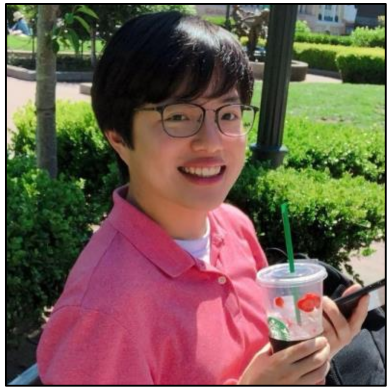
THyMe$^+$: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting

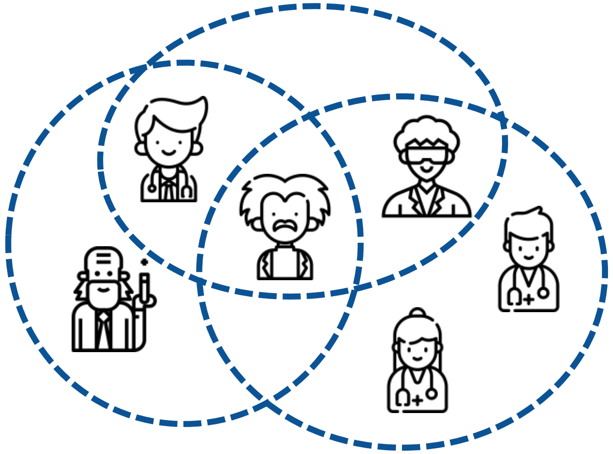
Geon Lee



Kijung Shin

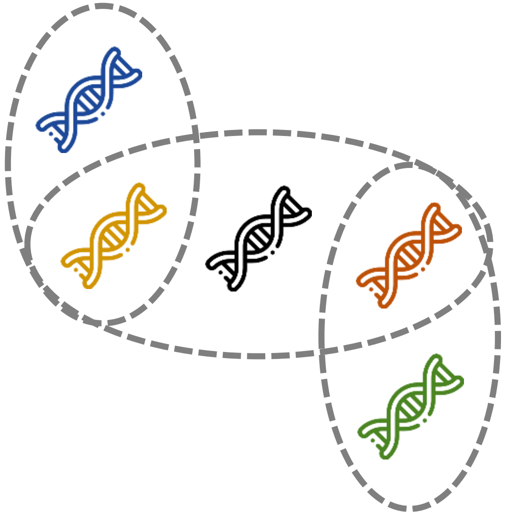
Hypergraphs are Everywhere

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



Collaborations of Researchers

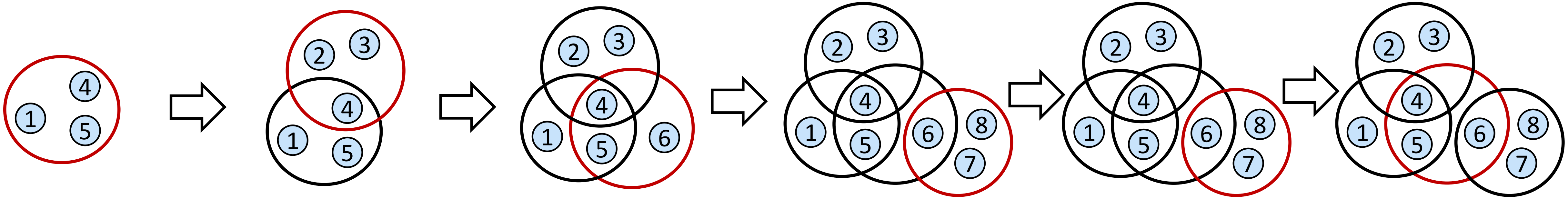
Co-purchases of Items



Joint Interactions of Proteins

Hypergraphs Evolve Over Time

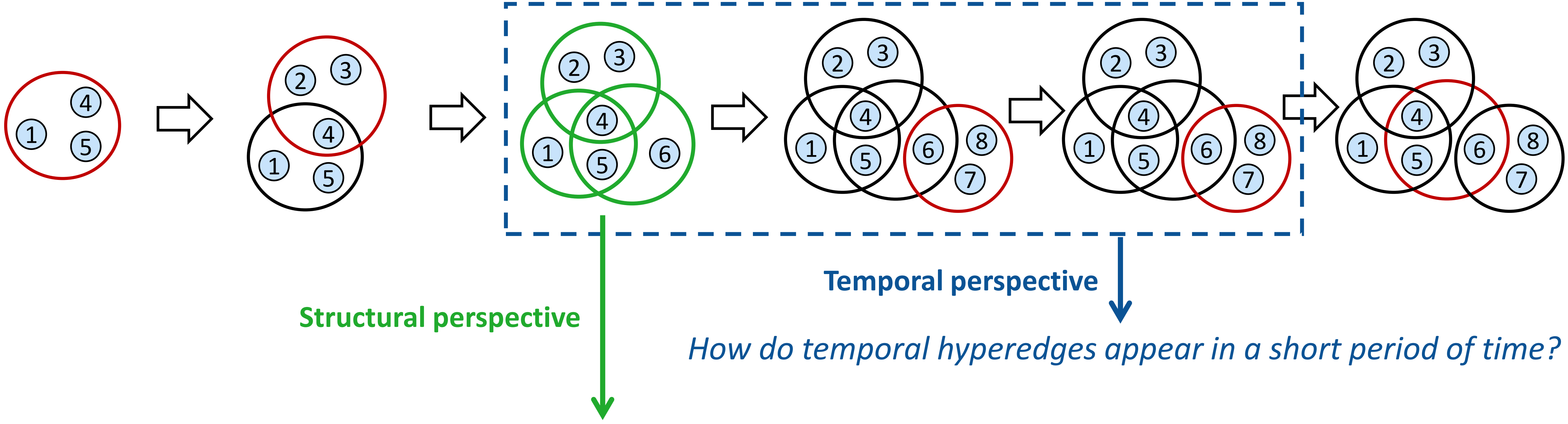
- In many real-world scenarios, hypergraphs **evolve over time**.
- **Temporal hypergraphs** consist of **temporal hyperedges**.



t_1=7 t_2=10 t_3=11 t_4=12 t_5=14 t_5=17

Our Question

What are **local structural & temporal properties** of real-world hypergraphs?



How do temporal hyperedges appear in a short period of time?

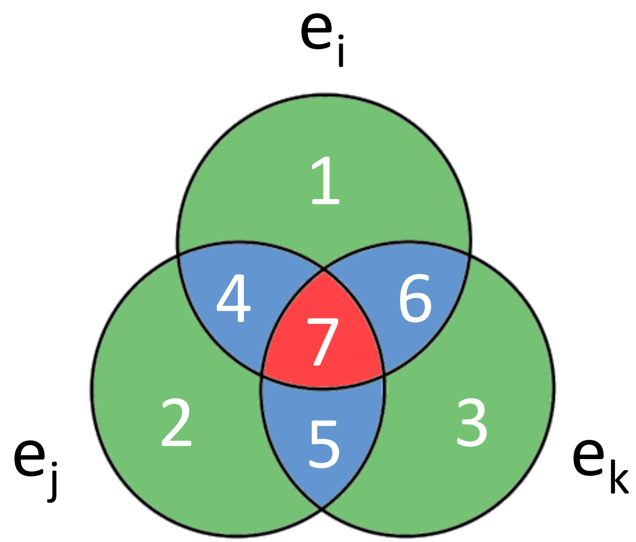
How do three hyperedges overlap each other?

Roadmap

1. **Backgrounds**
2. Concepts
3. Observations
4. Algorithms
5. Conclusion

Hypergraph Motifs

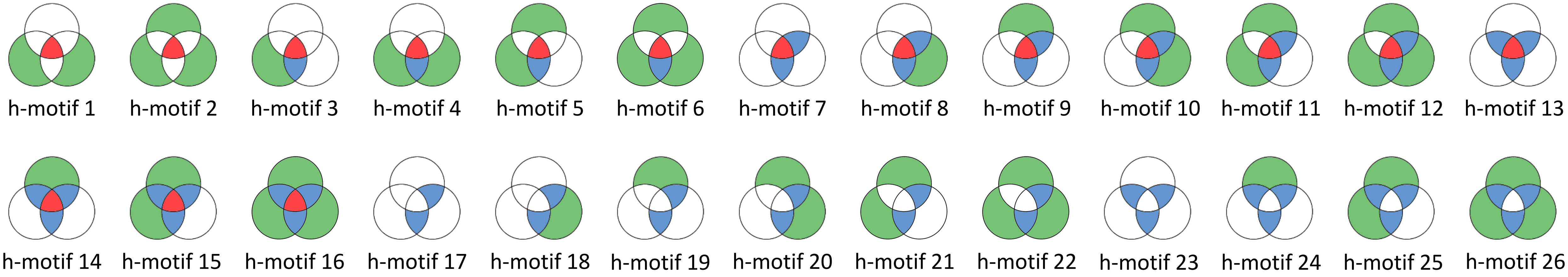
- Hypergraph motifs (h-motifs) describe connectivity patterns of three connected hyperedges in static hypergraphs.
- H-motifs describe the connectivity pattern of hyperedges e_i, e_j, and e_k by the emptiness of seven subsets.



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

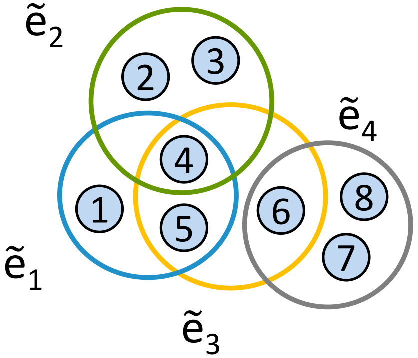
Hypergraph Motifs (cont.)

- 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

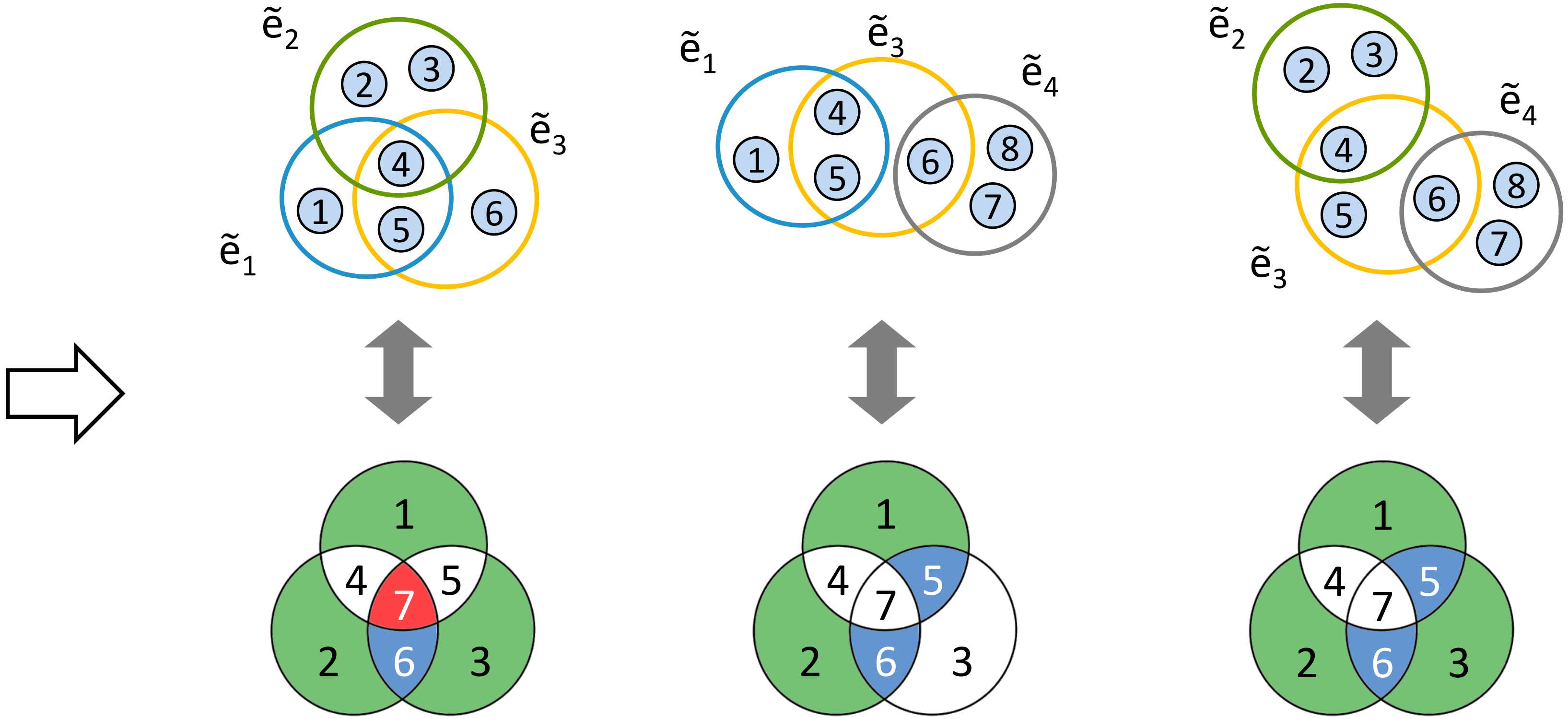


Hypergraph Motifs (cont.)

- Hypergraph motifs describe connectivity patterns of three connected hyperedges.
- For example:



Static Hypergraph



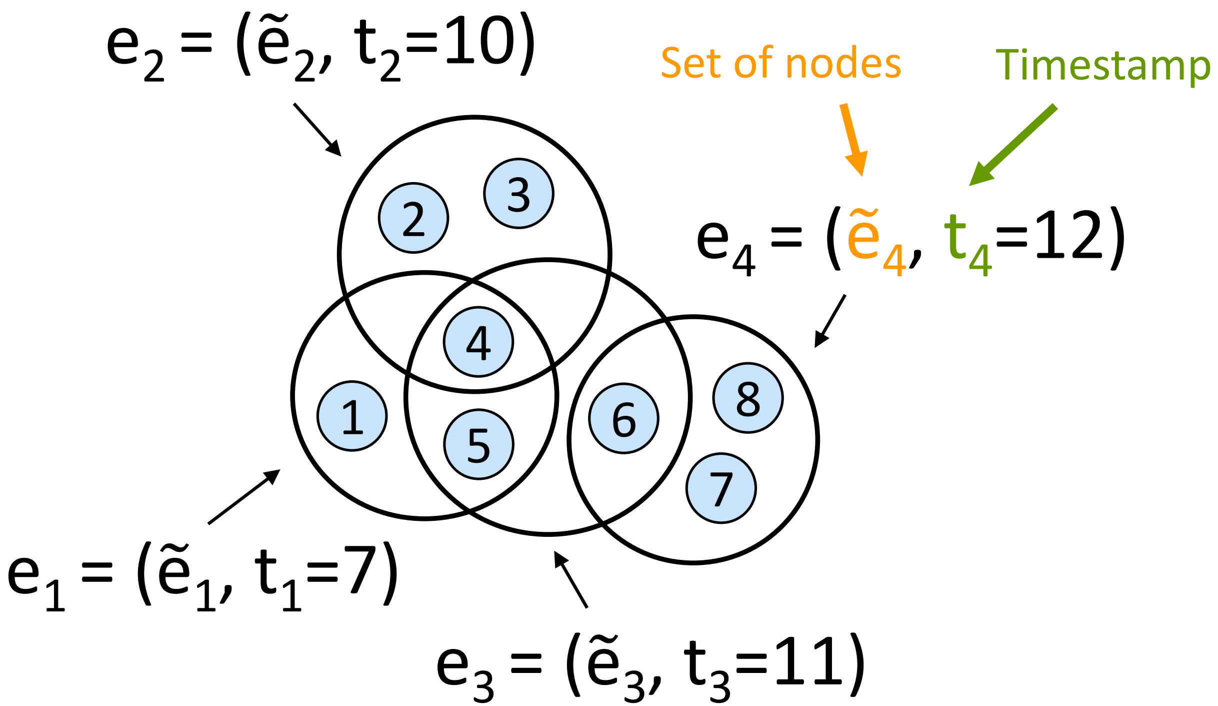
h-motif 6 h-motif 21 h-motif 22

Roadmap

1. Backgrounds
2. **Concepts**
3. Observations
4. Algorithms
5. Conclusion

Temporal Hypergraph Motifs

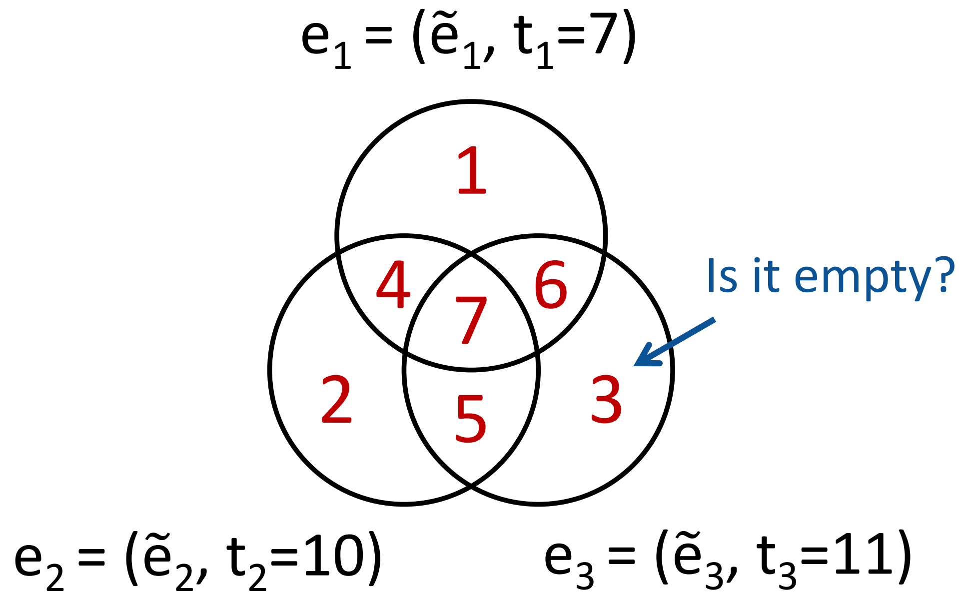
- How can we define *motifs* in temporal hypergraph?
- We define **temporal hypergraph motifs** (**TH-motifs**) that describe **structural** and **temporal** patterns in sequences of three connected temporal hyperedges.



Temporal Hypergraph Motifs (cont.)

Q1. How can we capture **structural properties** of hypergraphs?

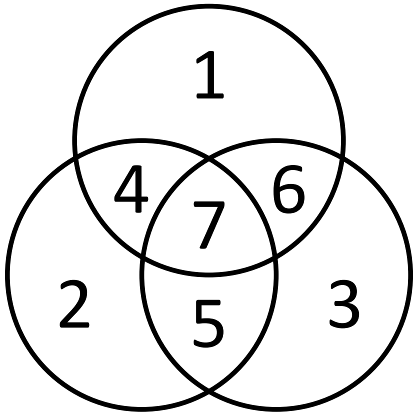
A1. We consider the **emptiness of seven subsets** of three temporal hyperedges.



Temporal Hypergraph Motifs (cont.)

Q2. How can we capture **temporal properties** of hypergraphs?

$e_1 = (\tilde{e}_1, t_1 = 7)$



$e_2 = (\tilde{e}_2, t_2 = 10)$ $e_3 = (\tilde{e}_3, t_3 = 11)$

A2-1. The three temporal hyperedges should **arrive within δ time**.

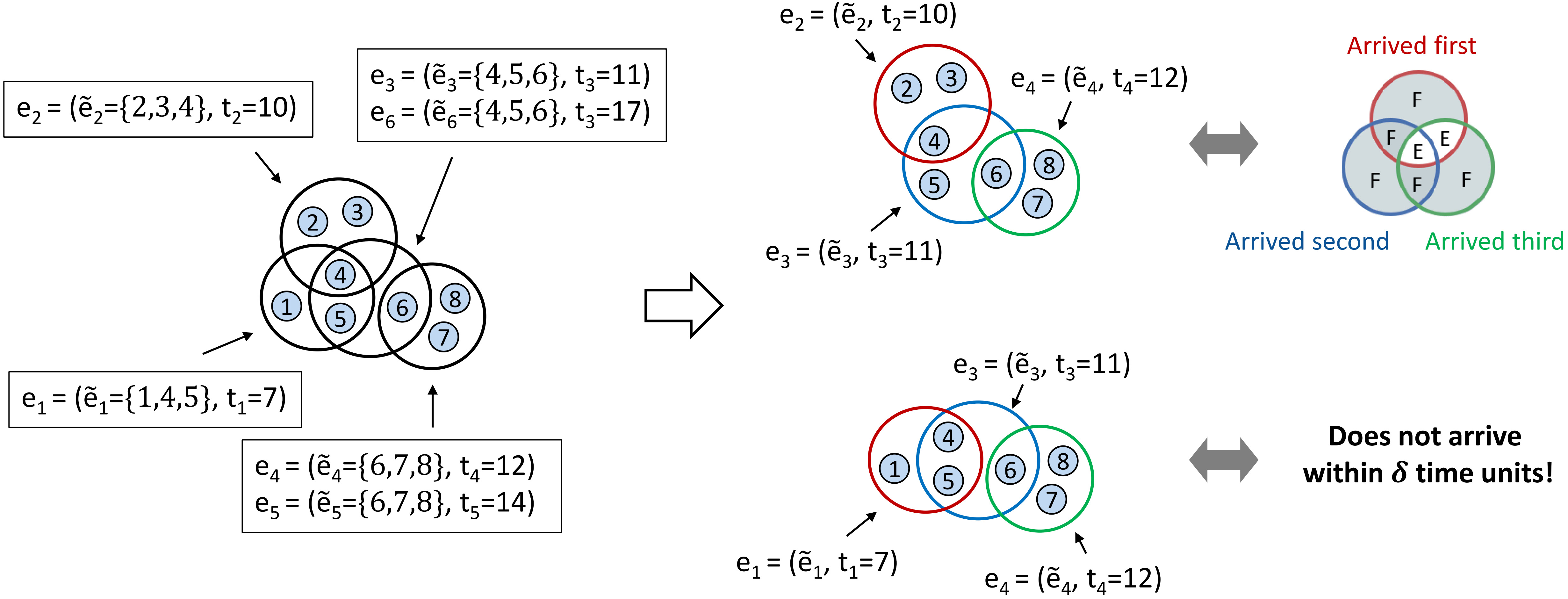
$$\max(t_1, t_2, t_3) - \min(t_1, t_2, t_3) \leq \delta$$

A2-2. The **order** of the three temporal hyperedges is considered.

$$e_1 \rightarrow e_2 \rightarrow e_3$$

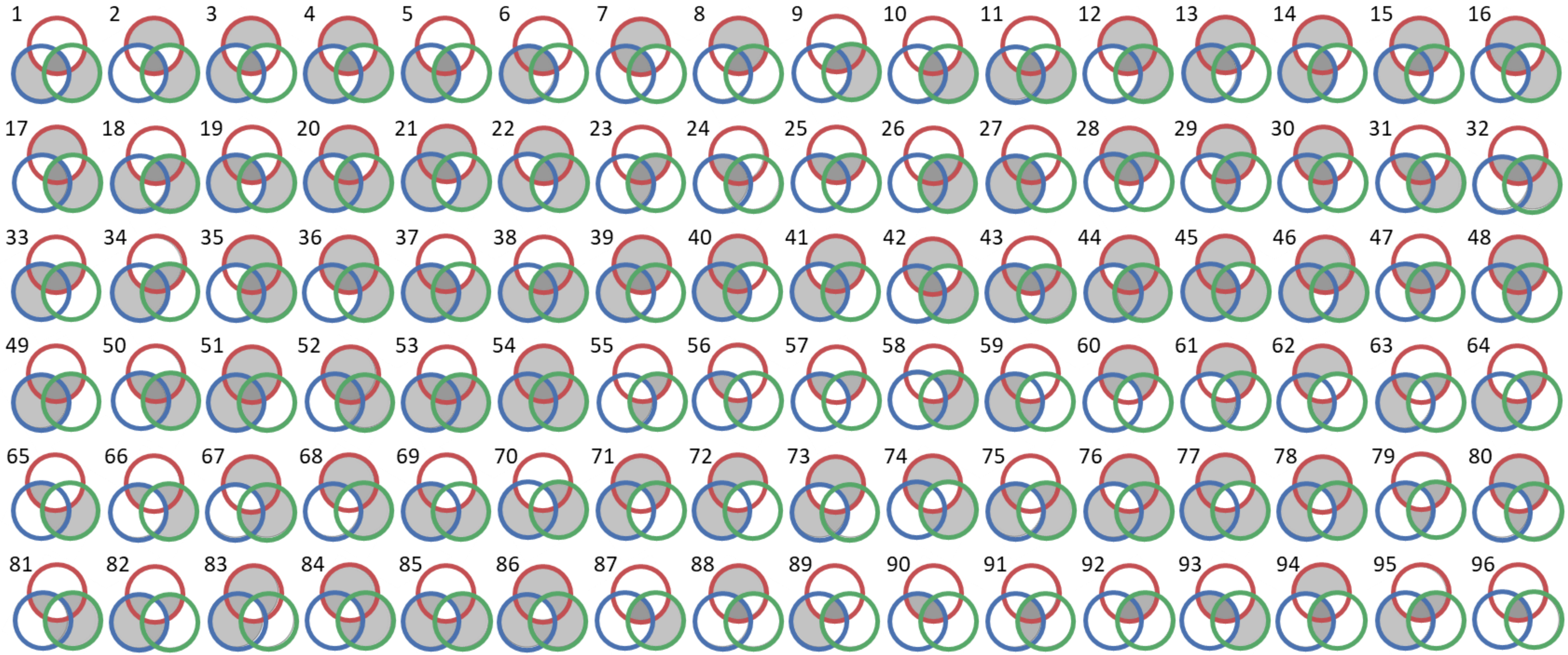
Temporal Hypergraph Motifs (cont.)

- For example, let $\boldsymbol{\delta} = \mathbf{3}$.



Temporal Hypergraph Motifs (cont.)

- We define **96 temporal hypergraph motifs** (**TH-motifs**).

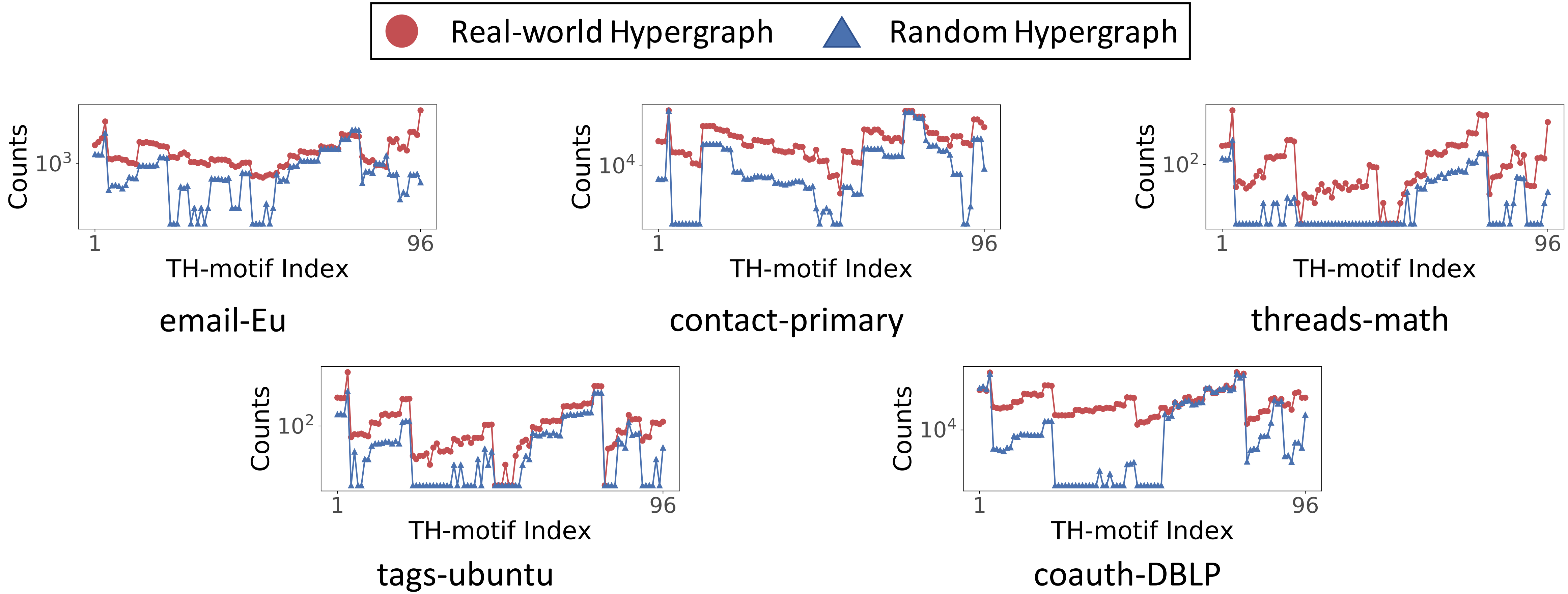


Roadmap

1. Backgrounds
2. Concepts
3. **Observations**
4. Algorithms
5. Conclusion

Observations: Real Hypergraphs are Not Random

Obs1. Real hypergraphs are clearly distinguished from randomized hypergraphs.



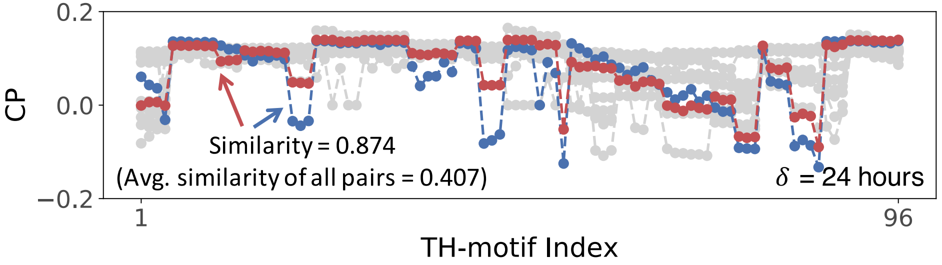
Observations: TH-motifs Distinguish Domains

Obs2. TH-motifs play a key role in capturing structural & temporal patterns.

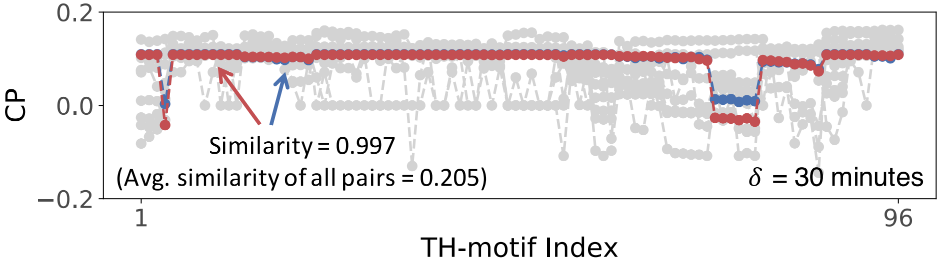
Characteristic Profile (CP): Relative significance of each TH-motif.

$$CP_t := \frac{\Delta_t}{\sqrt{\sum_{t=1}^{96} \Delta_t^2}} \quad \text{where} \quad \Delta_t := \frac{M[t] - M_{rand}[t]}{M[t] + M_{rand}[t] + \epsilon}$$

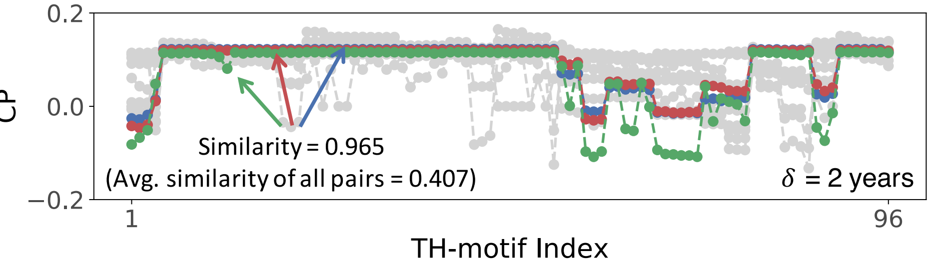
Email domain



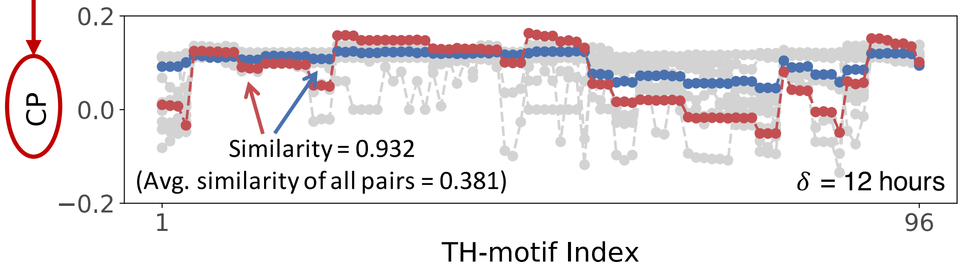
Contact domain



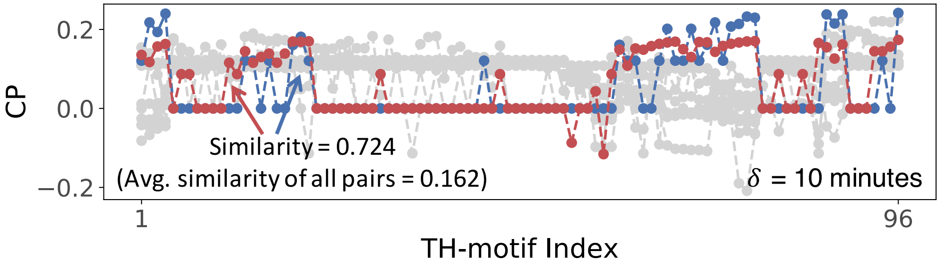
Coauthorship domain



Tags domain

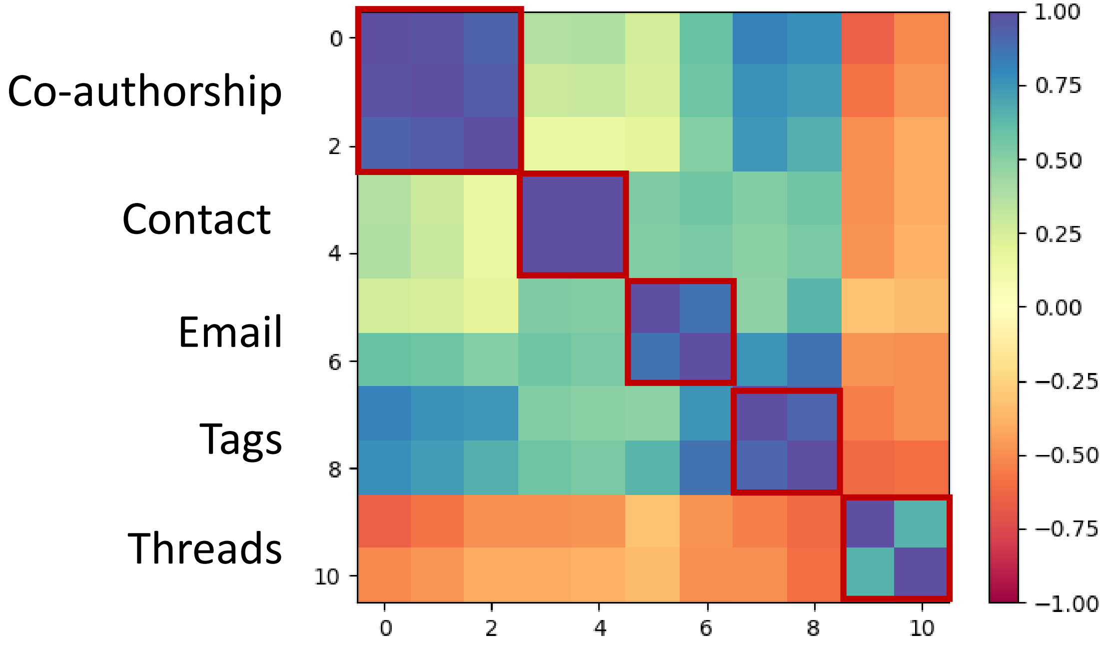


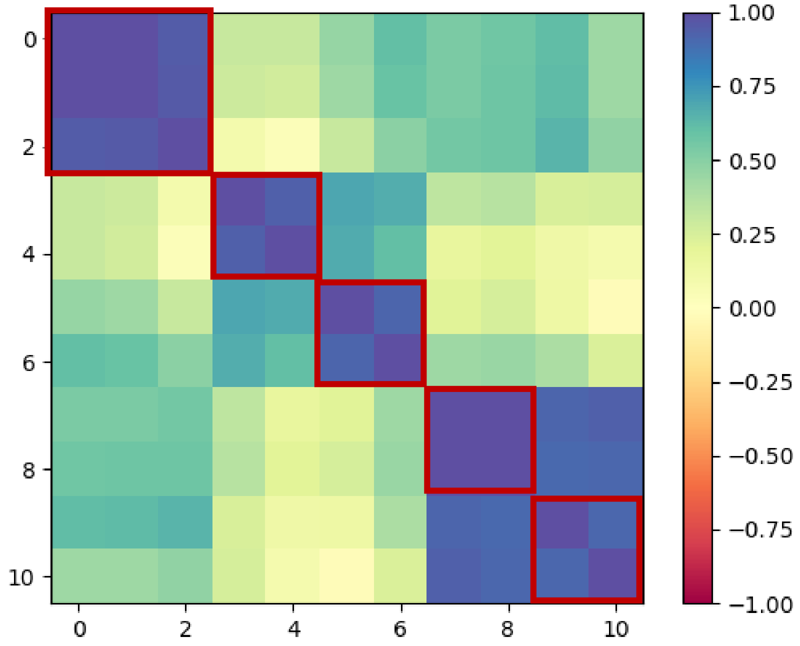
Threads domain



Observations: TH-motifs Distinguish Domains (cont.)

Obs2. TH-motifs play a key role in capturing structural & temporal patterns.





Temporal Hypergraph Motif

- Within-domain: 0.900
- Between-domain: 0.141

Gap: **0.759**
Times: **6.38X**

Static Hypergraph Motif

- Within-domain: 0.951
- Between-domain: 0.434

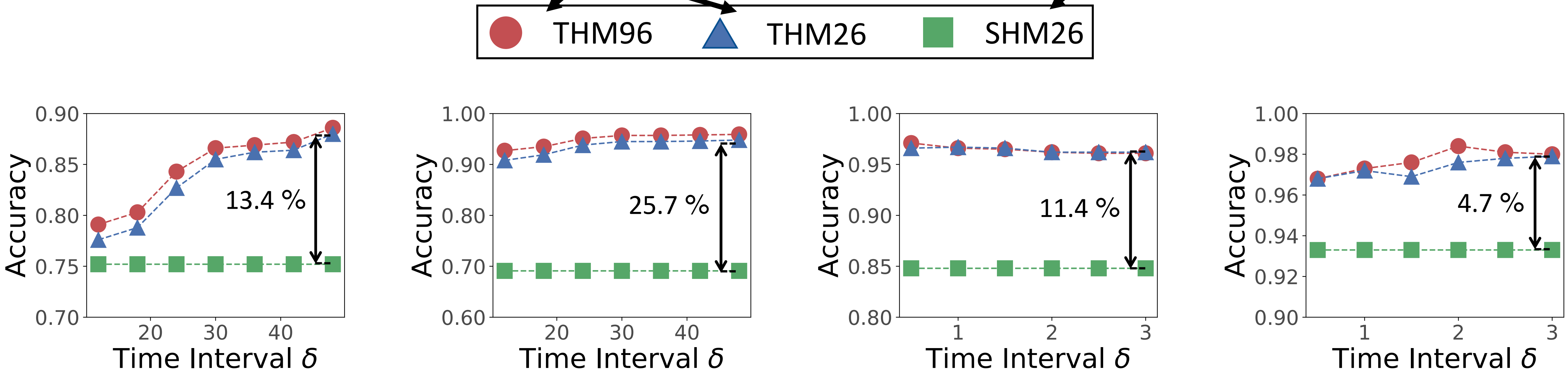
Gap: 0.517
Times: 2.19X

Observations: TH-motifs Help Predict Future Hyperedges

Obs3. TH-motifs can be used as powerful features for predicting future hyperedges.

\# of each **TH-motifs**' instances that each hyperedge is contained.

\# of each **static h-motifs**' instances that each hyperedge is contained.



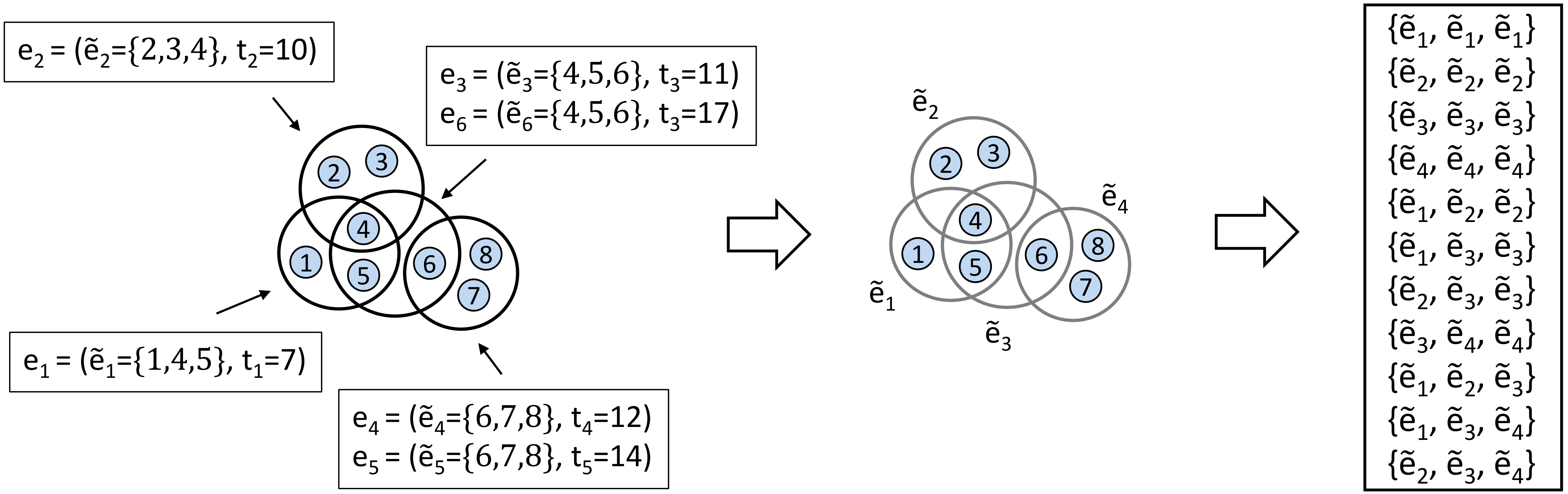
email-Enron | email-Eu | contact-primary | contact-high

Roadmap

1. Backgrounds
2. Concepts
3. Observations
4. **Algorithms**
5. Conclusion

DP: Naïve Approach Using Dynamic Programming

- DP enumerates the instances of static h-motifs in the induced static hypergraph.



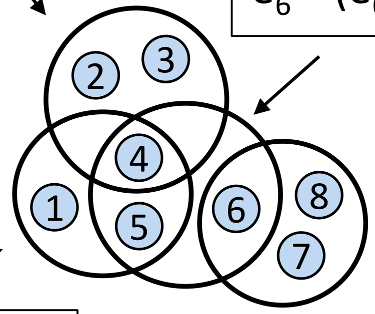
Temporal hypergraph | Induced static hypergraph | Instances of static h-motifs ($\leq$ 3 hyperedges)

DP: Naïve Approach Using Dynamic Programming (cont.)

- DP counts the instances of TH-motifs from instances of static h-motifs using **dynamic programming**.

e_2 = ($\tilde{e}_2$={2,3,4}, t_2=10)

e_3 = ($\tilde{e}_3$={4,5,6}, t_3=11)
e_6 = ($\tilde{e}_6$={4,5,6}, t_3=17)



e_1 = ($\tilde{e}_1$={1,4,5}, t_1=7)

e_4 = ($\tilde{e}_4$={6,7,8}, t_4=12)
e_5 = ($\tilde{e}_5$={6,7,8}, t_5=14)

{$\tilde{e}_1$, $\tilde{e}_1$, $\tilde{e}_1$}
{$\tilde{e}_2$, $\tilde{e}_2$, $\tilde{e}_2$}
{$\tilde{e}_3$, $\tilde{e}_3$, $\tilde{e}_3$}
{$\tilde{e}_4$, $\tilde{e}_4$, $\tilde{e}_4$}
{$\tilde{e}_1$, $\tilde{e}_2$, $\tilde{e}_2$}
{$\tilde{e}_1$, $\tilde{e}_3$, $\tilde{e}_3$}
{$\tilde{e}_2$, $\tilde{e}_3$, $\tilde{e}_3$}
{$\tilde{e}_3$, $\tilde{e}_4$, $\tilde{e}_4$}
{$\tilde{e}_1$, $\tilde{e}_2$, $\tilde{e}_3$}
{$\tilde{e}_1$, $\tilde{e}_3$, $\tilde{e}_4$}
{$\tilde{e}_2$, $\tilde{e}_3$, $\tilde{e}_4$}

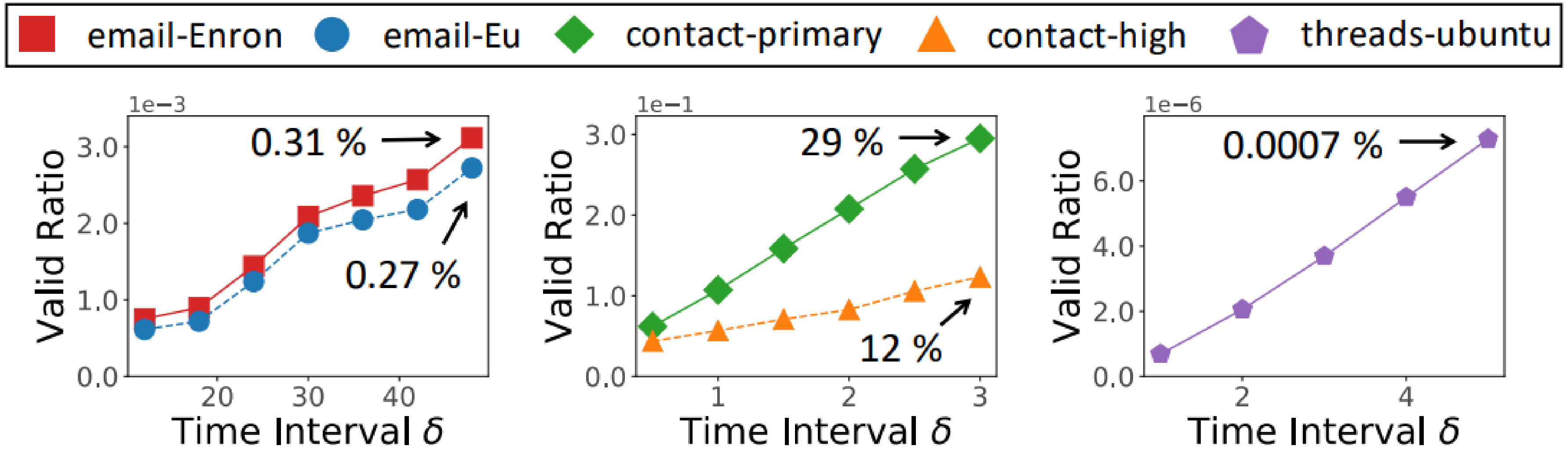
Instances of static h-motifs
($\leq$ 3 hyperedges)

{e_3, e_4, e_5}
{e_6, e_4, e_5}
{e_1, e_2, e_3}
{e_1, e_2, e_3}
{e_1, e_3, e_4}
{e_2, e_3, e_4}
{e_2, e_3, e_5}

Instances of temporal h-motifs
($\delta = 5$)

DP: Naïve Approach Using Dynamic Programming (cont.)

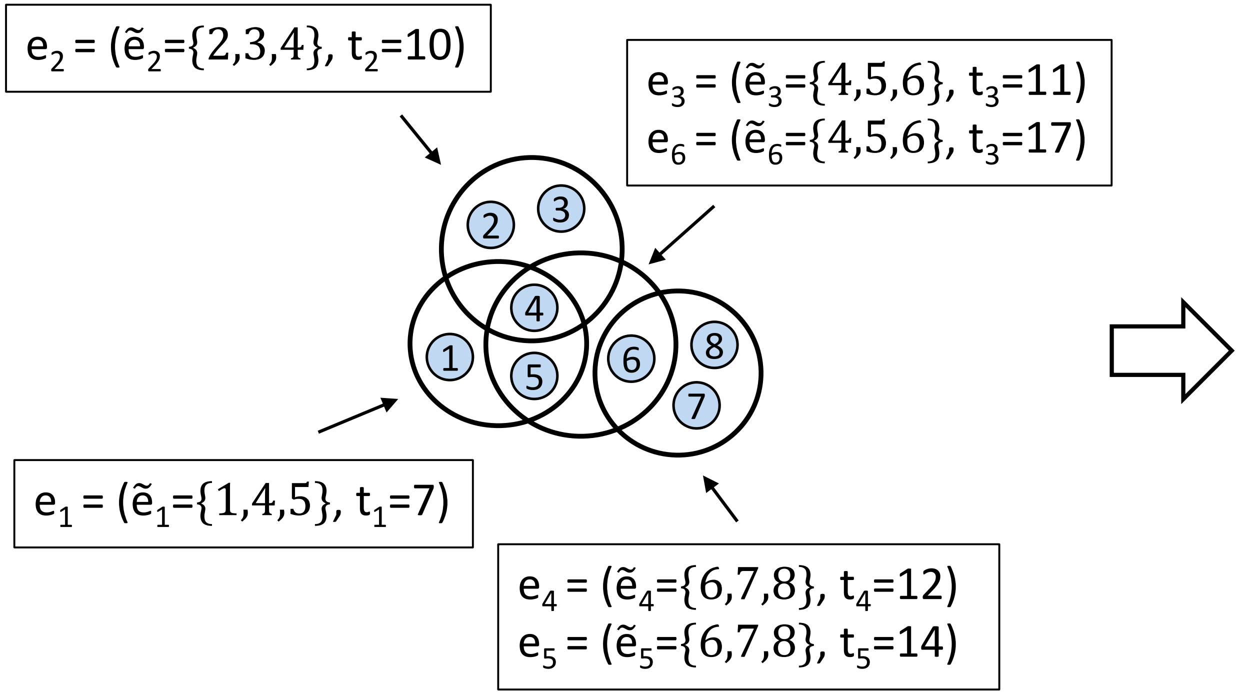
- DP enumerates all the instances of static h-motifs in the induced static hypergraph.
- However, only a small fraction of them are induced by any valid instance of TH-motifs.

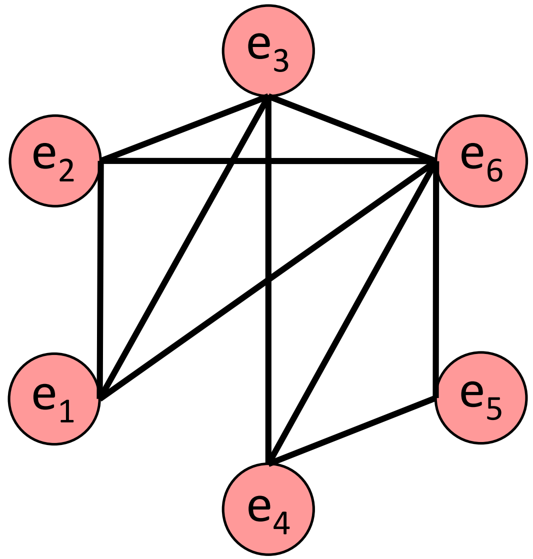


email contact threads

THyMe: Preliminary Version of Counting TH-Motifs

- THyMe exhaustively enumerates the instances of TH-motifs.
- THyMe **incrementally maintains** the projected graph $P = (V_P, E_P)$ where each **temporal hyperedge** is represented as a node.

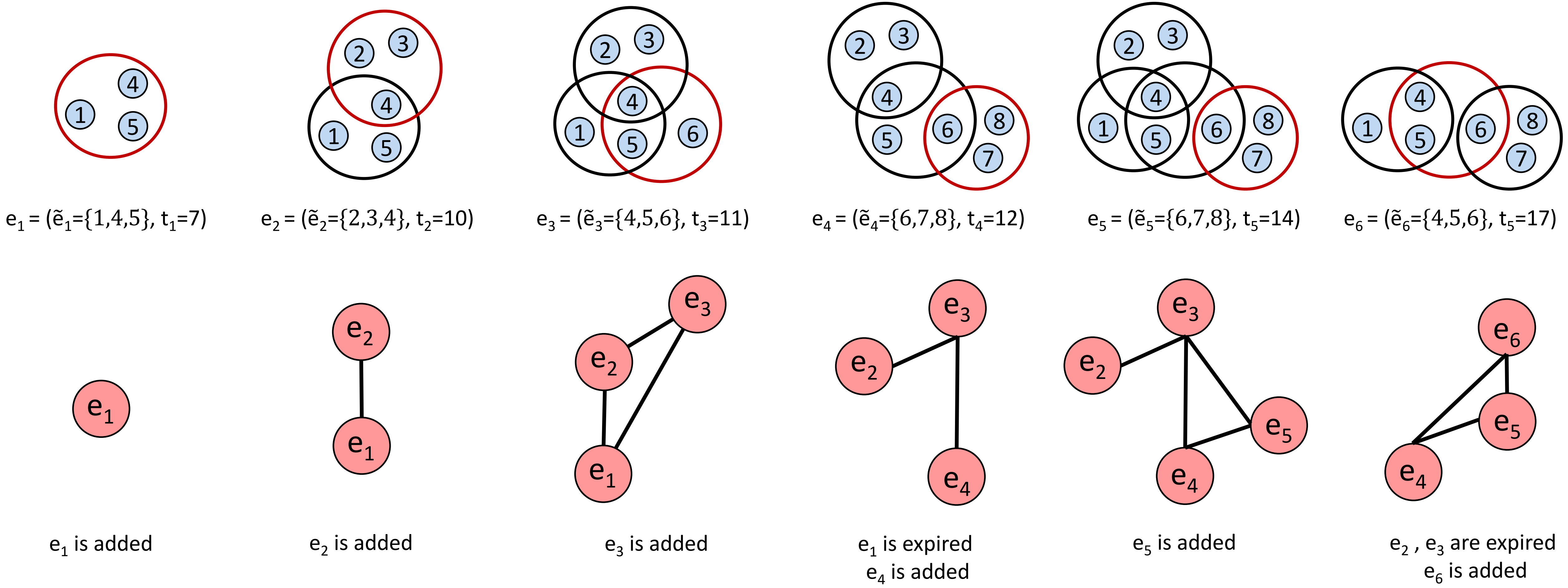




Projected graph P

THyMe: Preliminary Version of Counting TH-Motifs (cont.)

- THyMe **incrementally maintains** the projected graph $P = (V_P, E_P)$ (e.g., $\delta = 4$).

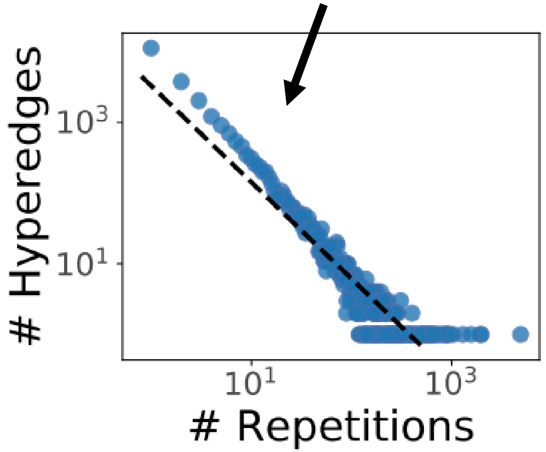


THyMe: Preliminary Version of Counting TH-Motifs (cont.)

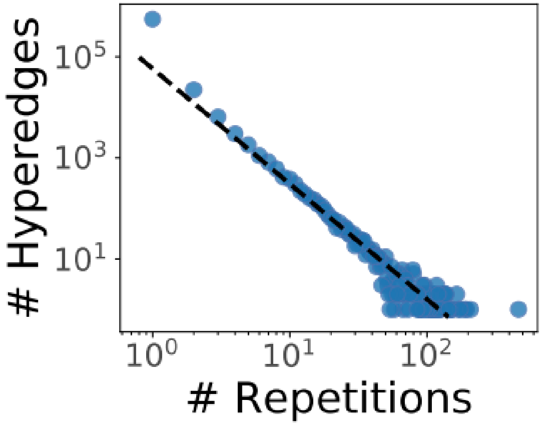
Obs4. Duplicated temporal hyperedges are common.

→ There can exist multiple *identical* nodes in the projected graph P.

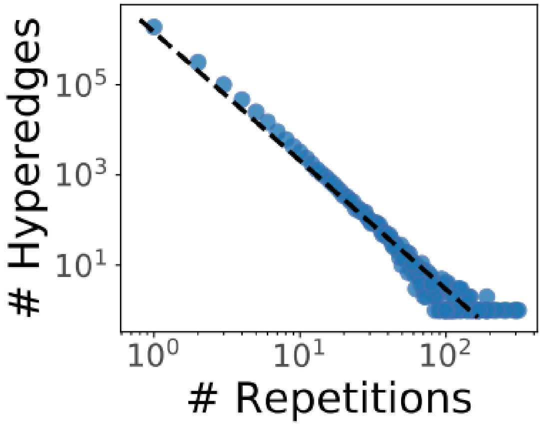
The number of repetitions follow a near power-law distribution.



email-Eu



threads-math

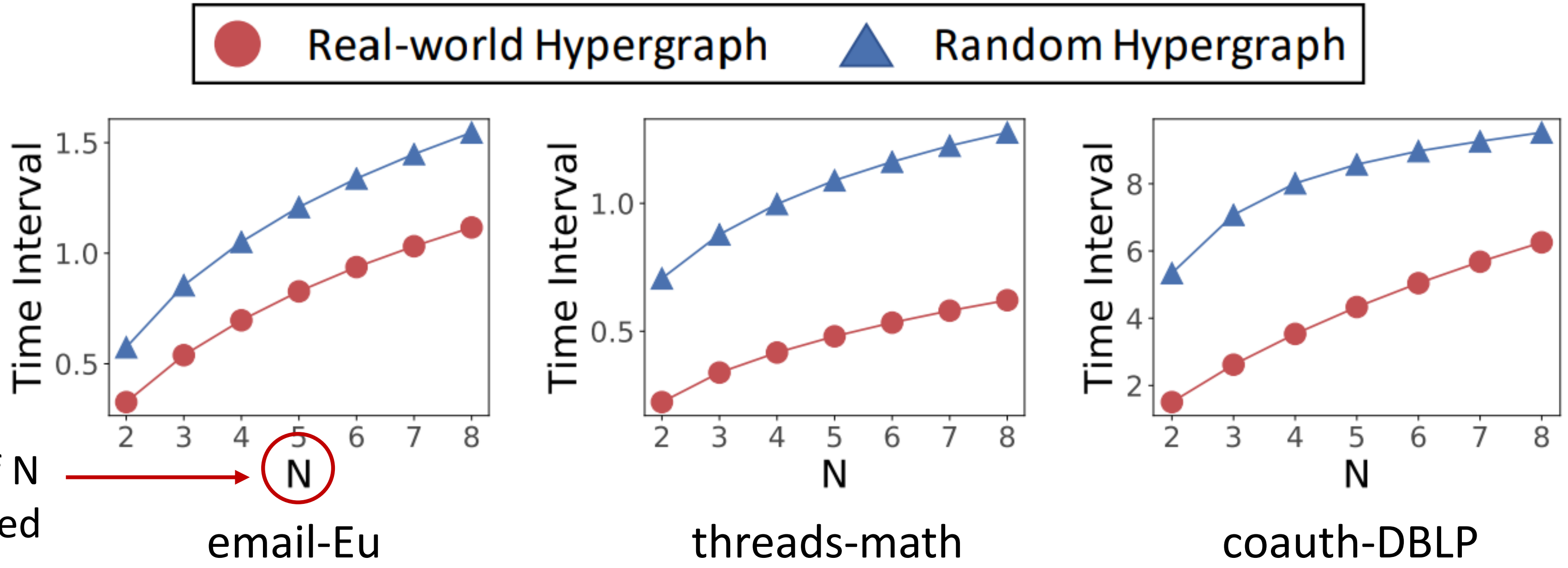


coauth-DBLP

THyMe: Preliminary Version of Counting TH-Motifs (cont.)

Obs5. Future temporal hyperedges are more likely to repeat recent hyperedges.

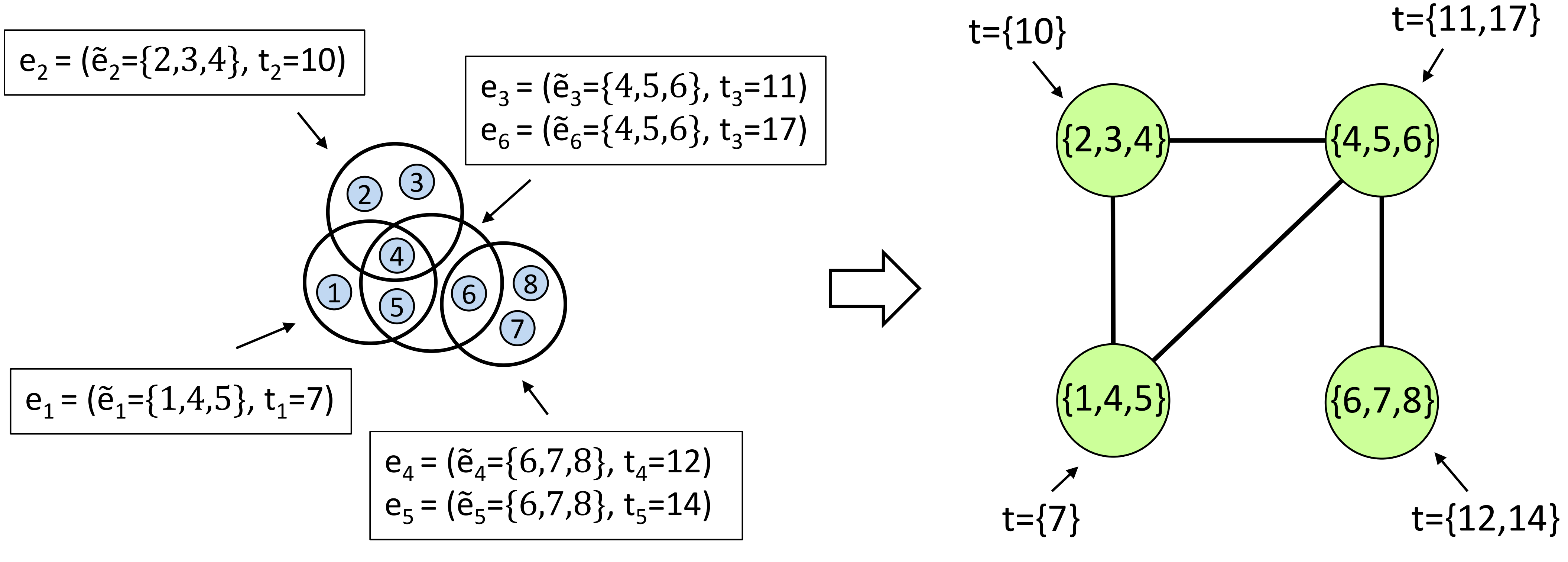
→ *Identical* nodes are more likely to exist within the window in the projected graph P.



The time intervals of N consecutive duplicated temporal hyperedges.

THyMe$^+$: Advanced Version of Counting TH-Motifs

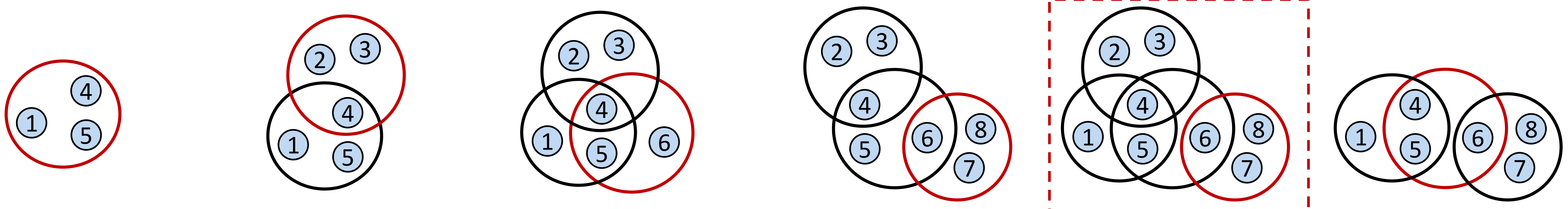
- THyMe$^+$ **incrementally maintains** the projected graph $Q = (V_Q, E_Q, t_Q)$ where each **induced static hyperedge** is represented as a node.



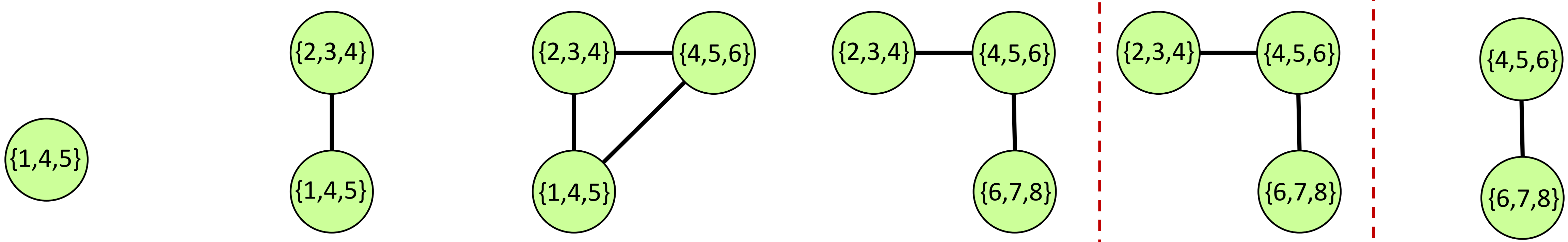
Projected graph Q

THyMe$^+$: Advanced Version of Counting TH-Motifs (cont.)

- THyMe$^+$ **incrementally maintains** the projected $Q = (V_Q, E_Q, t_Q)$ (e.g., $\delta = 4$).



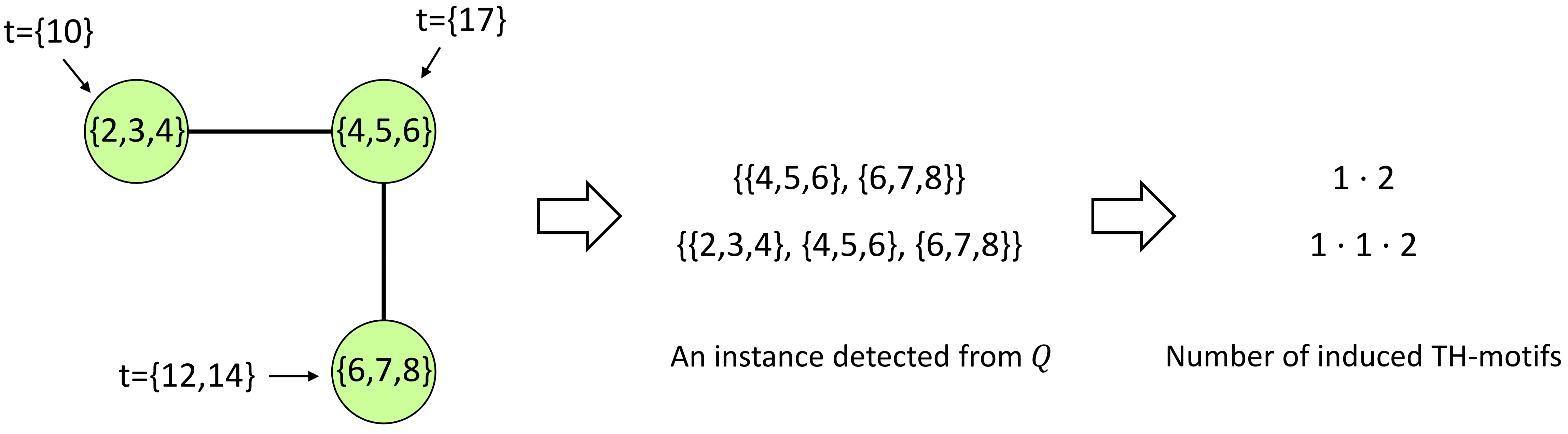
$e_1 = (\tilde{e}_1=\{1,4,5\}, t_1=7)$ | $e_2 = (\tilde{e}_2=\{2,3,4\}, t_2=10)$ | $e_3 = (\tilde{e}_3=\{4,5,6\}, t_3=11)$ | $e_4 = (\tilde{e}_4=\{6,7,8\}, t_4=12)$ | $e_5 = (\tilde{e}_5=\{6,7,8\}, t_5=14)$ | $e_6 = (\tilde{e}_6=\{4,5,6\}, t_5=17)$



{1,4,5} is added | {2,3,4} is added | {4,5,6} is added | {1,4,5} is expired {6,7,8} is added | **no changes** | {2,3,4} is expired

THyMe$^+$: Advanced Version of Counting TH-Motifs (cont.)

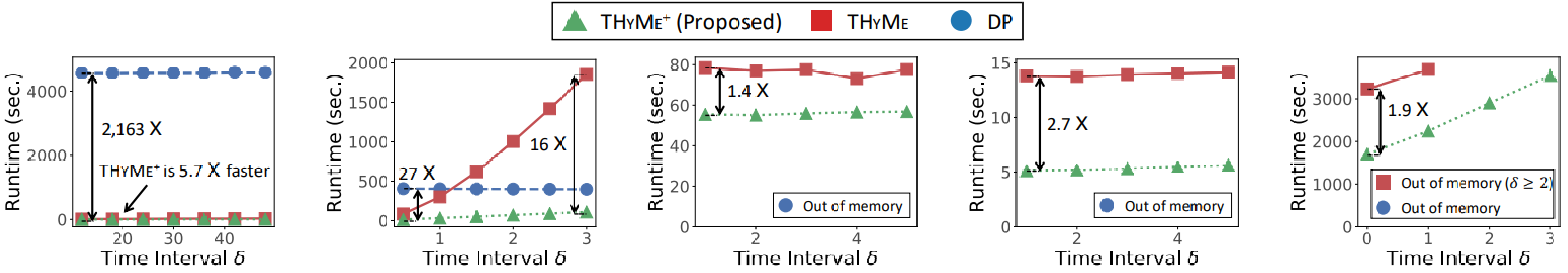
- THyMe$^+$ avoids the exhaustive enumeration of instances of TH-motifs by counting them based on the timestamps of the nodes V_Q of Q.



Projected graph Q

Speed and Efficiency of THyMe$^+$

- THyMe$^+$ is **faster** and **more space efficient** than DP and THyMe.



email-Eu contact-primary threads-math tags-ubuntu coauth-DBLP

Roadmap

1. Backgrounds
2. Concepts
3. Observations
4. Algorithms
5. **Conclusion**

Conclusion

- We propose **temporal hypergraph motifs** (**TH-motifs**) for describing structural and temporal patterns of real-world temporal hypergraphs.

Our contributions are:

✓ **New Concept**: We define 96 temporal hypergraph motifs.

✓ **Fast & Exact Algorithm**: We develop THyMe$^+$ for counting instances of TH-motifs.

✓ **Empirical Discoveries**: TH-motifs reveal interesting structural & temporal patterns.

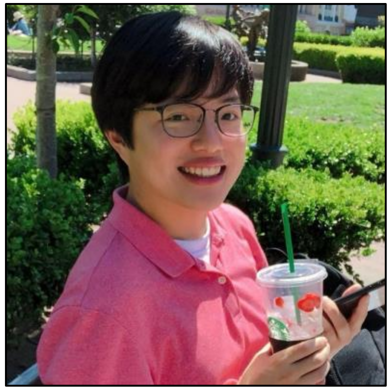
Code & datasets: https://github.com/geonlee0325/THyMe

KAIST AI

ICDM 2021
IEEE International Conference on Data Mining
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

THyMe$^+$: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting

Geon Lee



Kijung Shin