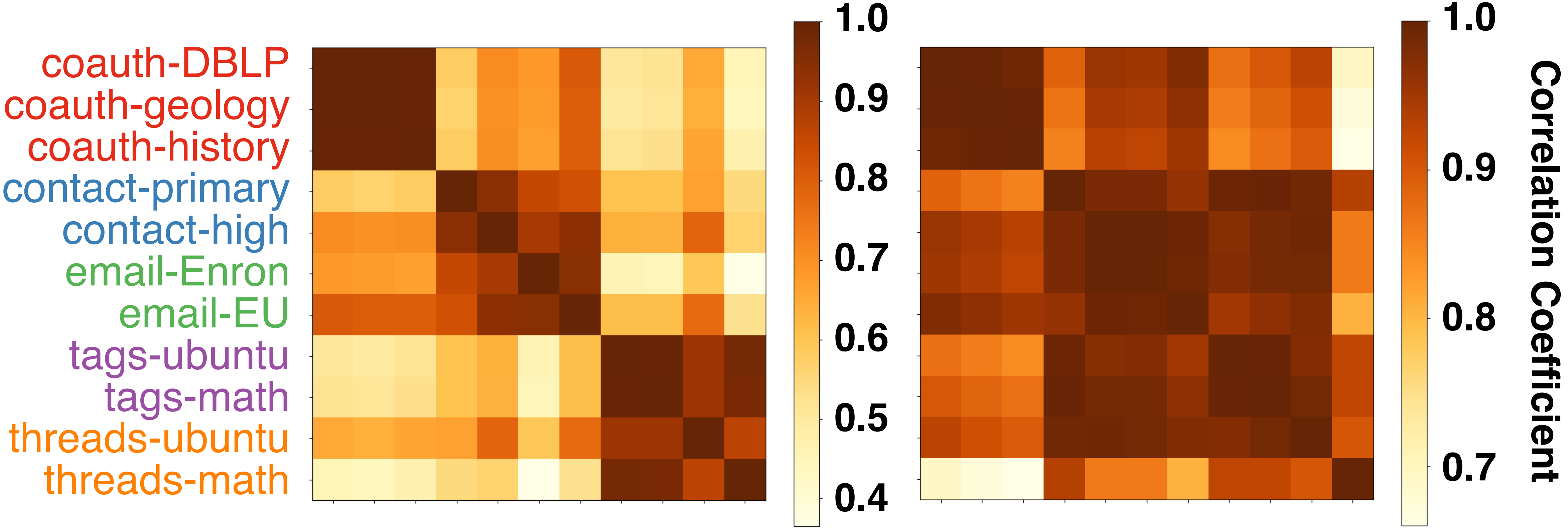






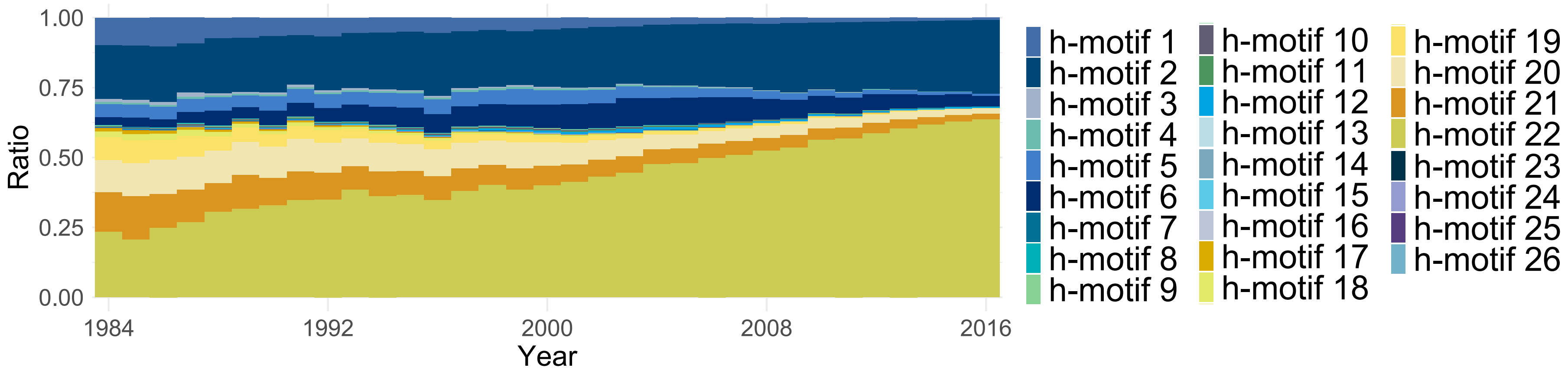
EXP2. Comparison across Domains (cont.)

- Characteristic profiles (CPs) based on hypergraph motifs (h-motifs) capture local structural patterns more accurately than CPs based on network motifs.

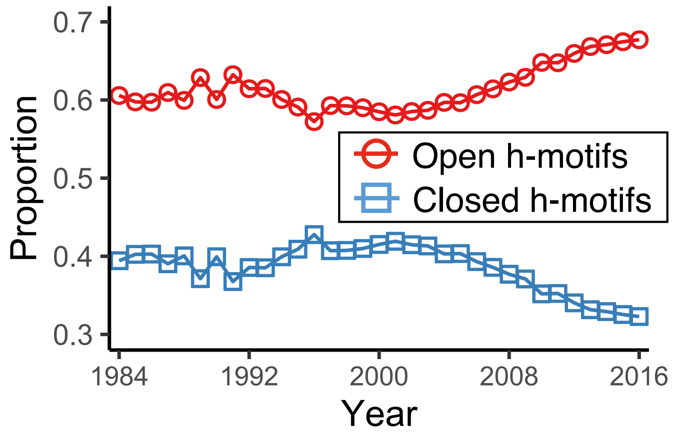


EXP3. Observations and Applications

- Trends in the formation of collaborations are captured by h-motifs.
  - (a) The fractions of the instances of h-motifs 2 and 22 have increased rapidly.
  - (b) The fraction of the instances of open h-motifs increased steadily since 2001.



(a) Fraction of the instances of each h-motif in the coauth-DBLP over time.



(b) Open and closed h-motifs.

EXP3. Observations and Applications (cont.)

- Hyperedge prediction: classify real hyperedges and fake ones.
- H-motifs give informative features.

		HM26	HM7	HC
Logistic Regression	**ACC**	0.754	0.656	0.636
	AUC	0.813	0.693	0.691
Random Forest	**ACC**	0.768	0.741	0.639
	AUC	0.852	0.779	0.692
Decision Tree	**ACC**	0.731	0.684	0.613
	AUC	0.732	0.685	0.616
K-Nearest Neighbors	**ACC**	0.694	0.689	0.640
	AUC	0.750	0.743	0.684
MLP Classifier	**ACC**	0.795	0.762	0.646
	AUC	0.875	0.841	0.701

ACC: accuracy, AUC: area under the ROC curve

HM26 ($\in \mathbb{R}^{26}$)

The number of each h-motif's instances that contain each hyperedge.

HM7 ($\in \mathbb{R}^{7}$)

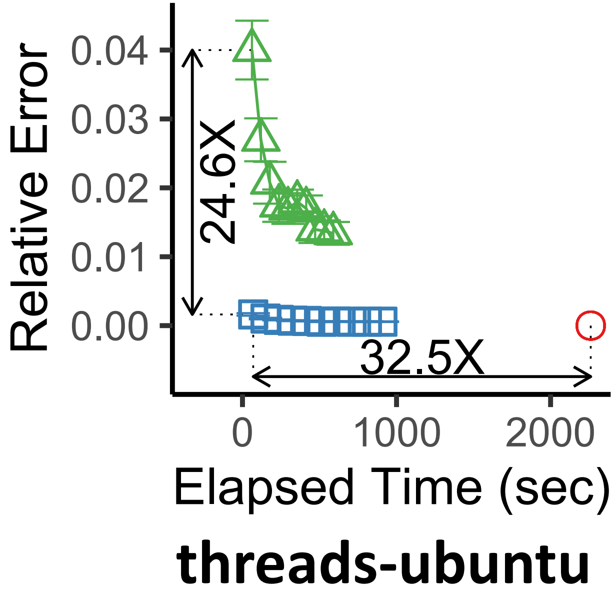
The seven features with the largest variance among those in HM26.

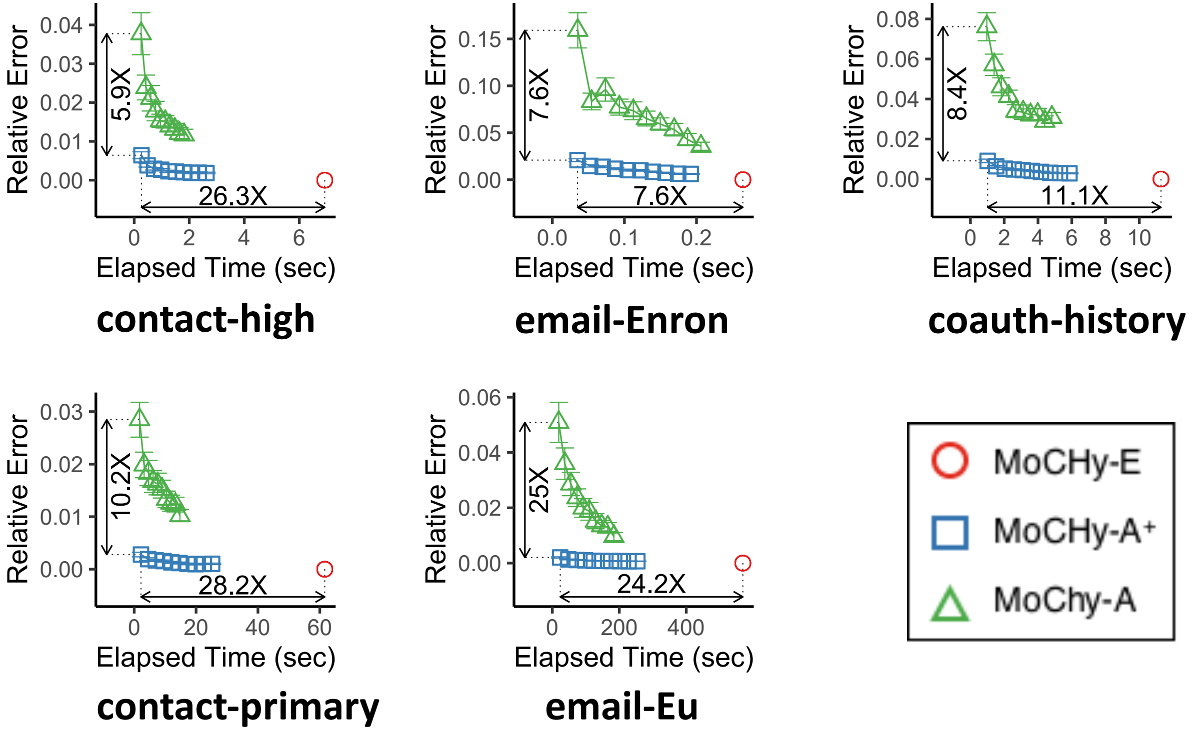
HC ($\in \mathbb{R}^{7}$)

The mean, maximum, and minimum degree and the mean, maximum, and minimum number of neighbors of the nodes in each hyperedge and its size.

EXP4. Performance of Counting Algorithms

- MoCHy-A$^+$ is up to 25X more accurate than MoCHy-A.
- MoCHy-A$^+$ is up to 32.5X faster than MoCHy-E.



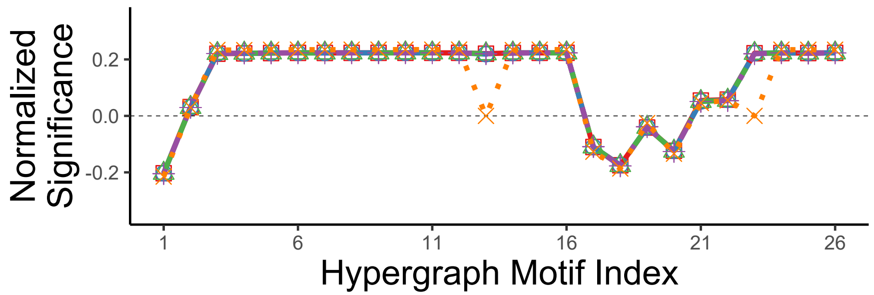


Definition

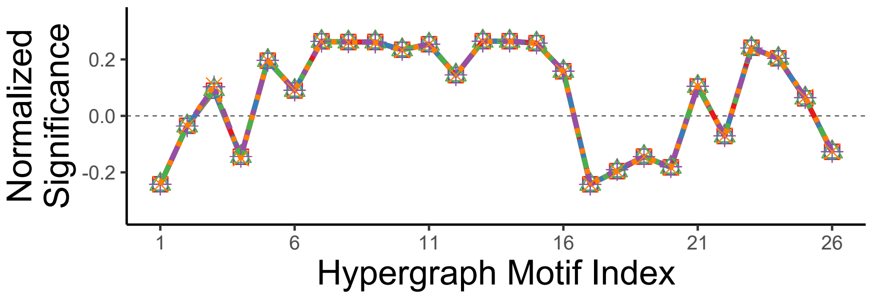
Relative Error: $\frac{\sum_{t=1}^{26}|M[t]-\bar{M}[t]|}{\sum_{t=1}^{26} M[t]}$

EXP4. Performance of Counting Algorithms (cont.)

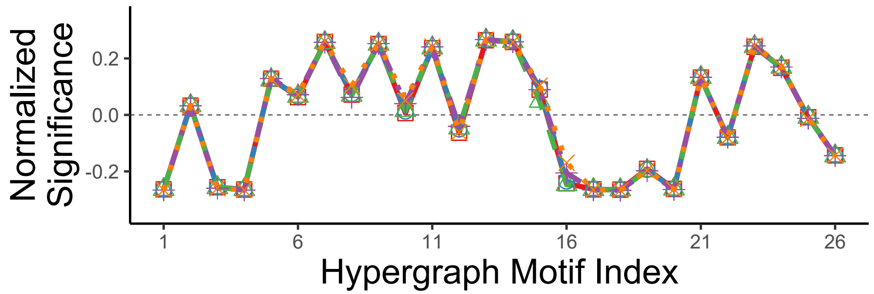
- CPs obtained by MoCHy-A^{+} are estimated near perfectly even with a smaller number of samples.



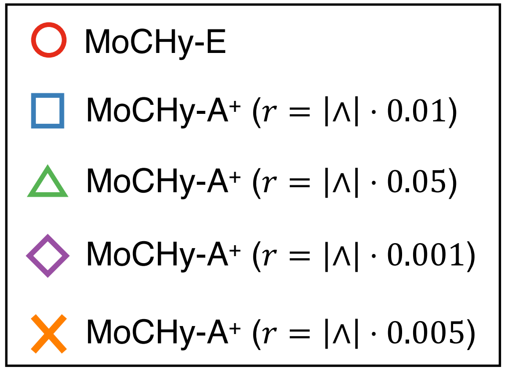
coauth-history



email-Eu

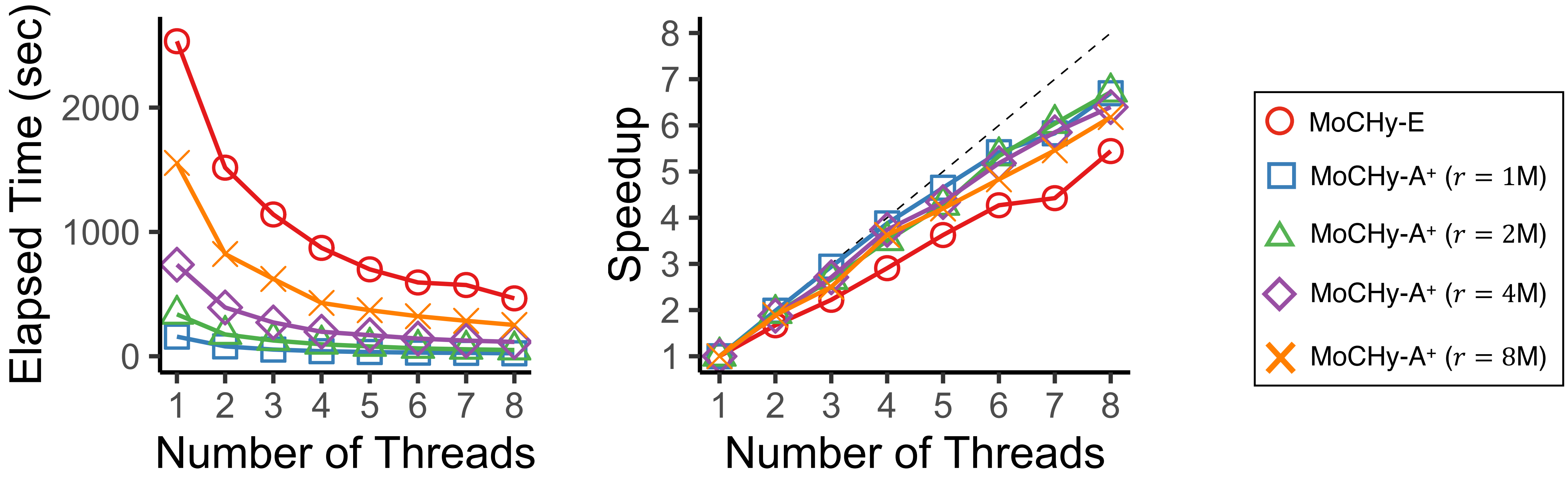


contact-primary



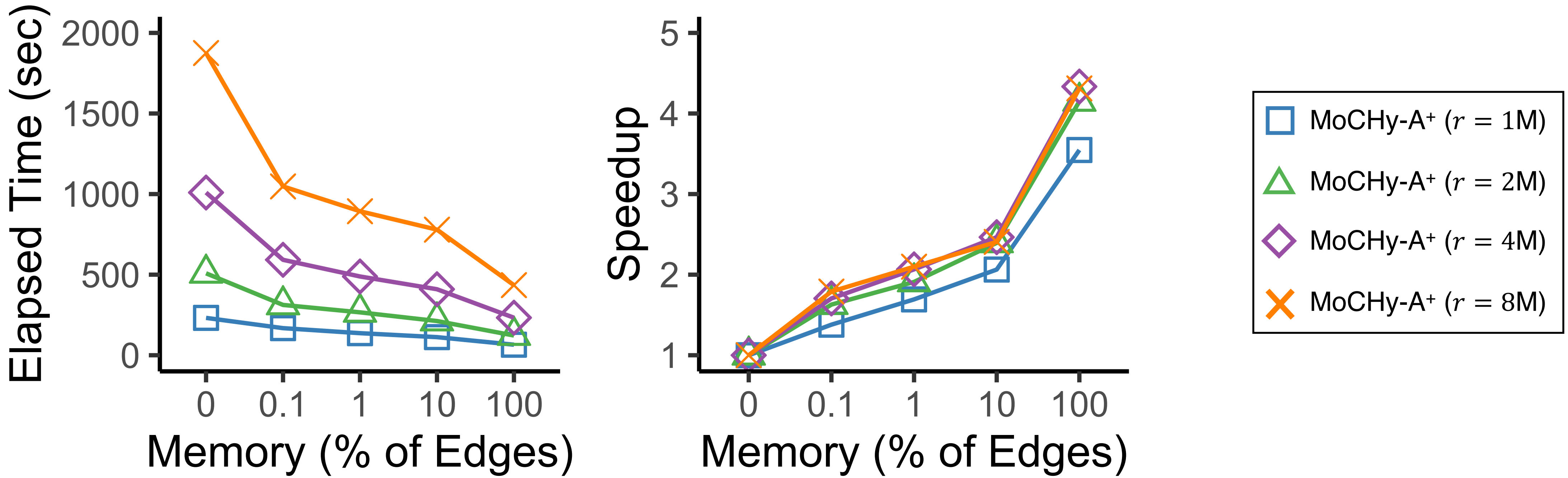
EXP4. Performance of Counting Algorithms (cont.)

- Both MoCHy-E and MoCHy-A^{+} achieve significant speedups with multiple threads.



EXP4. Performance of Counting Algorithms (cont.)

- Memoizing a small fraction of projected graphs leads to significant speedups of MoCHy-A$^+$.



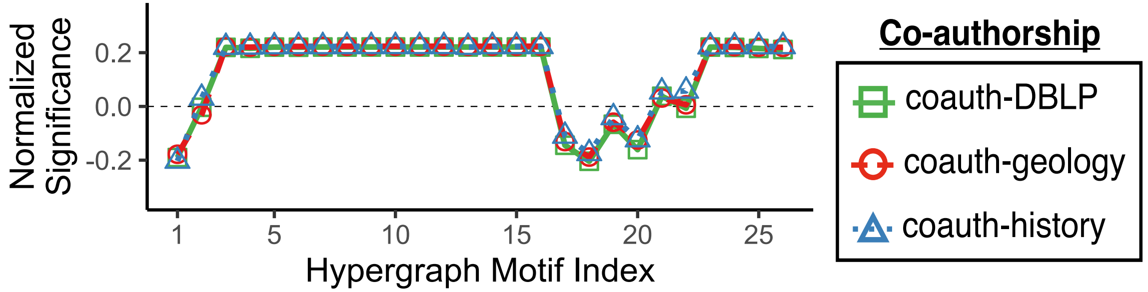
Roadmap

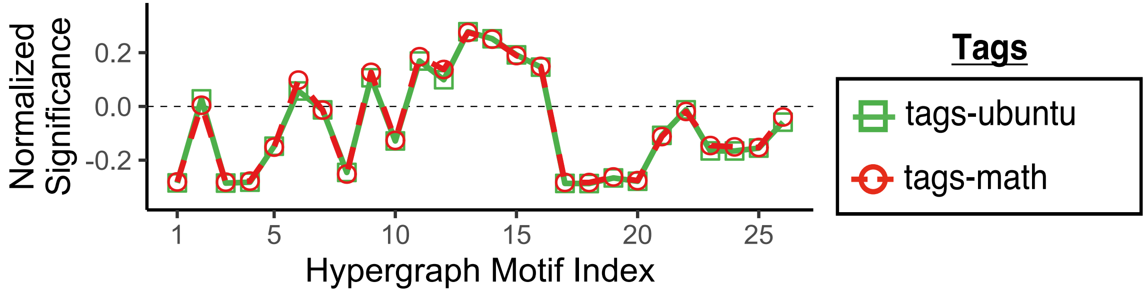
- Hypergraph Motif
- Proposed Method: MoCHy
- Experimental Results
- Conclusions

Conclusions

- We propose hypergraph motifs (h-motifs) for describing the connectivity patterns of hypergraphs.

- Novel Concepts
- Fast and Provable Algorithms
- Discoveries in 11 Real-world Hypergraphs



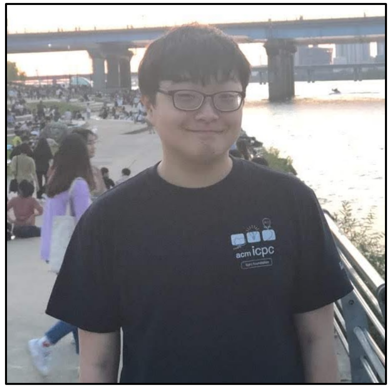


KAIST AI

VLDB2020 Tokyo

Hypergraph Motifs: Concepts, Algorithms, and Discoveries

Geon Lee



Jihoon Ko

Kijung Shin