Mining of Real-world Hypergraphs: Concepts, Patterns, and Generators

Part 2. Dynamic Structural Patterns

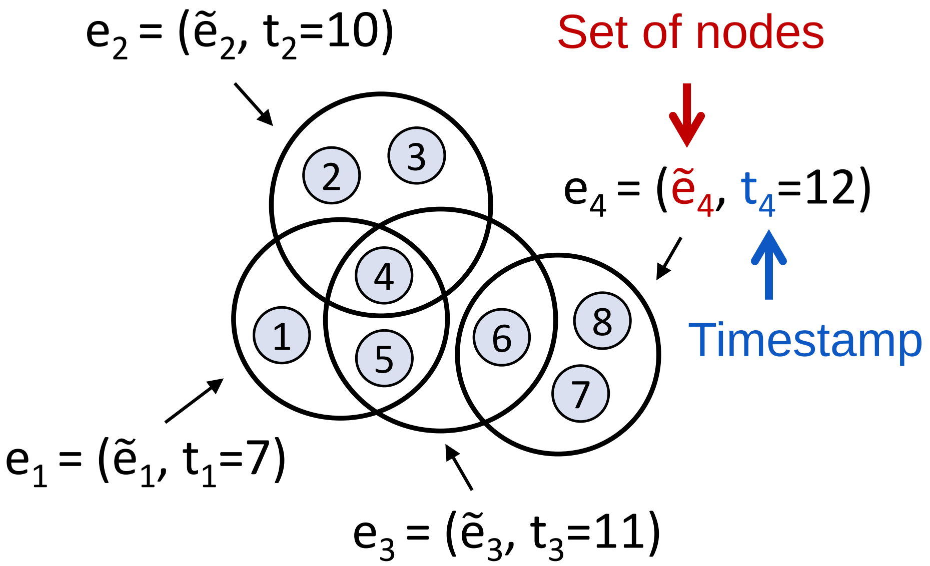
Geon Lee

Jaemin Yoo

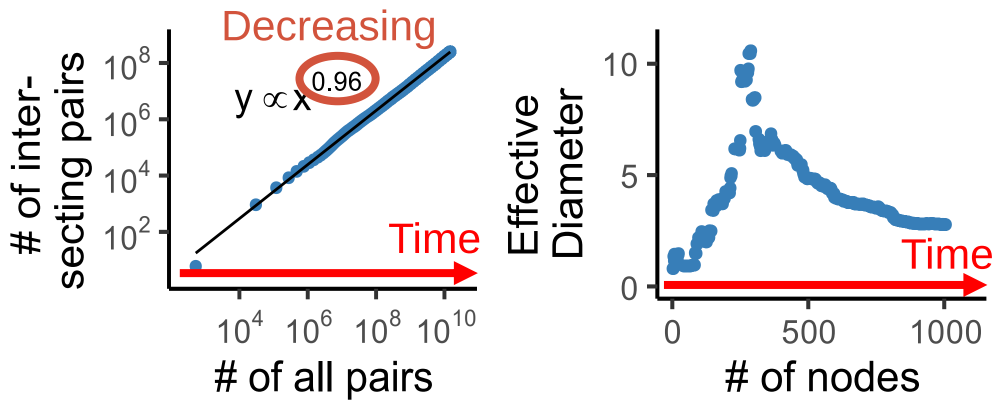
Kijung Shin

Part 2. Dynamic Structural Patterns

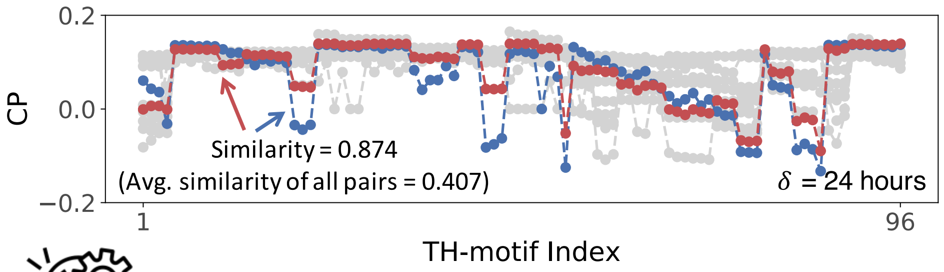
*“Given a **dynamic** hypergraph, how can we analyze its structure?”*



Input Temporal Hypergraph



Basic Patterns (Part 2-1)



Advanced Patterns (Part 2-2)

Roadmap

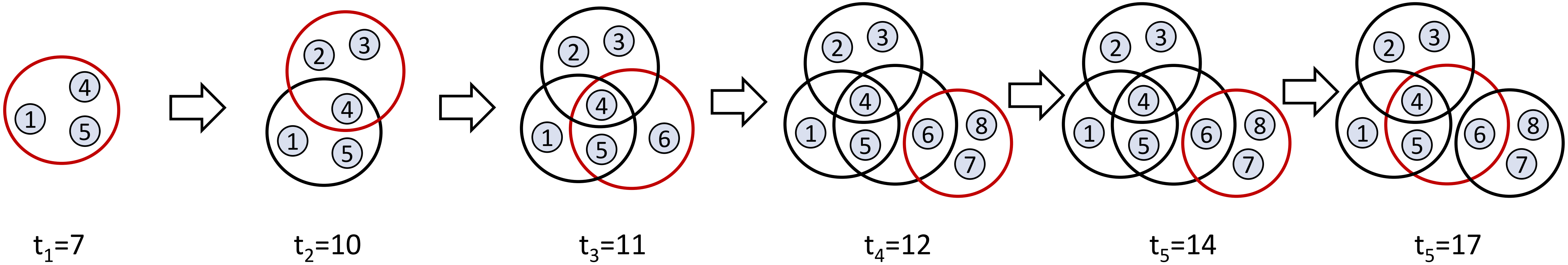
- **Part 1. Static Structural Patterns**
  - Basic Patterns
  - Advanced Patterns
- **Part 2. Dynamic Structural Patterns**
  - Basic Patterns **<<**
  - Advanced Patterns
- **Part 3. Generative Models**
  - Static Hypergraph Generator
  - Dynamic Hypergraph Generator

Part 2-1. Basic Dynamic Structural Patterns

		Part 1. Static Patterns	Part 2. Dynamic Patterns
Basic Patterns	Node-Level	DYHS20, KKS20, LCS21	BKT18, CS22
	Hyperedge-Level	KKS20, LCS21	BKT18, LS21, GLLB23, CBLK21
	Hypergraph-Level	BASJK18, DYHS20, KKS20	KKS20
Advanced Patterns	Sub-hypergraph-Level	KBCYK23, BASJK18, LMMB22, LKS20, LCS21, LL23, BLS23	BASJK18, CJ21, LS21, CS22

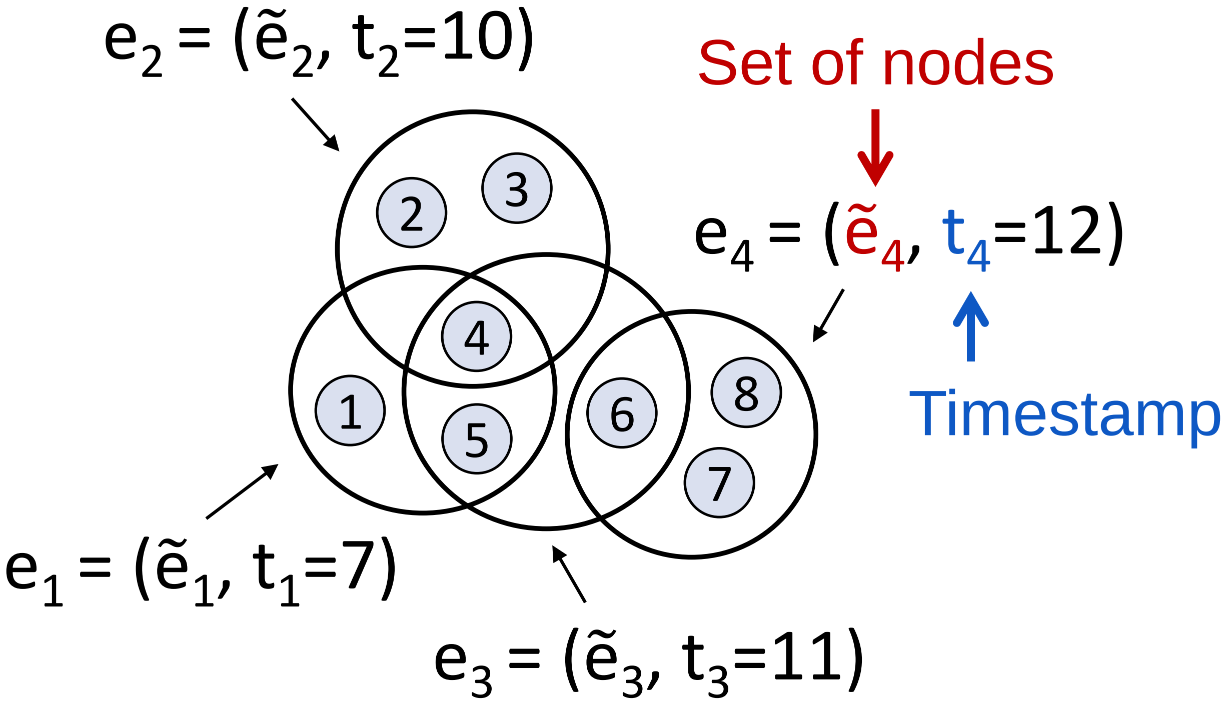
Background

- **Temporal hypergraphs** consist of **temporal hyperedges**.
  - Temporal hyperedges consisting of the same set of nodes can appear at different timestamps.



Background (cont.)

- Each **temporal hyperedge** consists of the set of nodes and the timestamp it arrived at.



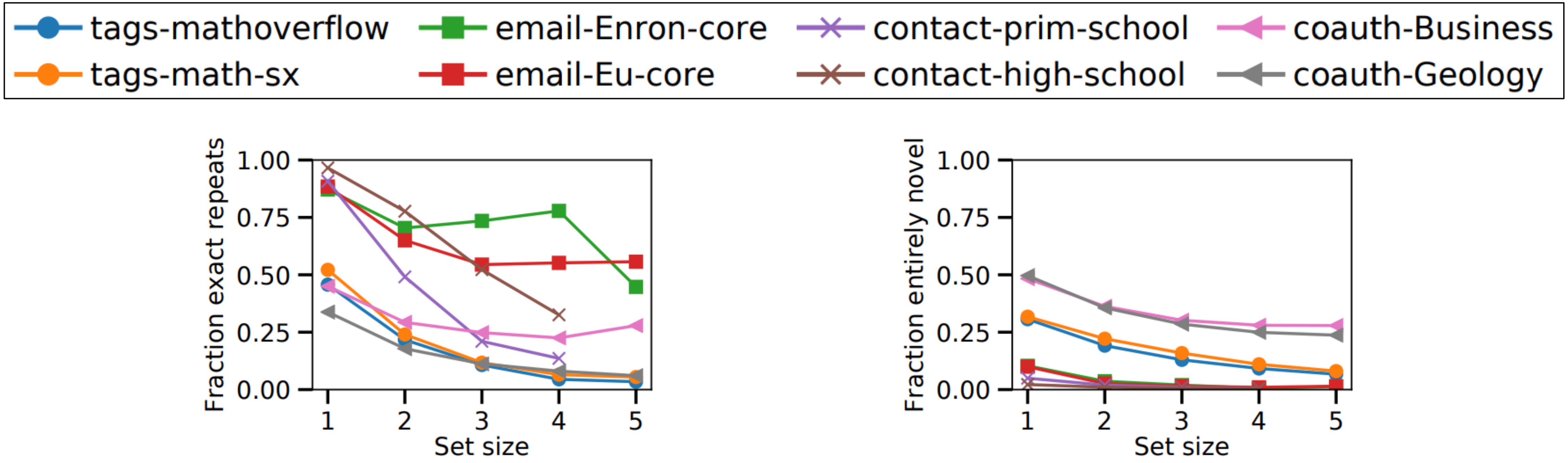
BKT18: Three Basic Dynamic Patterns

- **P1.** Repeat behavior
- **P2.** Subset correlation
- **P3.** Recency bias

	Node		P2
	Hyperedge		P1, P3
	Hypergraph		
	Sub-hypergraph		

Repeat Behavior

- Temporal hyperedges tend to **repeat** previous ones.



Exactly repeated hyperedges **Entirely novel hyperedges**

Subset Correlation

- Subsets of nodes tend to be **correlated**.
  - Size-2 and size-3 subsets co-appear in multiple temporal hyperedges.

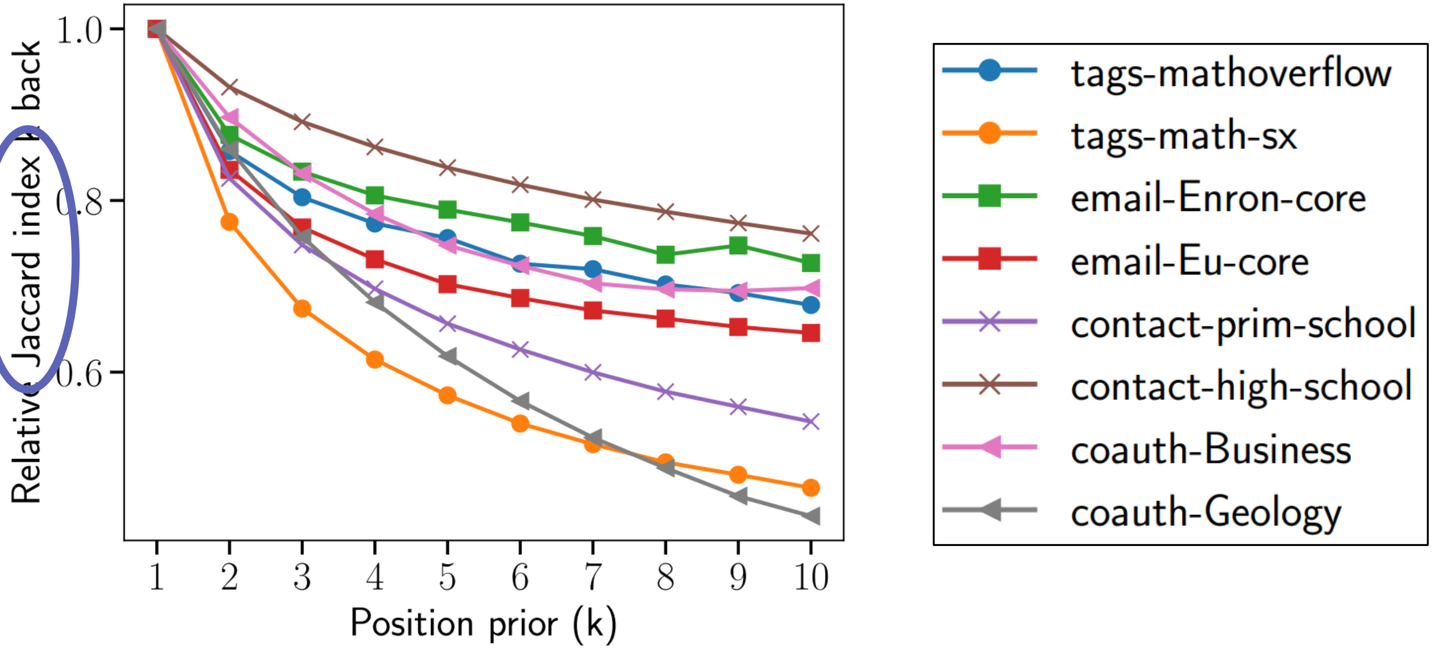
	size-2 subset counts		size-3 subset counts	
Dataset	data	null model	data	null model
email-Enron-core	5.82	4.34 ± 0.043	4.23	2.67 ± 0.038
email-Eu-core	4.46	3.11 ± 0.008	3.23	2.08 ± 0.007
contact-prim-school	2.36	1.87 ± 0.003	1.35	1.09 ± 0.002
contact-high-school	4.49	3.26 ± 0.007	2.09	1.35 ± 0.004
tags-mathoverflow	1.49	1.41 ± 0.002	1.18	1.15 ± 0.002
tags-math-sx	1.49	1.31 ± 0.001	1.21	1.12 ± 0.001
coauth-Business	1.50	1.30 ± 0.001	1.40	1.24 ± 0.001
coauth-Geology	1.29	1.15 ± 0.000	1.15	1.07 ± 0.000

Average number of times of size-2/3 subset appearance

Recency Bias

- Temporal hyperedges tend to be similar to **recent** ones.

Recent ⟷ Old



Similarity between two hyperedges

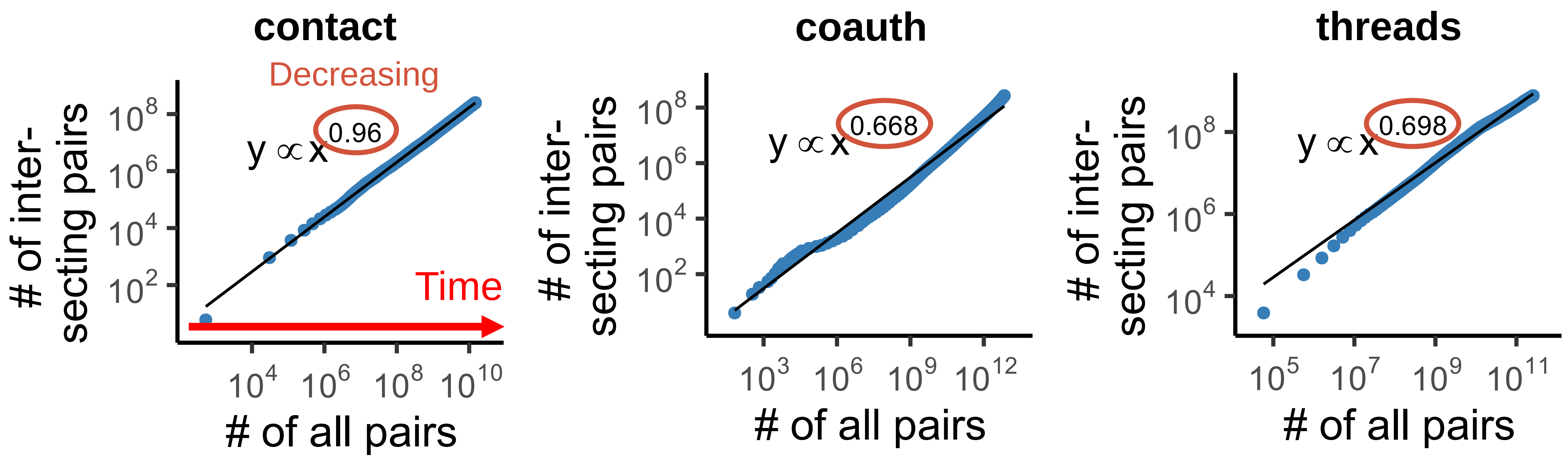
KKS20: Three Basic Dynamic Patterns

- **P1.** Diminishing overlaps
- **P2.** Densification
- **P3.** Shrinking diameter

	Node		
	Hyperedge		
	Hypergraph		P1-P3
	Sub-hypergraph		

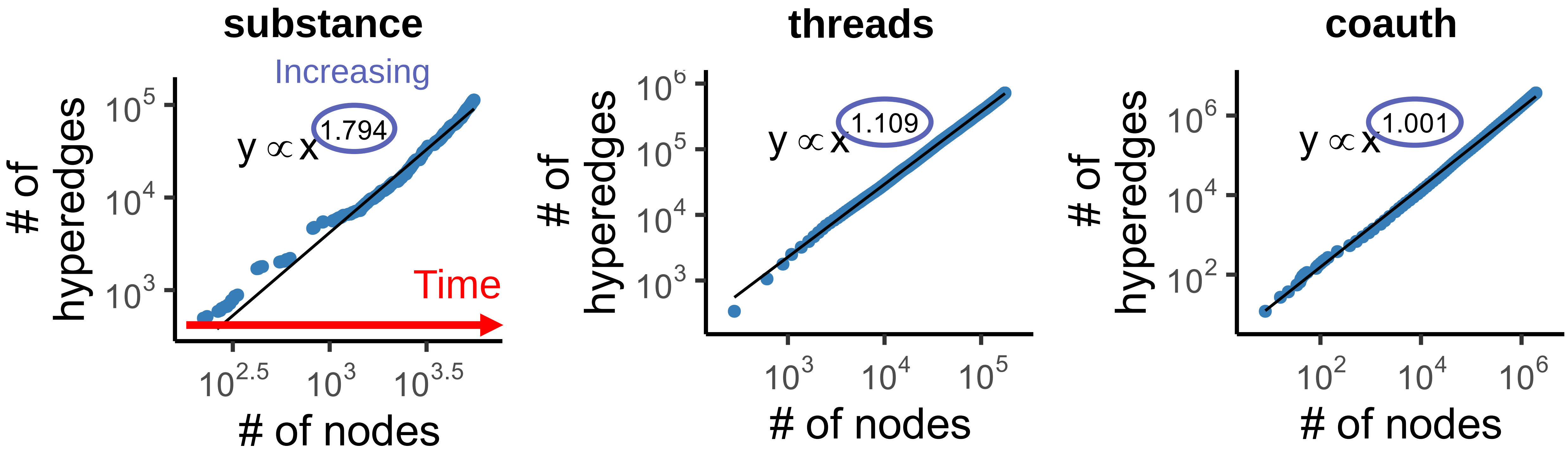
Diminishing Overlaps

- The overlap between hyperedges **decreases** over time.



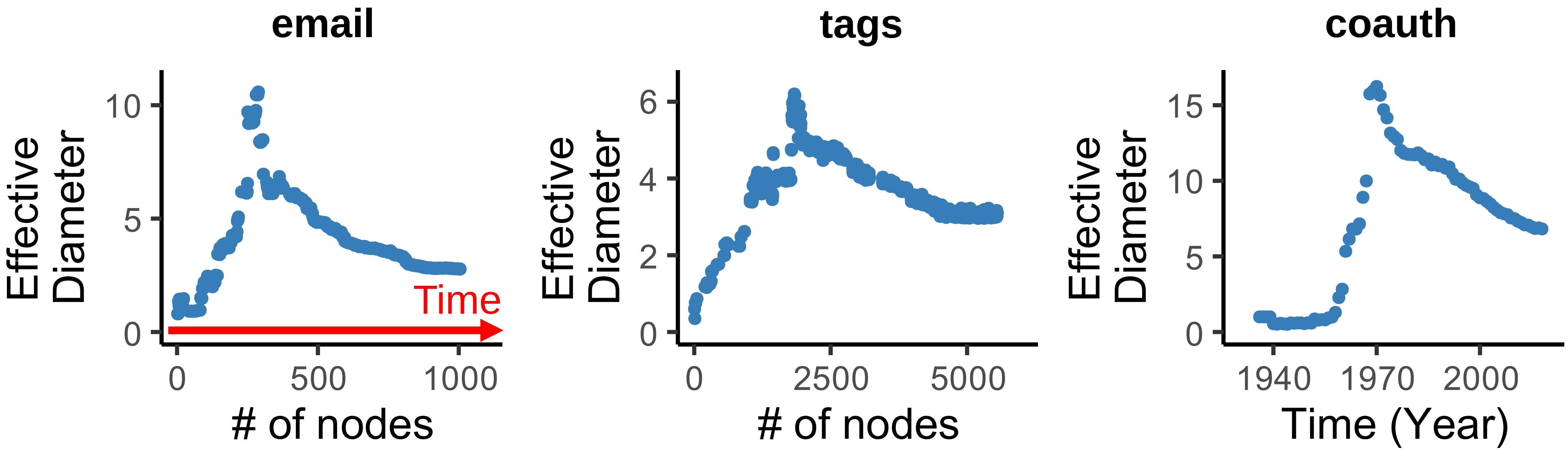
Densification

- Average node degree of the hypergraph **increases** over time.



Shrinking Diameter

- Effective diameters eventually **decrease** over time.



LS21: Two Basic Dynamic Patterns

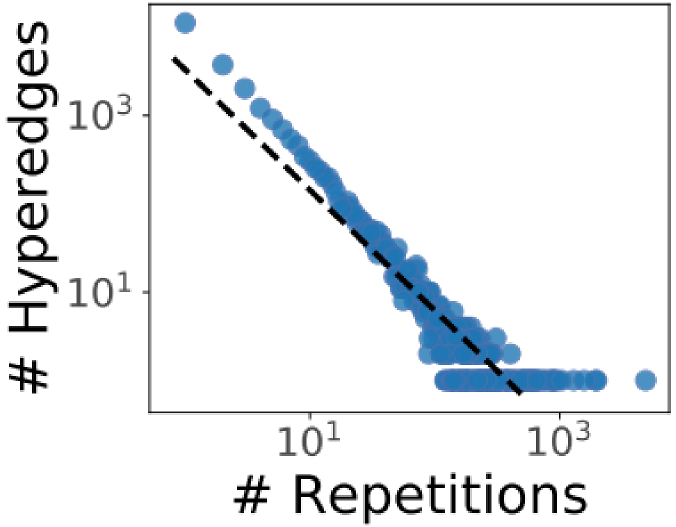
- **P1.** Hyperedge repetition
- **P2.** Temporal locality

	Node		
	Hyperedge		P1, P2
	Hypergraph		
	Sub-hypergraph		

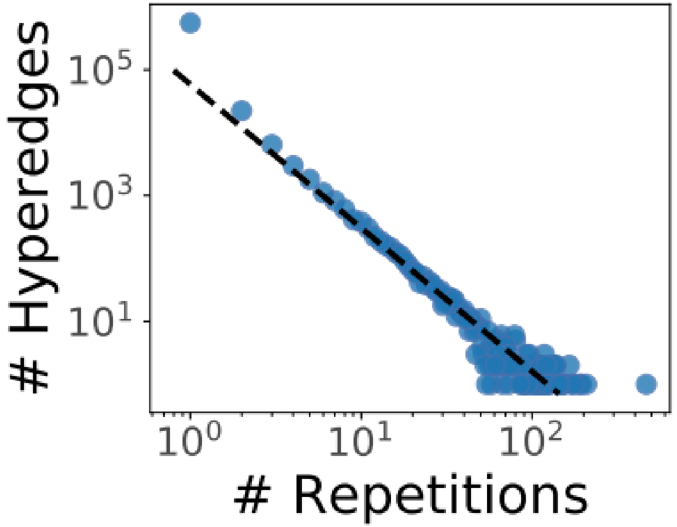
Hyperedge Repetition

- Temporal hyperedges in real-world hypergraphs are **repetitive**.
- The number of repetitions is **heavy-tailed**.

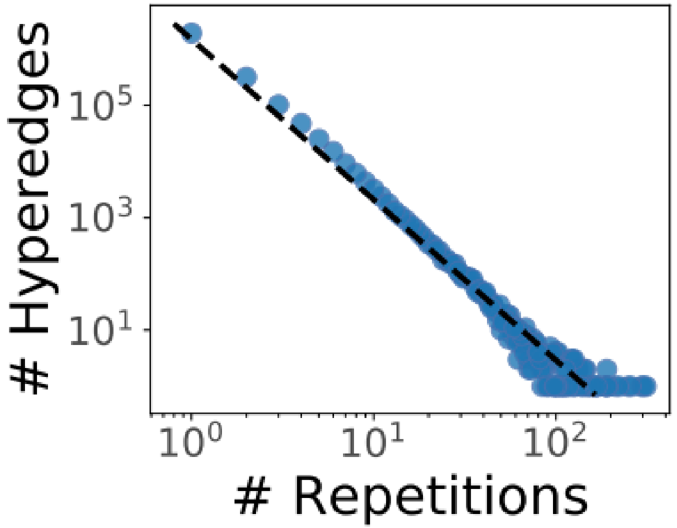
email-EU



threads-math



coauth-DBLP

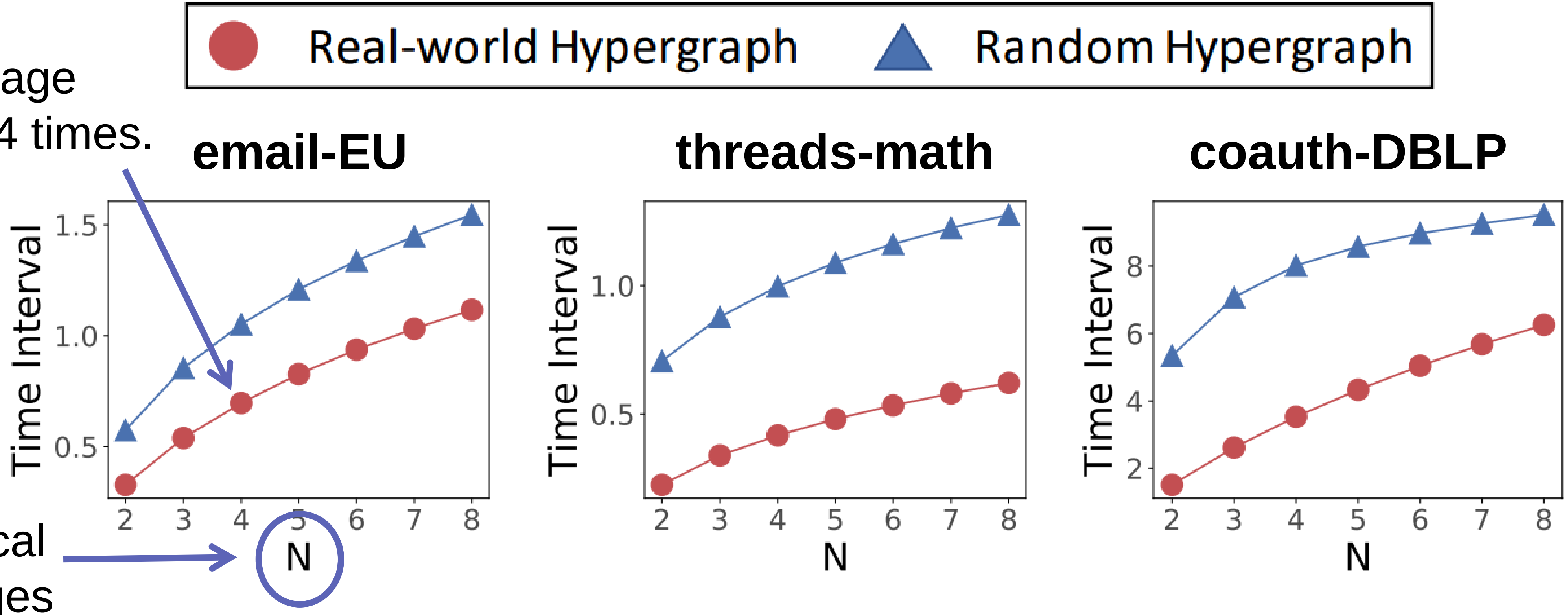


Temporal Locality

- Future temporal hyperedges are more likely to **repeat recent** ones.

e.g., It took in average 0.7 time to repeat 4 times.

Number of consecutive identical temporal hyperedges



GLLB23: Two Basic Dynamic Patterns

- **P1.** Temporal correlation between intra-order hyperedges
- **P2.** Temporal correlation between cross-order hyperedges

	Node		
	Hyperedge		P1, P2
	Hypergraph		
	Sub-hypergraph		

Temporal Correlations

?

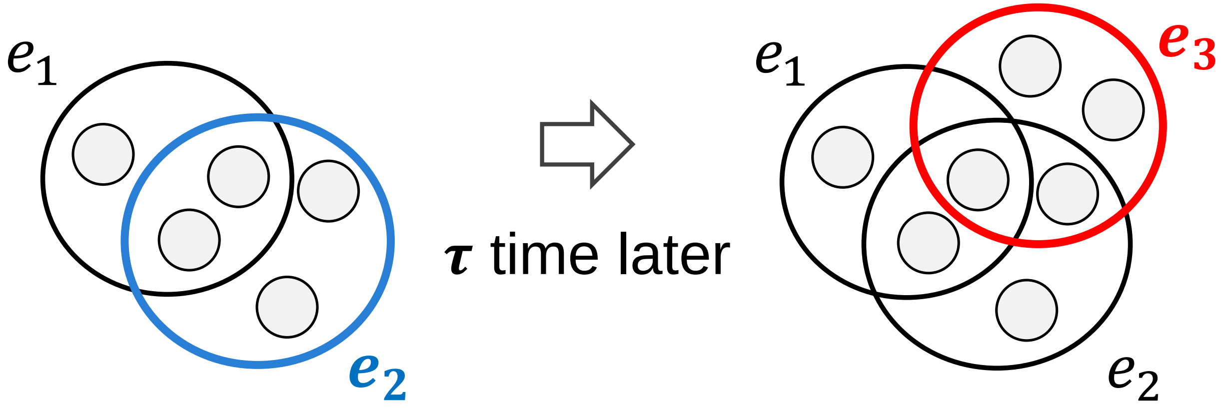
Question:

How are hyperedges **temporally correlated**?

Answer:

Temporal correlations between hyperedges are computed.

How does $\boldsymbol{e_2}$ *(before* $\boldsymbol{\tau}$ *time) affect the formation of* $\boldsymbol{e_3}$*?*

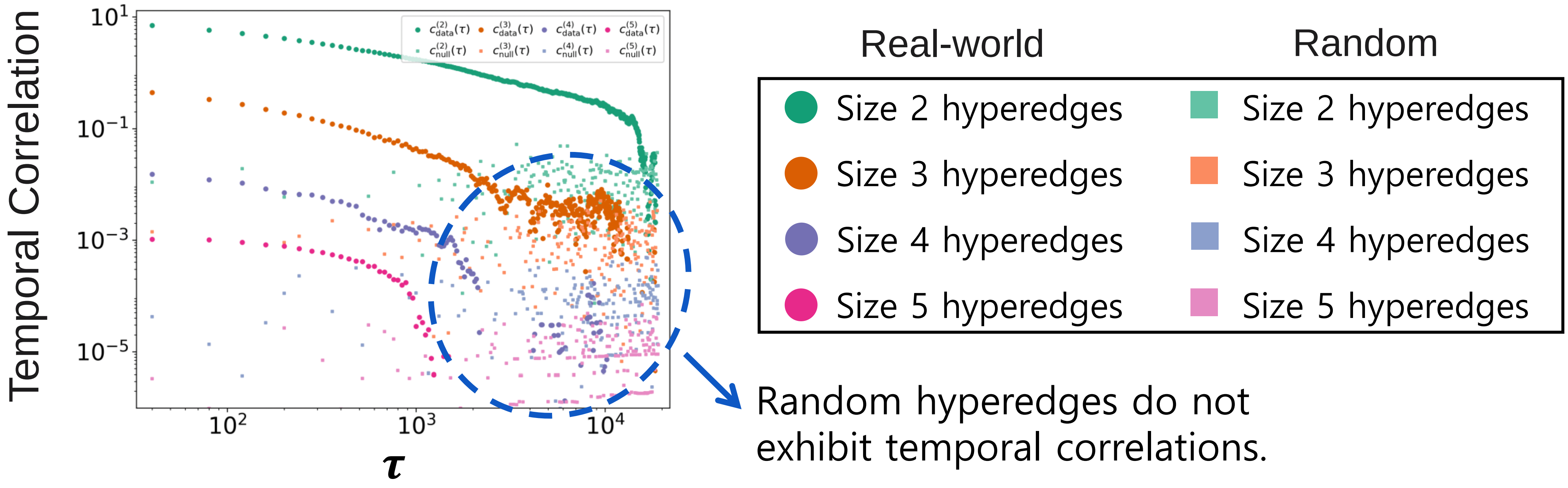


!

L. Gallo, L. Lacasa, V. Latora, and F. Battiston, "Higher-order Correlations Reveal Complex Memory in Temporal Hypergraphs", **arXiv (2023)**

Intra-Order Temporal Correlations

- Temporal correlations exist between hyperedges with the **same size**.
- Correlations decay slowly with the time interval $\boldsymbol{\tau}$.

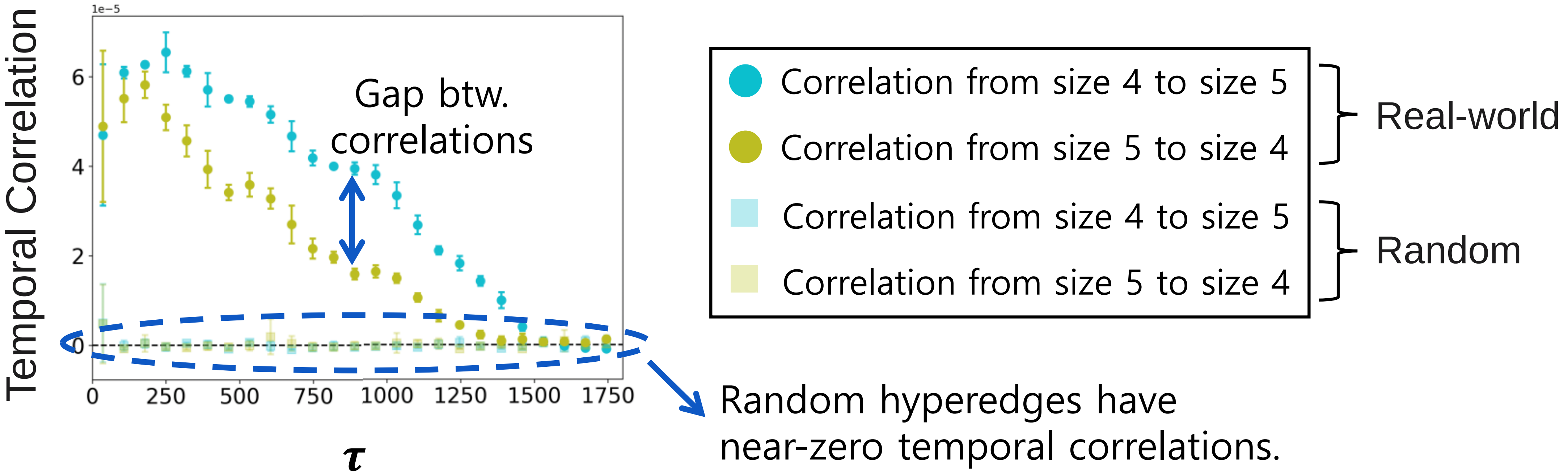


Random hyperedges do not exhibit temporal correlations.

L. Gallo, L. Lacasa, V. Latora, and F. Battiston, "Higher-order Correlations Reveal Complex Memory in Temporal Hypergraphs", **arXiv (2023)**

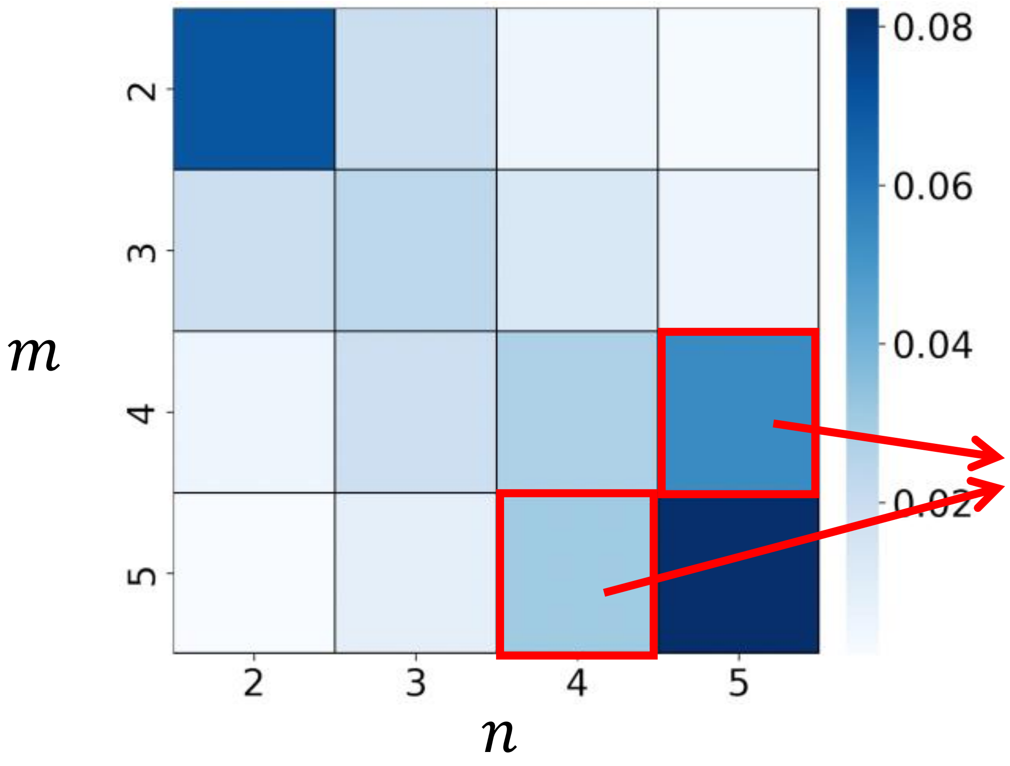
Cross-Order Temporal Correlations

- Temporal correlations exist between hyperedges with **different sizes**.



Cross-Order Temporal Correlations (cont.)

- Cross-order temporal correlations are **asymmetric**.

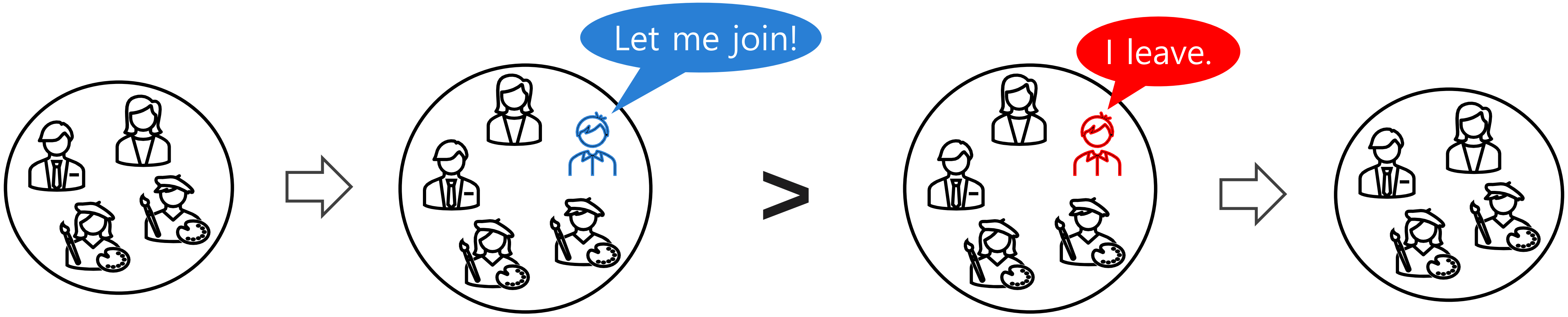


Temporal correlation **from size 4 to 5** is higher than that of **from size 5 to 4**.

L. Gallo, L. Lacasa, V. Latora, and F. Battiston, "Higher-order Correlations Reveal Complex Memory in Temporal Hypergraphs", **arXiv (2023)**

Cross-Order Temporal Correlations (cont.)

- The **asymmetry** suggests the formation or loss of group interactions.



Temporal correlation **from size 4 to size 5**

Temporal correlation **from size 5 to size 4**

L. Gallo, L. Lacasa, V. Latora, and F. Battiston, "Higher-order Correlations Reveal Complex Memory in Temporal Hypergraphs", **arXiv (2023)**

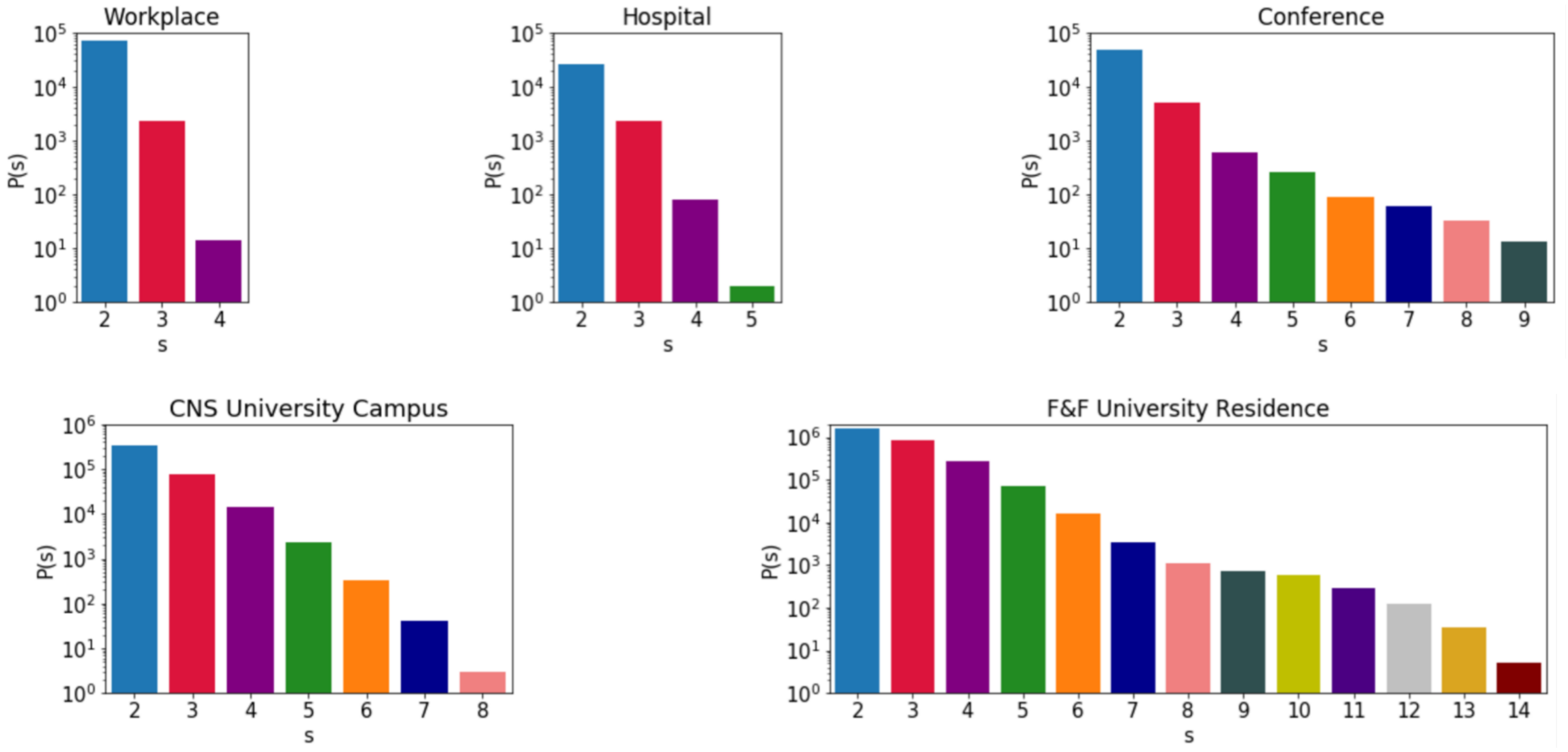
CBLK21: Four Basic Dynamic Patterns

- **P1.** Hyperedge sizes and frequencies
- **P2.** Temporal heterogeneity of hyperedges
- **P3.** Bursty behavior of hyperedges
- **P4.** Temporal reinforcement of hyperedges

	Node		
	Hyperedge		P1, P2, P3, P4
	Hypergraph		
	Sub-hypergraph		

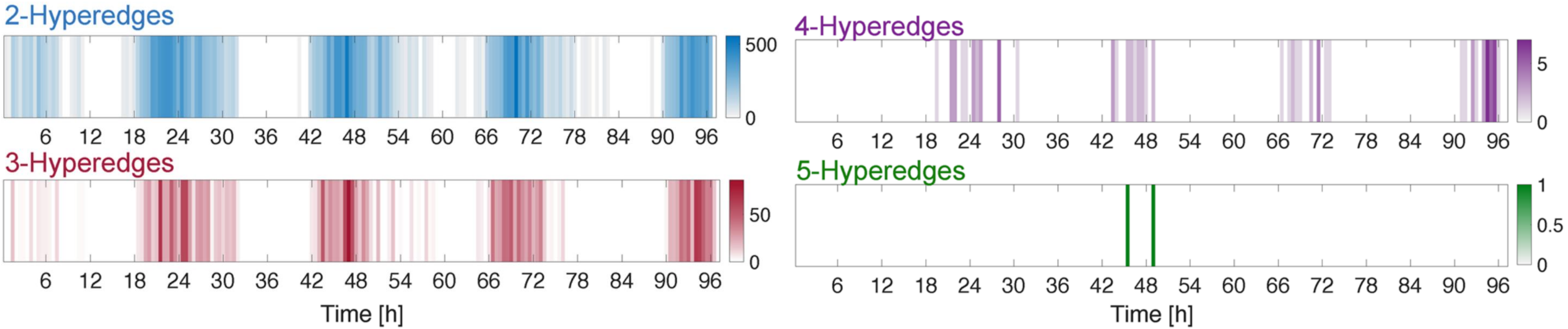
Hyperedge Sizes and Frequencies

- **Smaller hyperedges** are more numerous in a real-world dataset.



Temporal Heterogeneity of Hyperedges

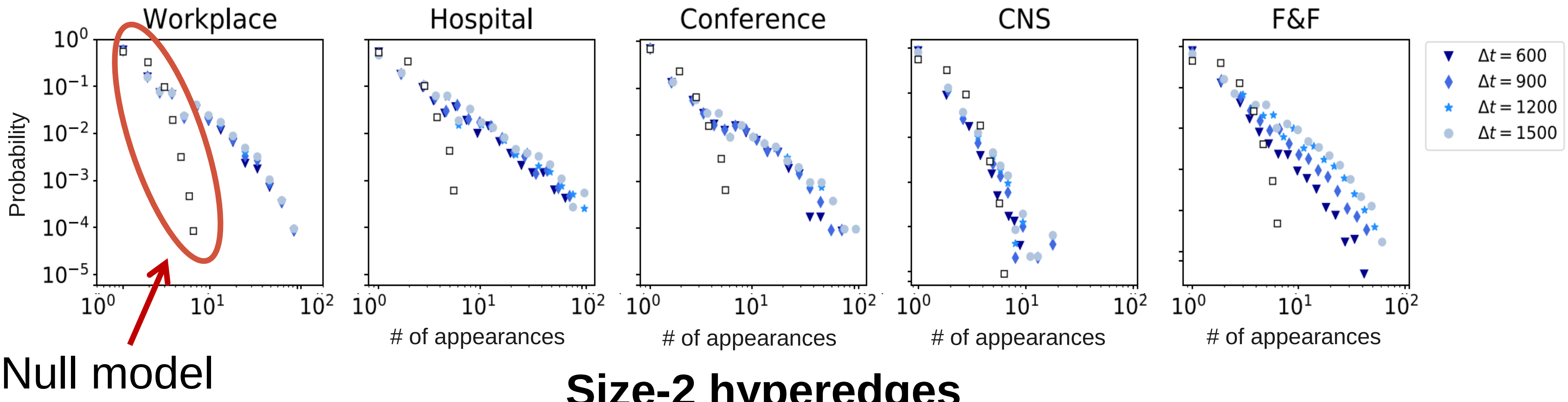
- Emergence of hyperedges is strongly **heterogeneous in time**.
  - Bursty patterns of hyperedges are not independent of size.



G. Cencetti, F. Battiston, B. Lepri, M. Karsai, “Temporal Properties of Higher-order Interactions in Social Networks”, **Scientific Report (2021)**

Bursty Behavior of Hyperedges

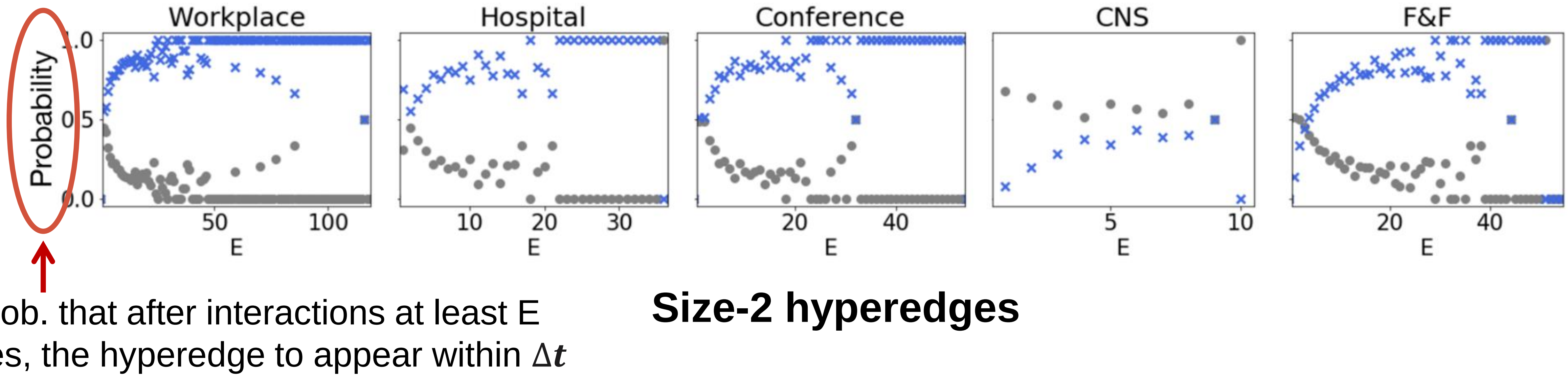
- Hyperedges in real-world hypergraphs are likely to **repeatedly appear within a short time (Δt)** and the distribution of their numbers of appearances are **heavy-tailed**.



Size-2 hyperedges

Temporal Reinforcement of Hyperedges

- There exists **temporal reinforcement** in real-world hypergraphs, i.e., the longer the length of an interaction, the higher chances the interaction will not break down.



Size-2 hyperedges

G. Cencetti, F. Battiston, B. Lepri, M. Karsai, "Temporal Properties of Higher-order Interactions in Social Networks", **Scientific Report (2021)**

Roadmap

- **Part 1. Static Structural Patterns**
  - Basic Patterns
  - Advanced Patterns
- **Part 2. Dynamic Structural Patterns**
  - Basic Patterns
  - Advanced Patterns **<<**
- **Part 3. Generative Models**
  - Static Hypergraph Generator
  - Dynamic Hypergraph Generator

Part 2-2. Advanced Dynamic Structural Patterns

		Part 1. Static Patterns	Part 2. Dynamic Patterns
Basic Patterns	Node-Level	DYHS20, KKS20, LCS21	BKT18, CS22
	Hyperedge-Level	KKS20, LCS21	BKT18, LS21, GLLB23, CBLK21
	Hypergraph-Level	BASJK18, DYHS20, KKS20	KKS20
Advanced Patterns	Sub-hypergraph-Level	KBCYK23, BASJK18, LMMB22, LKS20, LCS21, LL23, BLS23	BASJK18, CJ21, LS21, CS22

BASJK18: One Advanced Dynamic Pattern

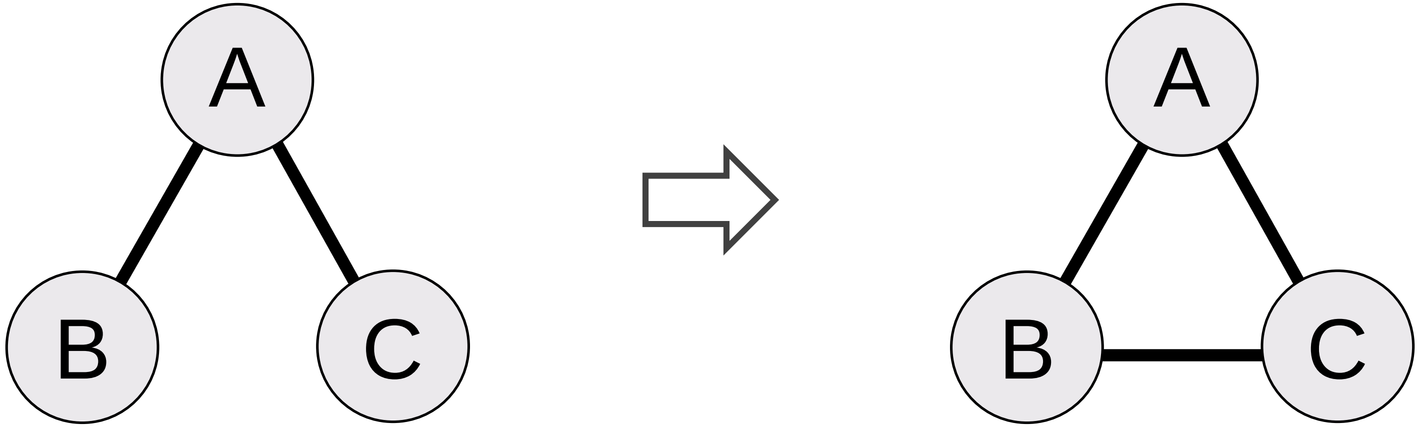
- **P1.** Simplicial closure

	Node		
	Hyperedge		
	Hypergraph		
	Sub-hypergraph		P1

A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, "Simplicial Closure and Higher-order Link Prediction", **PNAS (2018)**

Background

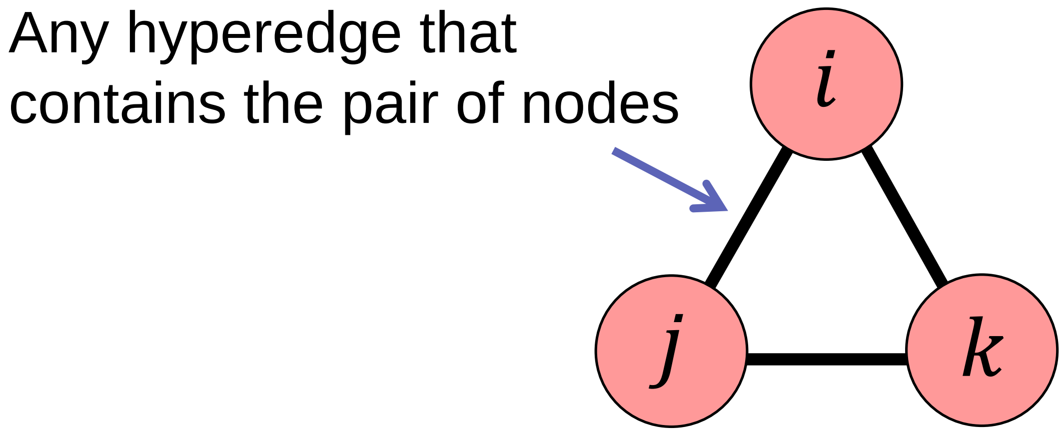
- **Triadic closure** in networks is a tendency of forming a triangle.



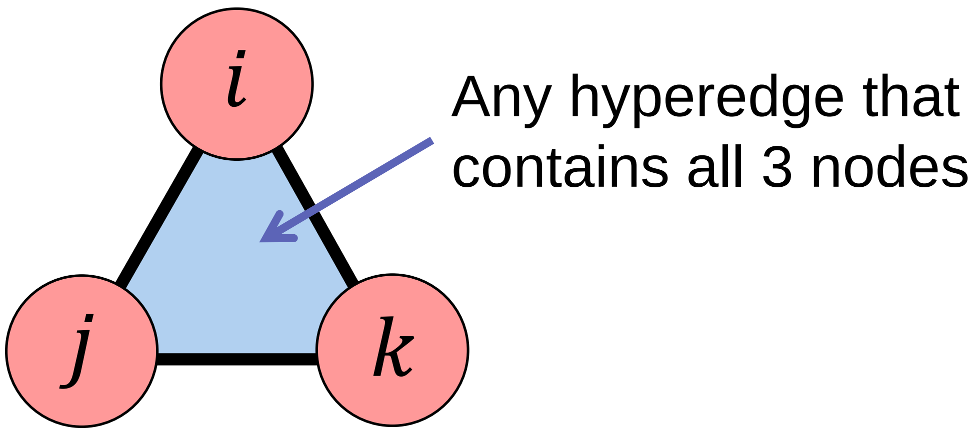
A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, “Simplicial Closure and Higher-order Link Prediction”, **PNAS (2018)**

Recap: Open and Closed Triangles

- There are two types of **triangles** in hypergraphs.
  - Static observations can be found at Part 1.



Open Triangle

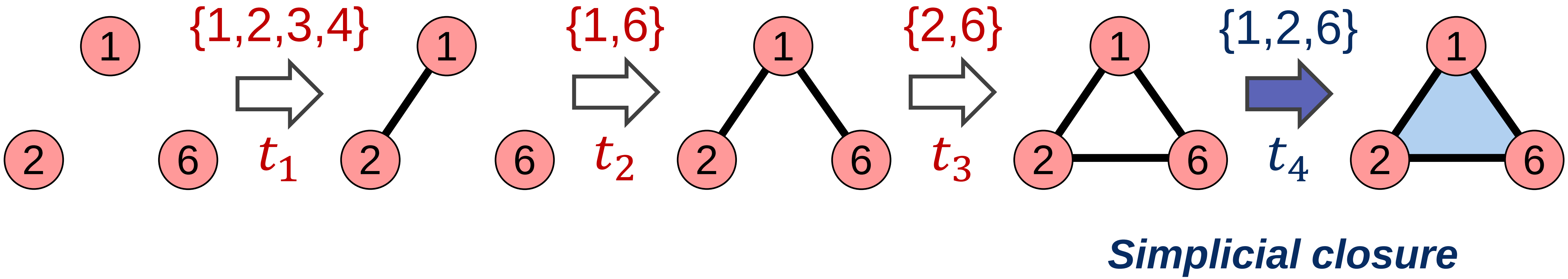


Closed Triangle

A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, "Simplicial Closure and Higher-order Link Prediction", **PNAS (2018)**

Simplicial Closure

- **Simplicial closure** is an event of forming a **closed triangle**.



A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, “Simplicial Closure and Higher-order Link Prediction”, **PNAS (2018)**

Structural Factors of Simplicial Closure

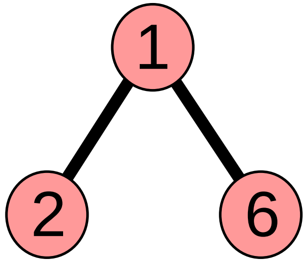
Questions:

How does the simplicial closure occur?

Can we predict it from the structure of the projected graph?

Approach:

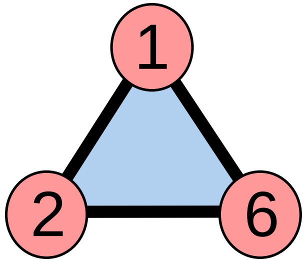
First 80% of the data (in time)



Triplets not in closed triangle

Compute the ratio

Last 20% of the data (in time)

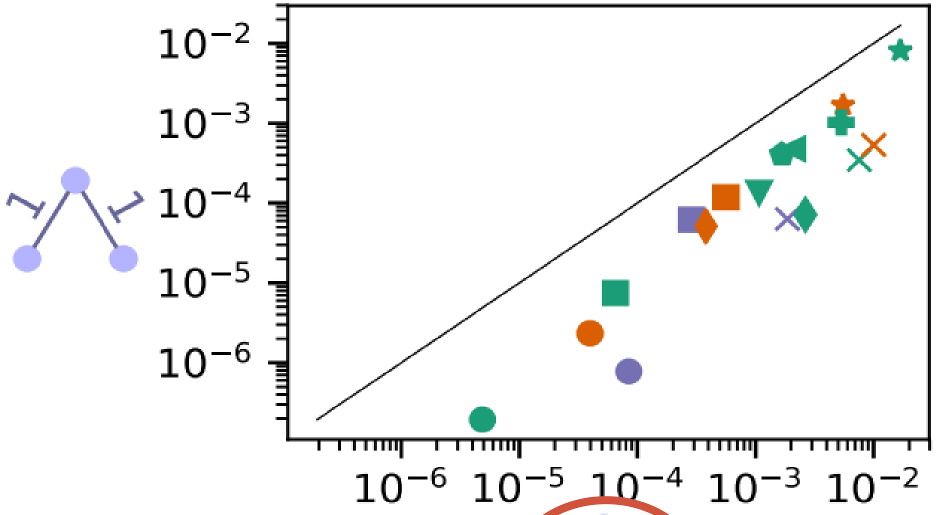


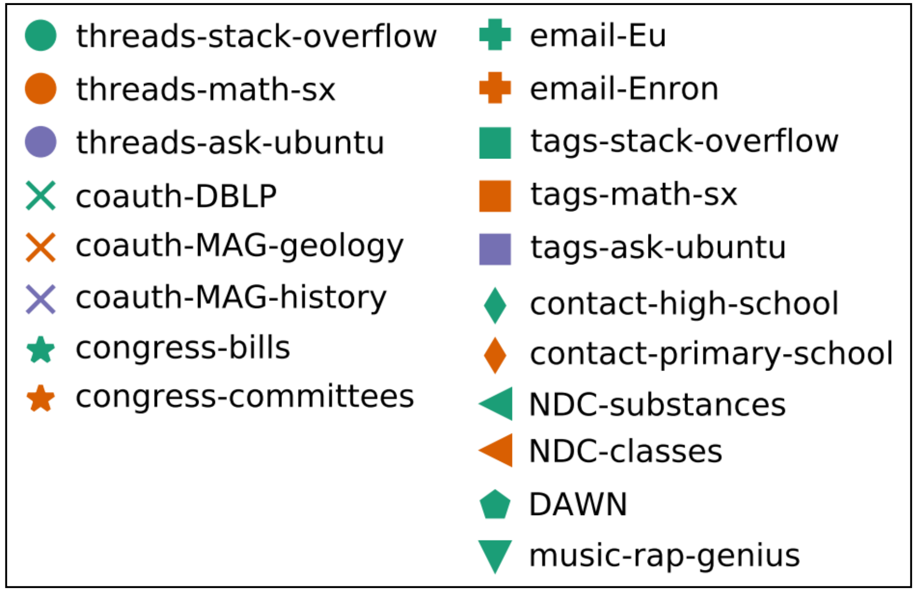
Closed triangle

A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, "Simplicial Closure and Higher-order Link Prediction", **PNAS (2018)**

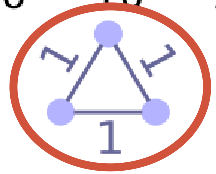
Edge Density vs. Simplicial Closure

- Increased **edge density** increases simplicial closure probability.



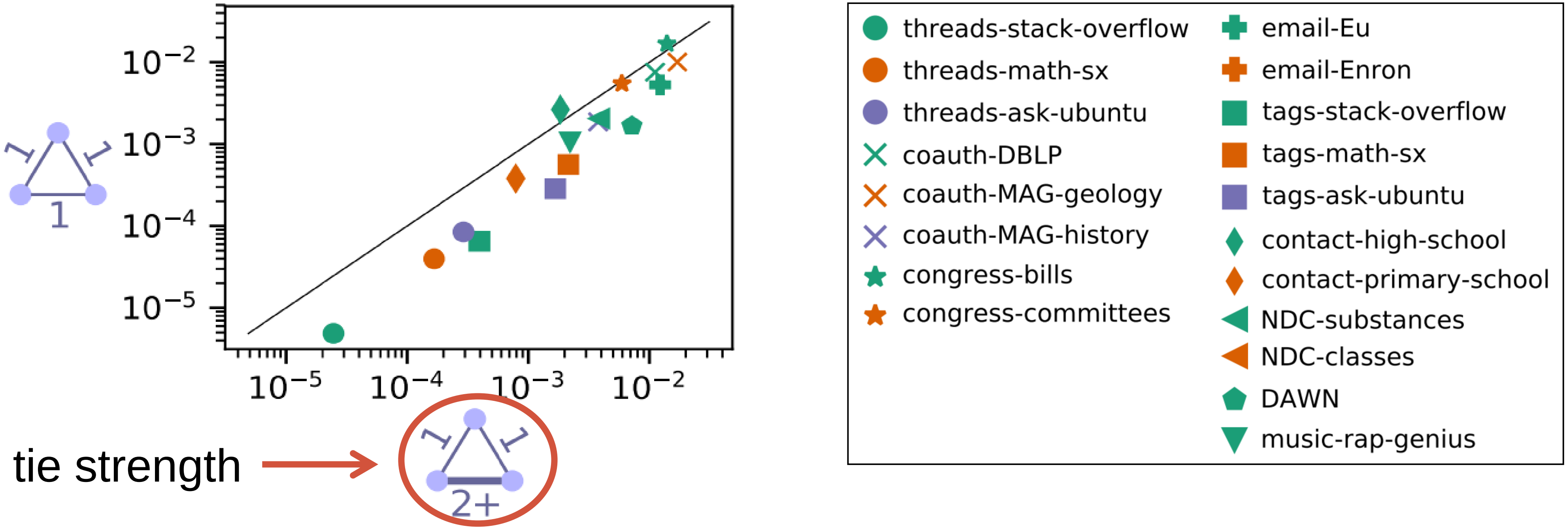


Higher edge density



Tie Strength vs. Simplicial Closure

- Increased **tie strength** increases simplicial closure probability.

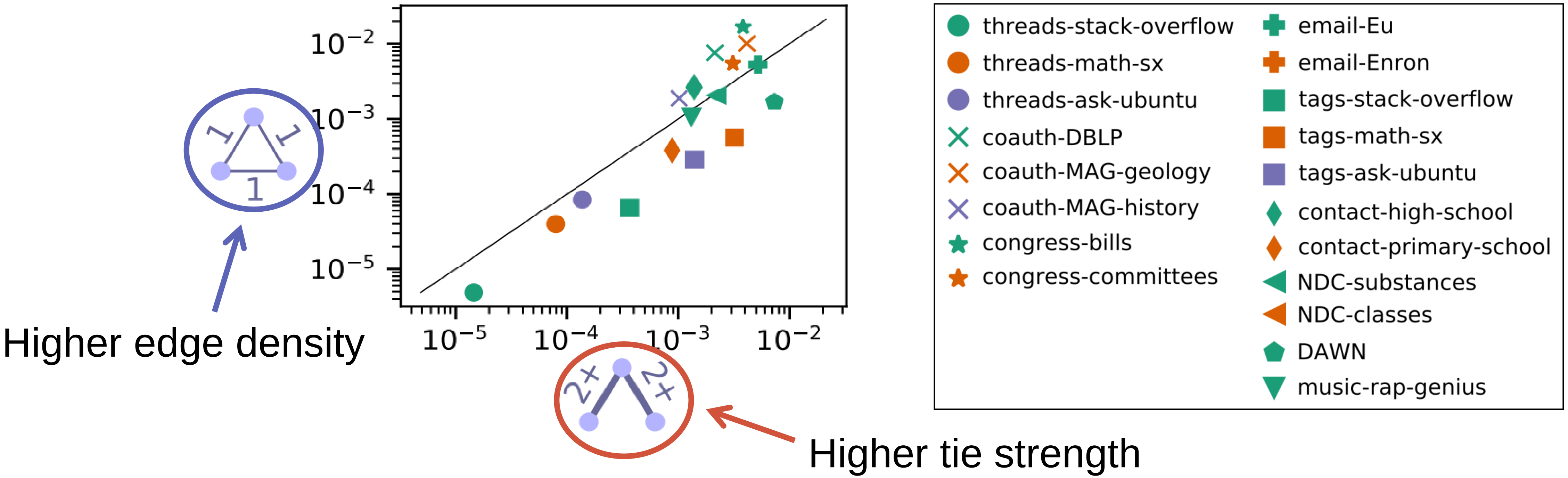


Higher tie strength

A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, "Simplicial Closure and Higher-order Link Prediction", **PNAS (2018)**

Edge Density vs. Tie Strength

- Relative importance of **edge density** and **tie strength** depends on the dataset.



A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, "Simplicial Closure and Higher-order Link Prediction", **PNAS (2018)**

CS22: One Advanced Dynamic Pattern

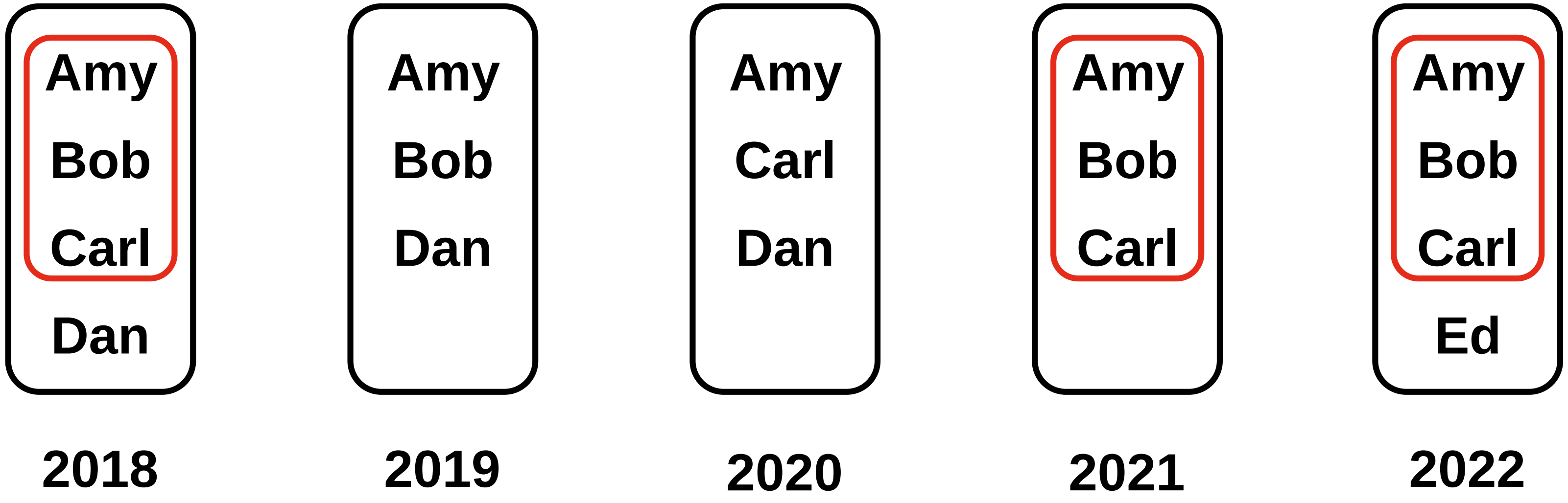
- **P1.** Persistence of higher-order interactions (HOIs)

	Node		
	Hyperedge		
	Hypergraph		
	Sub-hypergraph		P1

H. Choo, K. Shin, "On the Persistence of Higher-order Interactions in Real-world Hypergraphs", **SDM 2022**

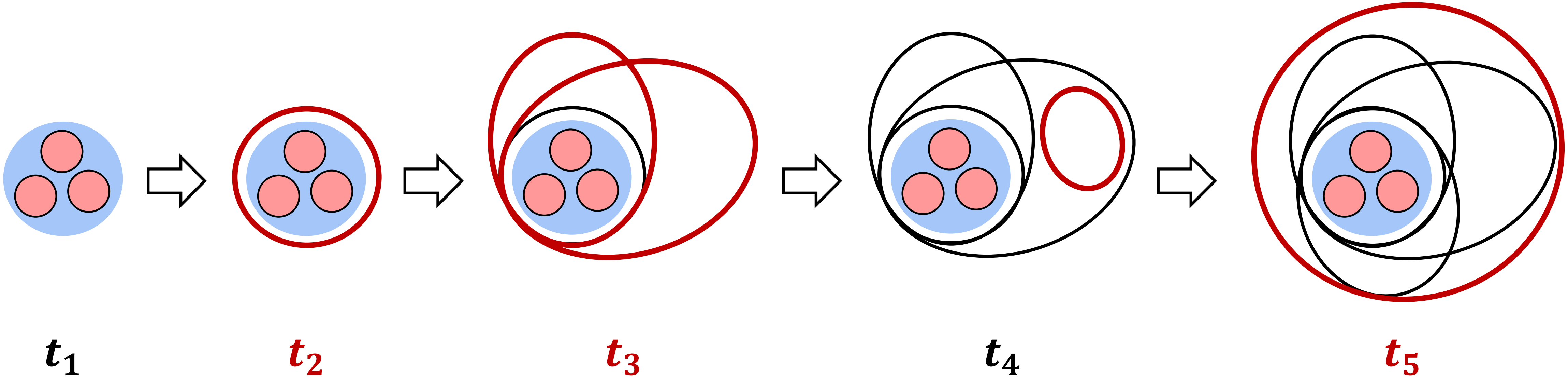
Higher-Order Interaction

- A **higher-order interaction (HOI)** is the co-appearance of a set of nodes in any hyperedge.
- HOIs can appear **repeatedly** over time.



Persistence of HOIs

- **Persistence** of an HOI $\boldsymbol{S}$ over time range $\boldsymbol{T}$ is the number of time units in $\boldsymbol{T}$ when $\boldsymbol{S}$ co-appears in any hyperedge.
- Example: **Persistence = 3**



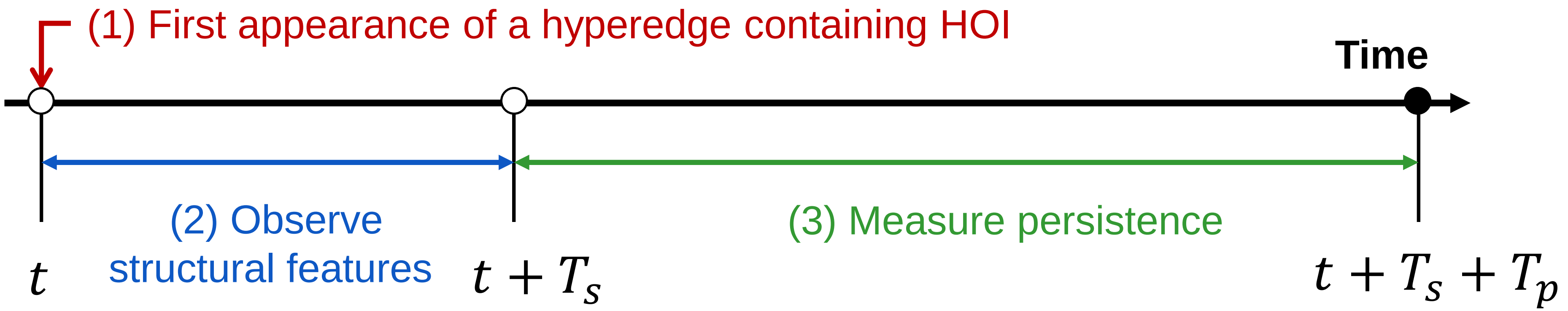
Group Features and Persistence

Questions:

What structural features are related to HOI persistence?

Answer:

We examine relations between various structural features and HOI persistence with the following setting:



Group Features and Persistence (cont.)

- Basic structural features of each HOI S:
  - $\#$: number of hyperedges including S
  - Σ: sum of sizes of hyperedges containing S
  - $\cup$: number of hyperedges overlapping S
  - $\Sigma \cup$: sum of sizes of hyperedges overlapping S
  - $\cap$: number of common neighbors of S
  - $\mathcal{H}$: entropy in sizes of hyperedges containing S
- Group structural features of each HOI S:

(1) $\#$, (2) $\#/\cup$, (3) $\Sigma/(\Sigma \cup)$, (4) $\cap$, (5) $\#/\cap$, (6) $\Sigma/\cap$, (7) $\Sigma/\#$, (8) $\mathcal{H}$

Group Features and Persistence (cont.)

- HOI persistence is **positively correlated** with:
  - The **number of hyperedges** containing the HOI
  - The **entropy of the sizes of hyperedges** containing the HOI

Normalized mutual information → MI

Pearson correlation coefficient → CC

	Size of HOIs	$\#$	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma \cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$
MI	2	0.13	0.11	<u>0.14</u>	0.05	0.10	0.12	0.10	**0.15**
	3	<u>0.11</u>	0.06	0.08	0.05	0.08	0.09	0.08	**0.12**
	4	<u>0.11</u>	0.05	0.07	0.06	0.07	0.10	0.07	**0.12**
	Avg.	<u>0.12</u>	0.08	0.10	0.05	0.08	0.11	0.08	**0.13**
CC	2	**0.36**	0.09	0.09	0.17	0.19	0.26	-0.08	<u>0.32</u>
	3	**0.31**	0.10	0.10	0.05	0.16	0.20	-0.09	<u>0.25</u>
	4	**0.30**	0.13	0.13	-0.01	0.17	0.20	-0.10	<u>0.24</u>
	Avg.	**0.32**	0.10	0.11	0.07	0.17	0.22	-0.09	<u>0.27</u>

H. Choo, K. Shin, "On the Persistence of Higher-order Interactions in Real-world Hypergraphs", **SDM 2022**

LS21: One Advanced Dynamic Pattern

- **P1.** Temporal hypergraph motifs (TH-motifs)

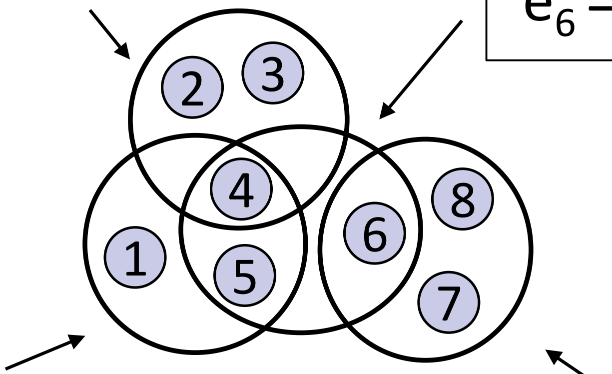
	Node		
	Hyperedge		
	Hypergraph		
	Sub-hypergraph		P1

Temporal Hypergraph Motifs: Definition

*"How can we define **motifs** in **temporal hypergraphs**?"*

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

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



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

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

Temporal Hypergraph Motifs: Definition (cont.)

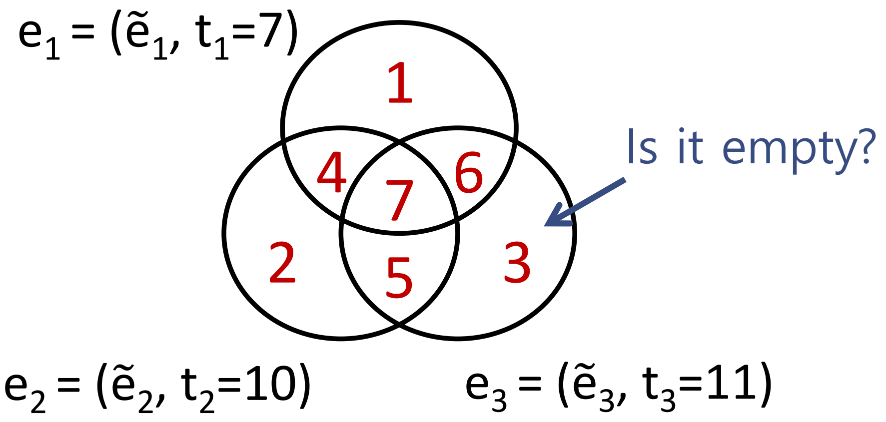
?

Question:

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

Answer:

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



!

Temporal Hypergraph Motifs: Definition (cont.)

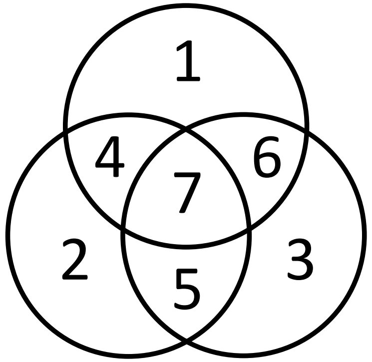
Question:

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

Answer:

The three temporal hyperedges should **arrive within δ time**.

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



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

$$\mathbf{max}(\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3}) - \mathbf{min}(\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3}) \leq \boldsymbol{\delta}$$

Temporal Hypergraph Motifs: Definition (cont.)

?

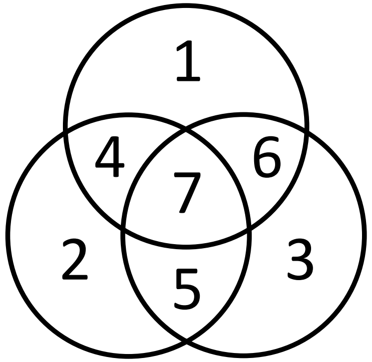
Question:

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

Answer:

The **order** of the three temporal hyperedges is considered.

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



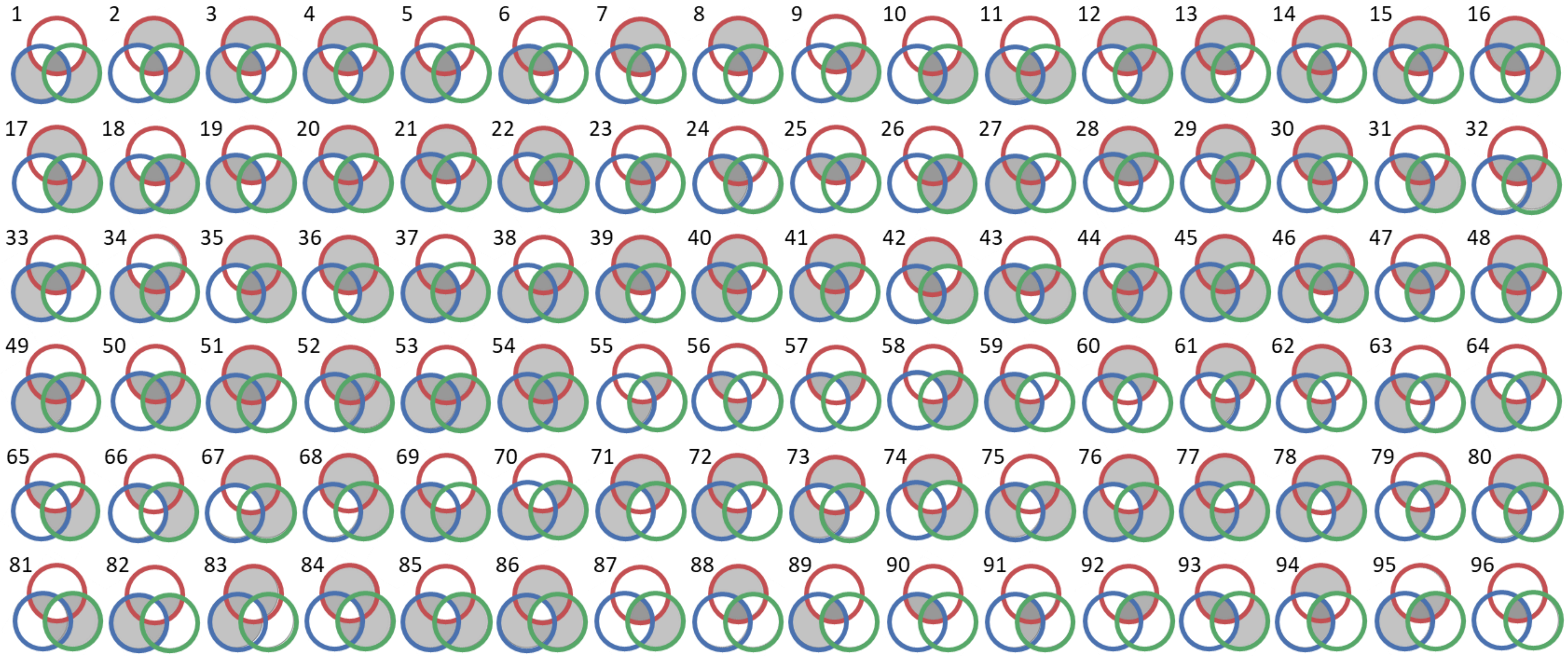
$e_2 = (\tilde{e}_2, t_2=10)$ $e_3 = (\tilde{e}_3, t_3=11)$

!

$\mathbf{e_1 \rightarrow e_2 \rightarrow e_3}$

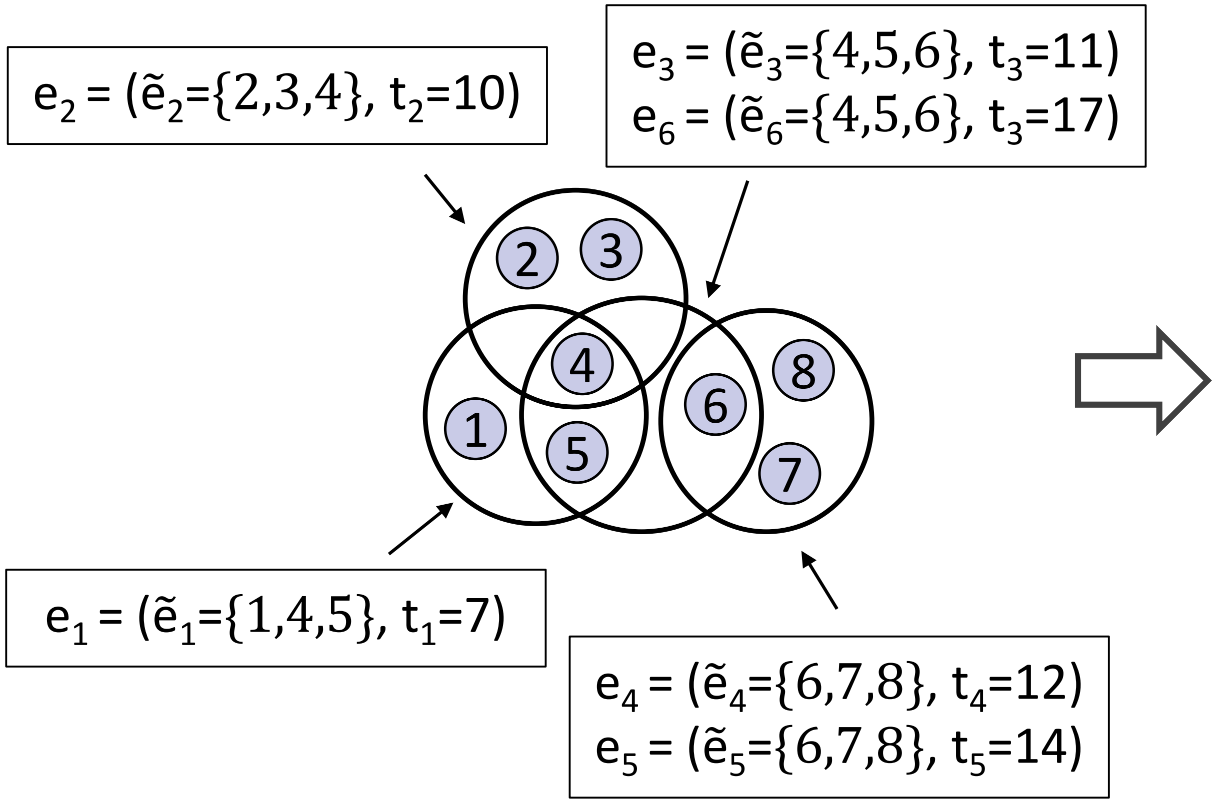
Temporal Hypergraph Motifs: Definition (cont.)

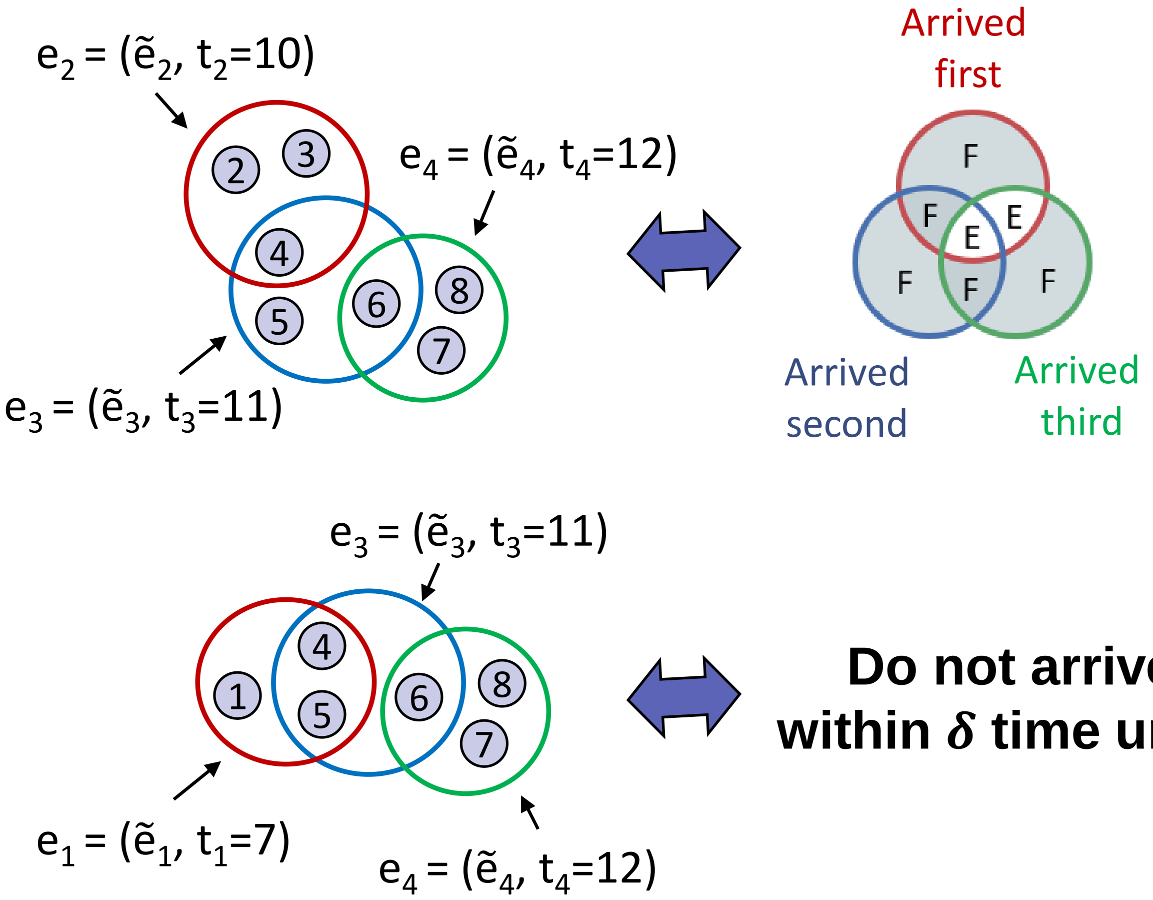
- There exist **96 temporal hypergraph motifs (TH-motifs)**.



Temporal Hypergraph Motifs: Example

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



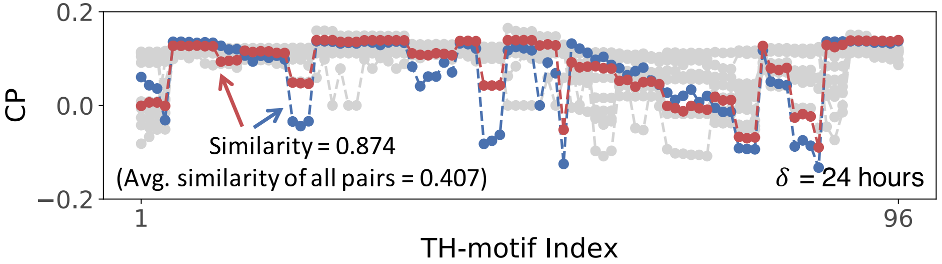


Do not arrive within δ time units!

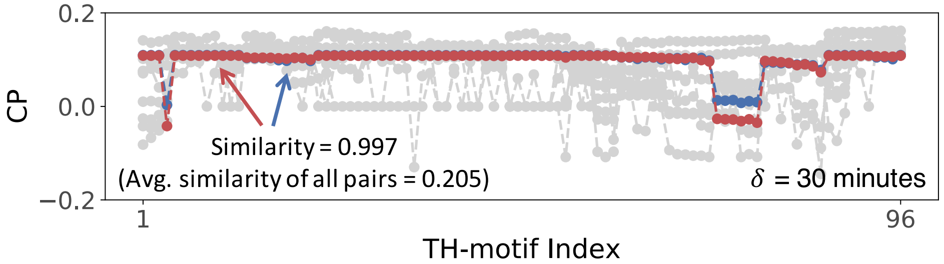
Comparison across Domains

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

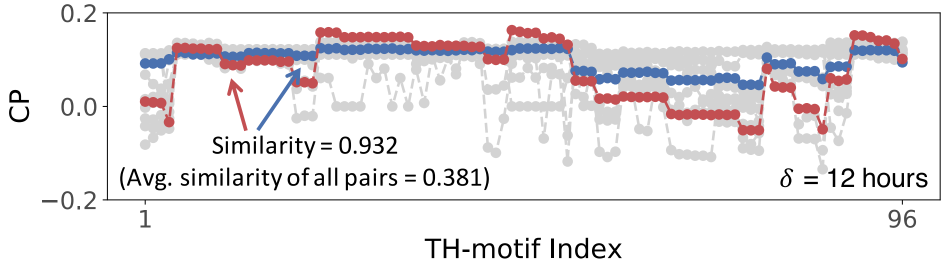
Email domain



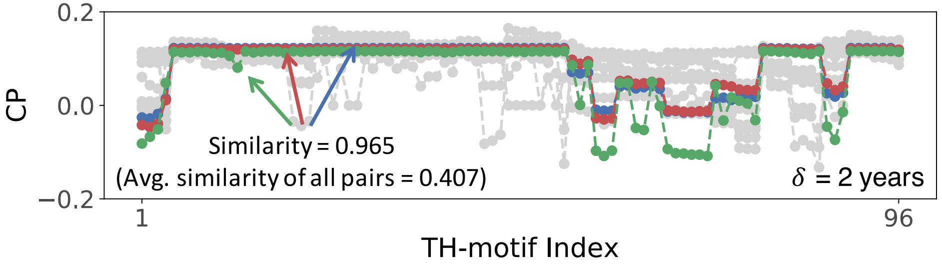
Contact domain



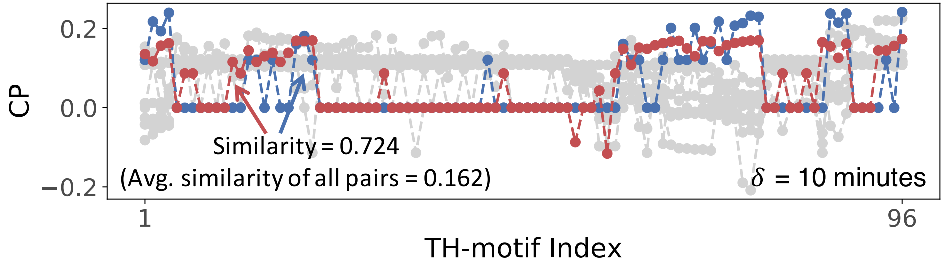
Tags domain



Coauthorship domain

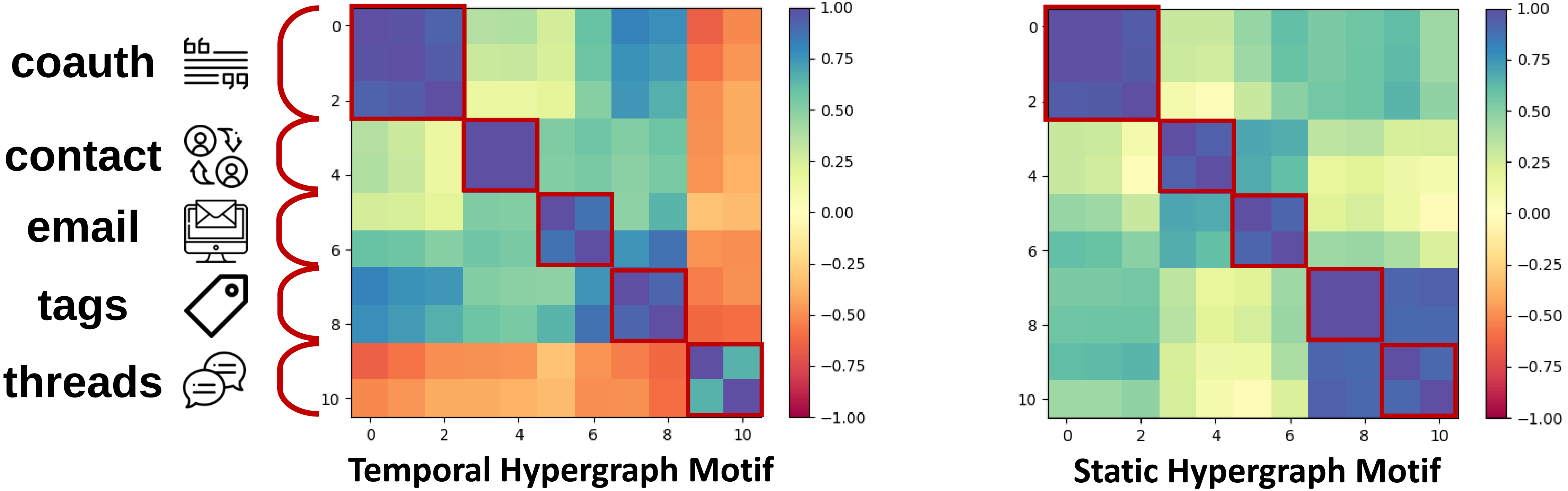


Threads domain



Comparison across Domains (cont.)

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



Temporal Hypergraph Motif

- Within-domain: 0.900 Gap: **0.759**
- Between-domain: 0.141 Times: **6.38X**

Static Hypergraph Motif

- Within-domain: 0.951 Gap: 0.517
- Between-domain: 0.434 Times: 2.19X

CK21: Four Advanced Dynamic Patterns

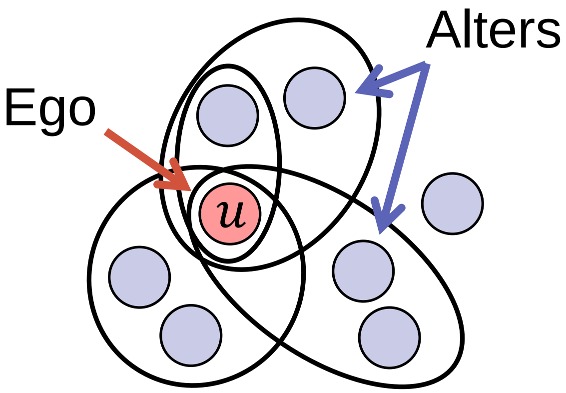
- **P1.** Intersection size of ego-networks
- **P2.** Spread of alter-networks
- **P3.** Anthropic principles of ego-networks
- **P4.** Novelty rate of ego-networks

	Node		
	Hyperedge		
	Hypergraph		
	Sub-hypergraph		P1, P2, P3, P4

Hypergraph Ego-networks

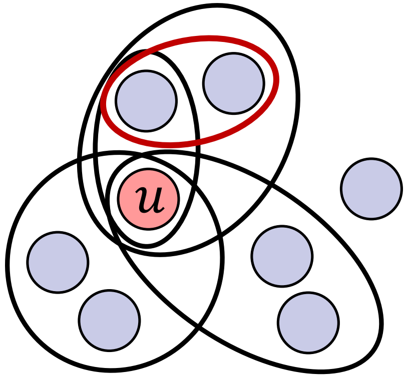
- There are three types of **hypergraph ego-networks**.

Star ego-network



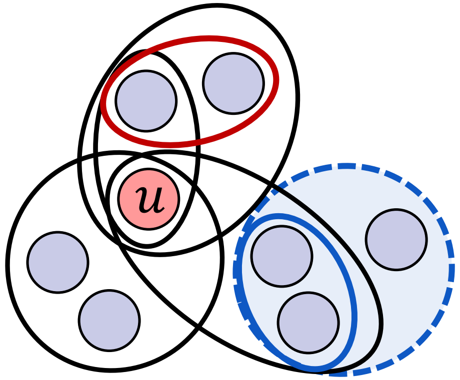
Hyperedges containing the ego-node.

Radial ego-network



Hyperedges containing ego-node **or** its neighbors.

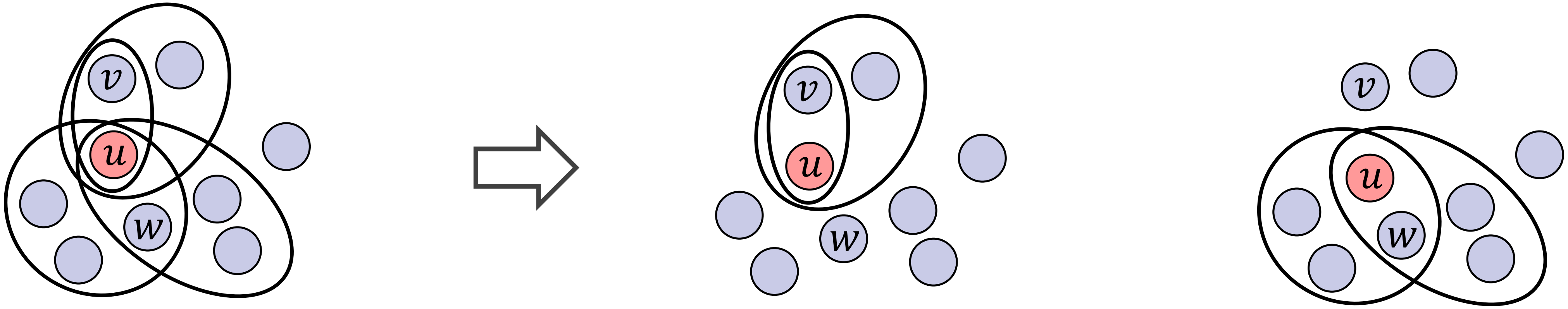
Contracted ego-network



Hyperedges containing ego-node or its neighbors **+ sub-hyperedges** containing them

Hypergraph Alter-networks

- An **alter-network** of node $\boldsymbol{v}$ in **ego-network** of node $\boldsymbol{u}$ is the collection of hyperedges in the ego-network that contain $\boldsymbol{v}$.



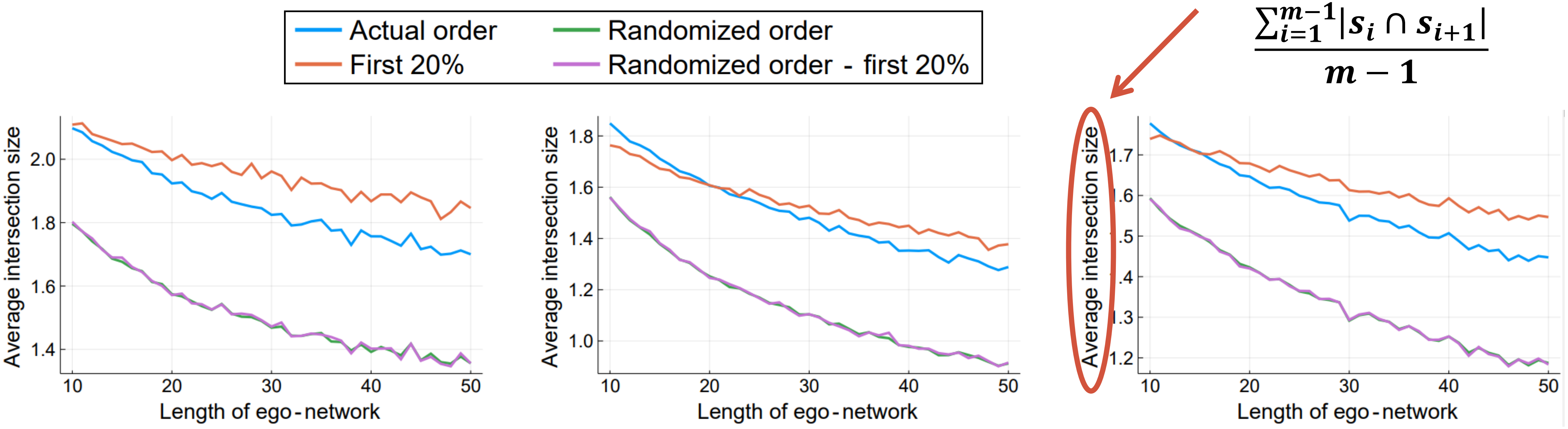
Ego-network of $\boldsymbol{u}$ **Alter-network of $\boldsymbol{v}$** **Alter-network of $\boldsymbol{w}$**

Intersection Size of Ego-networks

- **Temporally adjacent** hyperedges are similar.

Average intersection size of ego-network $\{s_1, \dots, s_m\}$:

$$\frac{\sum_{i=1}^{m-1} |s_i \cap s_{i+1}|}{m-1}$$



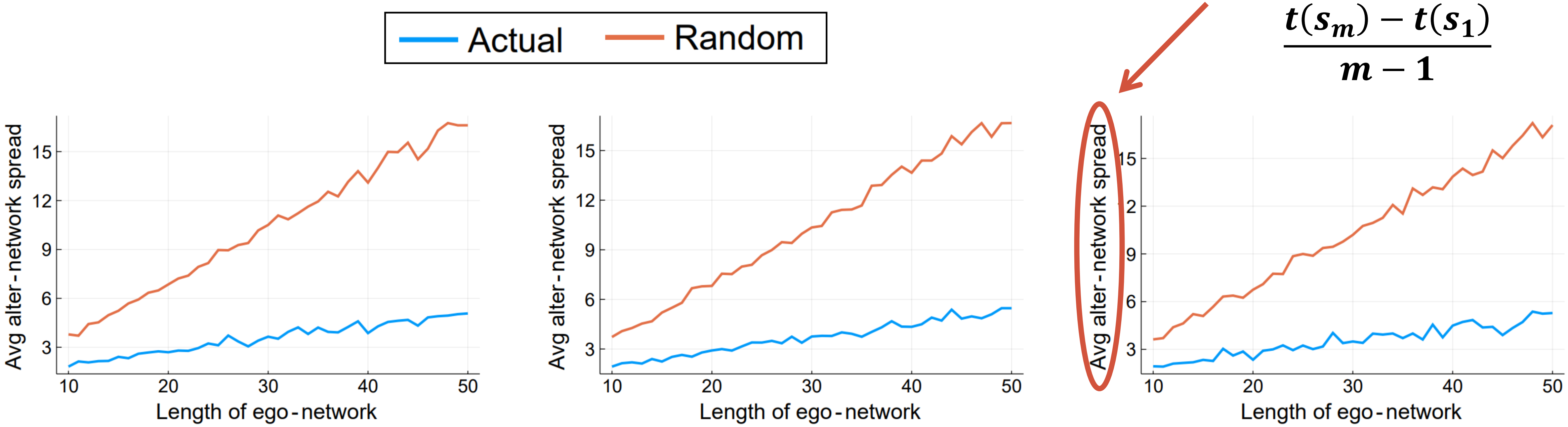
Star ego-network **Radial ego-network** **Contracted ego-network**

Spread of Alter-networks

- Spread of **alter networks** is temporally local.

Average spread time of alter-network $\{s_i, \dots, s_m\}$:

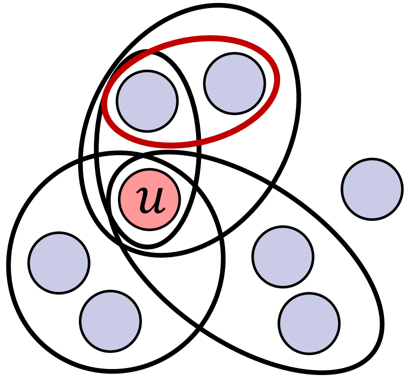
$$\frac{t(s_m) - t(s_1)}{m-1}$$



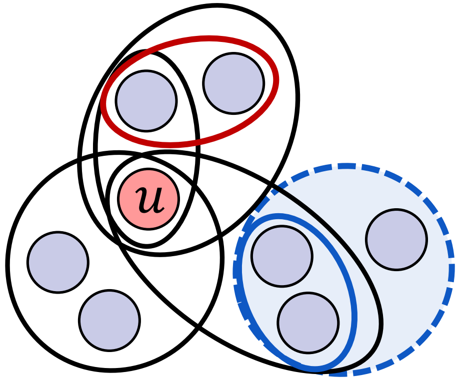
Star ego-network **Radial ego-network** **Contracted ego-network**

Anthropic Principle of Ego-networks

- **Radial & contracted ego-networks** may include hyperedges without ego-node u.
- These hyperedges may arrive **before** the ego-node entered it.



Radial ego-network



Contracted ego-network

Anthropic Principle of Ego-networks (cont.)

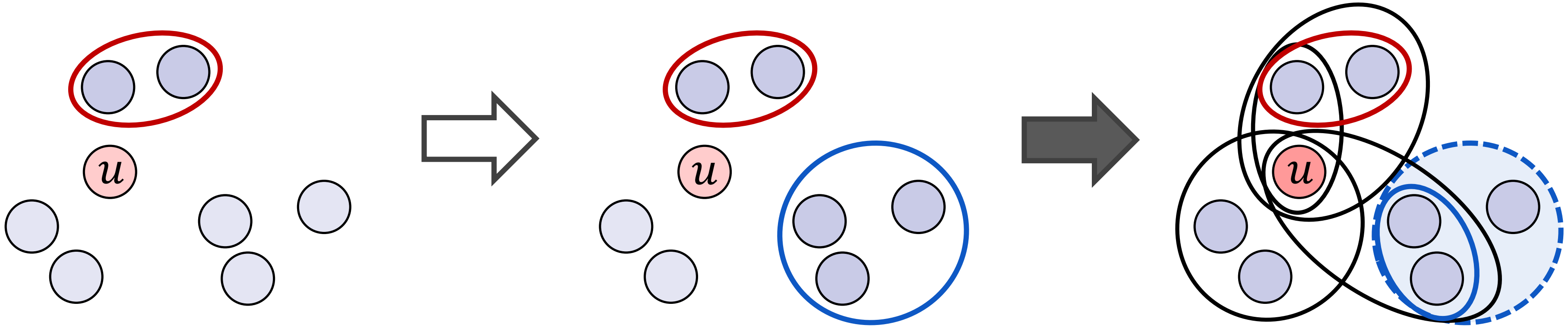
?

Question:

Across real-world radial & contracted ego-networks, at what timestamp does the ego-node arrive?

Answer:

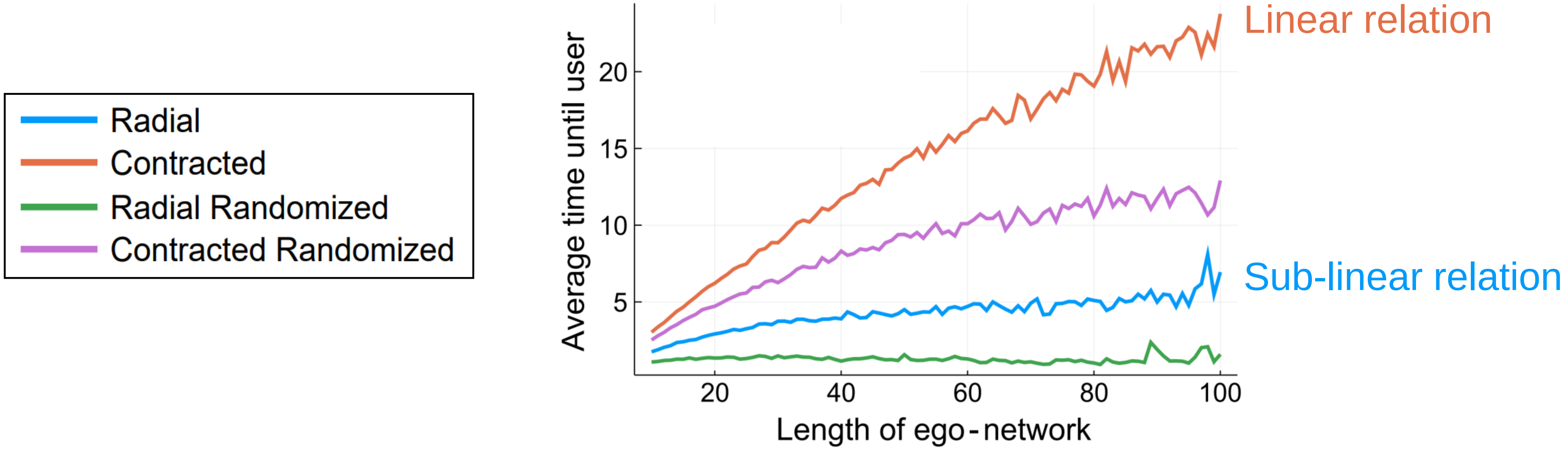
Timestamps of the ego-node arrival are observed.



!

Anthropic Principle of Ego-networks (cont.)

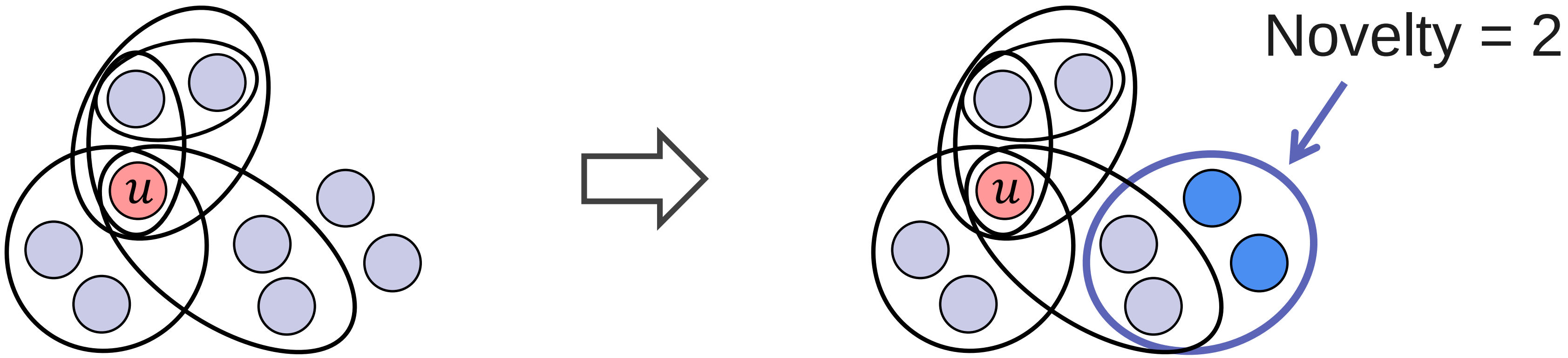
- The **ego-node** tends to arrive…
  - before or around the fifth timestamp in **radial ego-networks**,
  - at linear timestamp to the size in **contracted ego-networks**.



C. Comrie, J. Kleinberg, "Hypergraph Ego-networks and Their Temporal Evolution", **ICDM 2021**

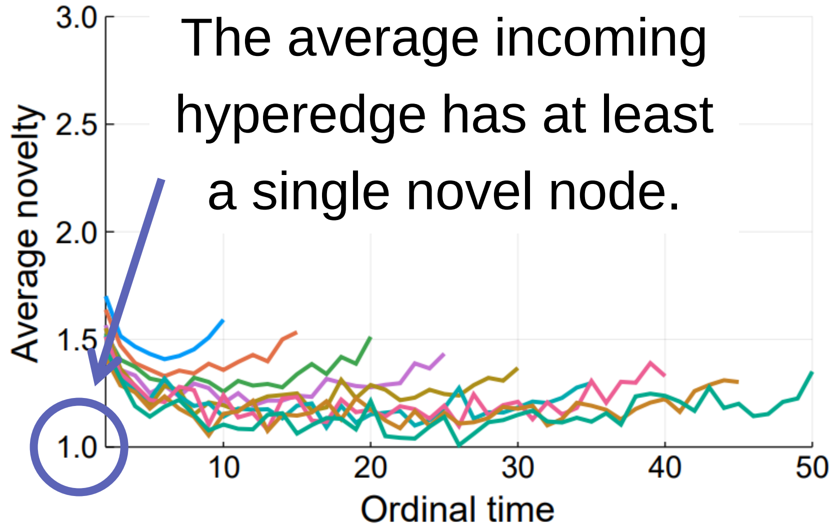
Novelty Rate of Ego-networks

- **Novelty** of a hyperedge is the number of nodes in the hyperedge that have never been contained in any previous hyperedges.

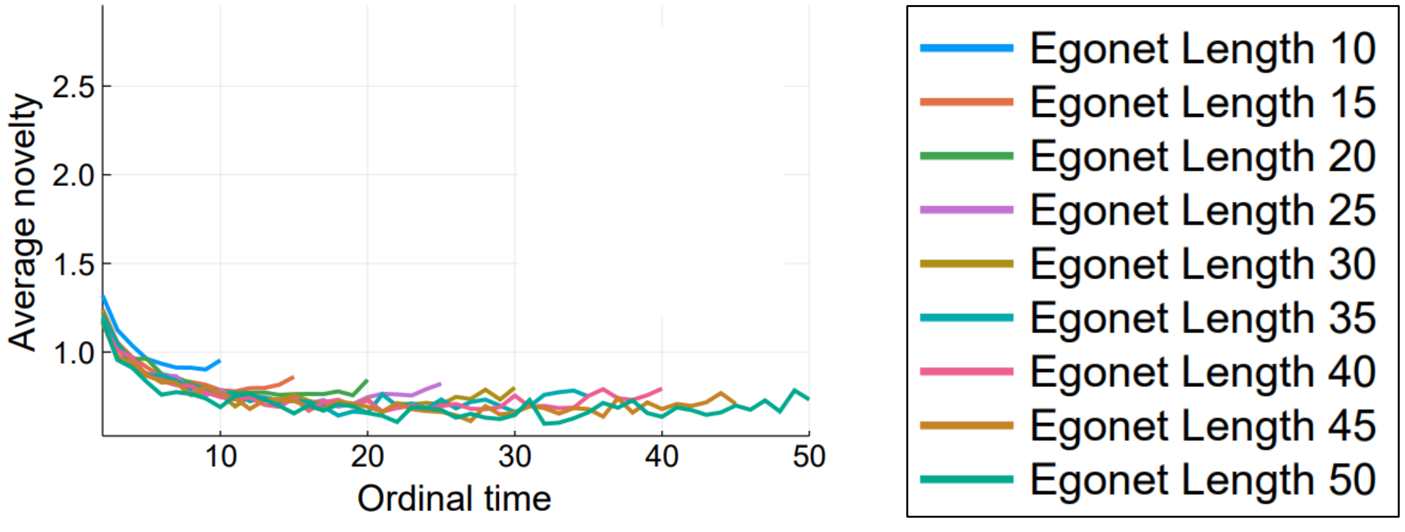


Novelty Rate of Ego-networks (cont.)

- Novelty slowly & gradually **decreases** in **star & radial ego-networks**.



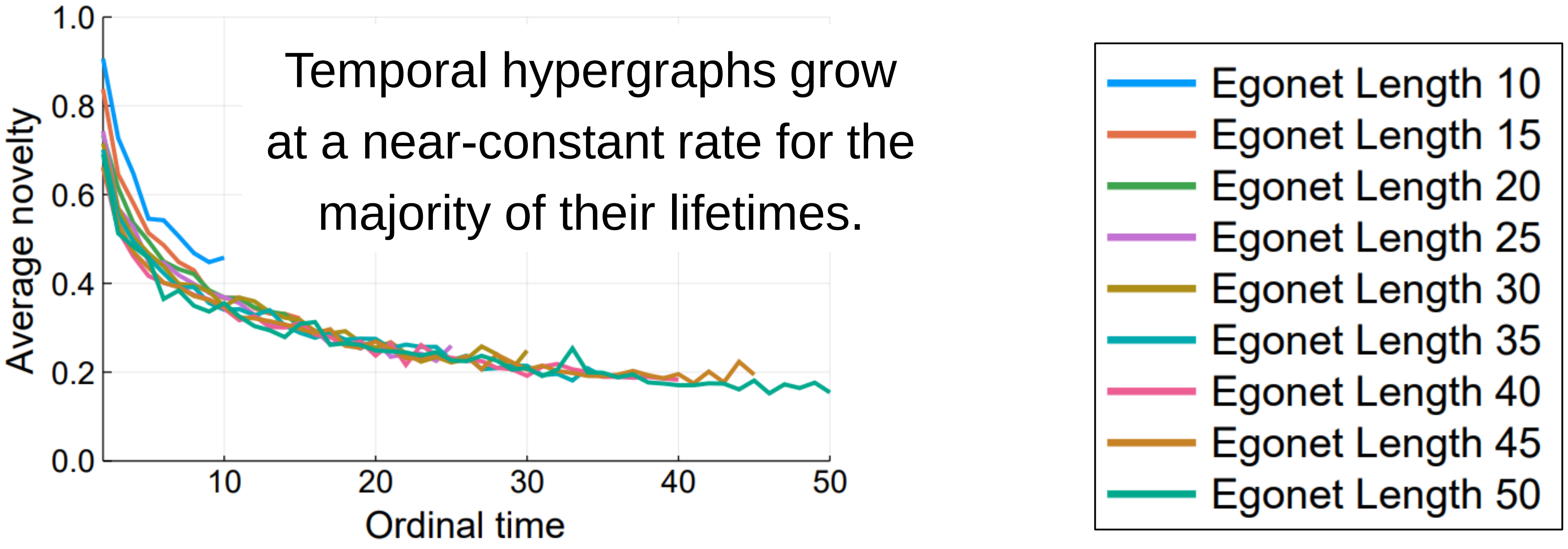
Star ego-network



Radial ego-network

Novelty Rate of Ego-networks (cont.)

- Novelty always **decreases** in **contracted ego-networks**.



Contracted ego-network

C. Comrie, J. Kleinberg, "Hypergraph Ego-networks and Their Temporal Evolution", **ICDM 2021**

References

1. [BASJK18] Benson, Austin R., et al. "Simplicial Closure and Higher-order Link Prediction." PNAS 115(48):E11221–E11230, 2018.
2. [BKT18] Benson, Austin R., Ravi Kumar, and Andrew Tomkins, "Sequences of Sets." KDD 2018.
3. [CBLK21] Cencetti, Giulia, et al. "Temporal Properties of Higher-order Interactions in Social Networks." Scientific reports 11(1):1–10, 2021.
4. [CK21] Comrie, Cazamere, and Jon Kleinberg. "Hypergraph Ego-networks and Their Temporal Evolution." ICDM 2021.
5. [CS22] Choo, Hyunjin, and Kijung Shin. "On the Persistence of Higher-order Interactions in Real-world Hypergraphs." SDM 2022.
6. [KKS20] Kook, Yunbum, Jihoon Ko, and Kijung Shin. "Evolution of Real-world Hypergraphs: Patterns and Models without Oracles." ICDM 2020.

References (cont.)

7. [LS21] Lee, Geon, and Kijung Shin. “THyMe+: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting.” ICDM 2021.

8. [GLLB23] Gallo, Luca, et al. “Higher-order Correlations Reveal Complex Memory in Temporal Hypergraphs. arXiv (2023).