AUGUST 6 - 10
KDD2023
LONG BEACH, CA
29TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING

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

## Part 1. Static Structural Patterns

Geon Lee

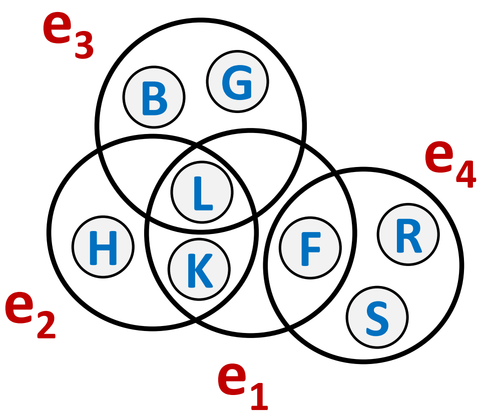
Jaemin Yoo

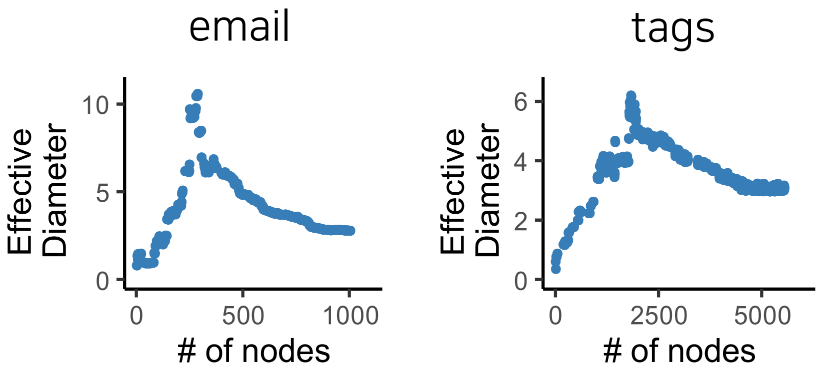
Kijung Shin

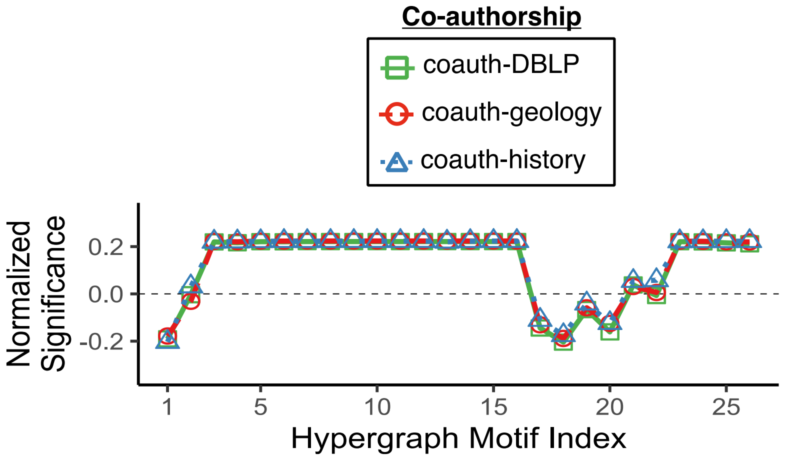
# Part 1. Static Structural Patterns

*“What do real-world hypergraphs look like?”*

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







**Input Hypergraph**

**Basic Patterns (Part 1-1)**

**Advanced Patterns (Part 1-2)**

# Roadmap

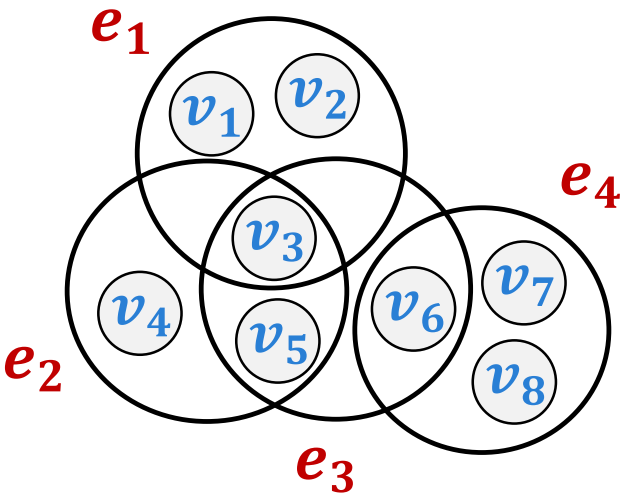
- **Part 1. Static Structural Patterns**
  - Basic Patterns **<<**
  - Advanced Patterns
- **Part 2. Dynamic Structural Patterns**
  - Basic Patterns
  - Advanced Patterns
- **Part 3. Generative Models**
  - Static hypergraph Generator
  - Dynamic hypergraph Generator

# Part 1-1. Basic Static 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 |

# Background

- **Degree** of a node $v$ is the number of hyperedges containing $v$.
- **Size** of a hyperedge $e$ is the number of nodes in $e$.



**Hypergraph**

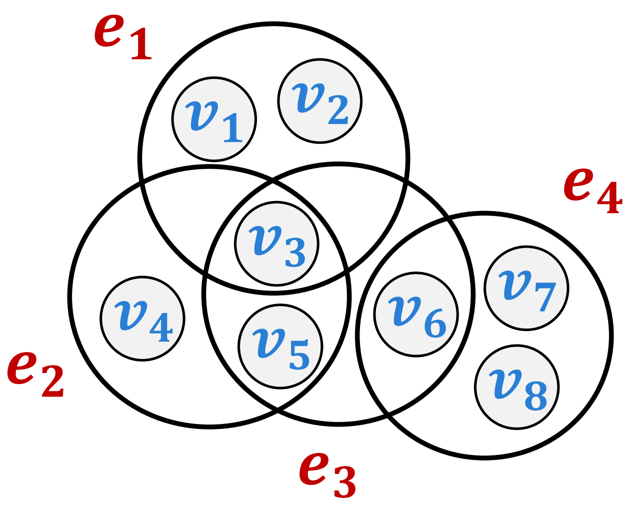
**Degree** of $v_5$ is 2.

**Size** of $e_2$ is 3.

**Example**

# Background (cont.)

- **Incidence matrix** $\boldsymbol{H} = \{\mathbf{0}, \mathbf{1}\}^{|\boldsymbol{V}|\times|\boldsymbol{E}|}$ of a hypergraph $\boldsymbol{\mathcal{G}} = (\boldsymbol{V}, \boldsymbol{E})$ is:



**Hypergraph**

$$\begin{array}{c|cccc} & e_1 & e_2 & e_3 & e_4 \\ \hline v_1 & 1 & 0 & 0 & 0 \\ v_2 & 1 & 0 & 0 & 0 \\ v_3 & 1 & 1 & 1 & 0 \\ v_4 & 0 & 1 & 0 & 0 \\ v_5 & 0 & 1 & 1 & 0 \\ v_6 & 0 & 0 & 1 & 1 \\ v_7 & 0 & 0 & 0 & 1 \\ v_8 & 0 & 0 & 0 & 1 \end{array}$$

$$\boldsymbol{H}[\boldsymbol{i}][\boldsymbol{j}] = \begin{cases} \mathbf{1}, & \text{if } \boldsymbol{v_i} \in \boldsymbol{e_j} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

**Incidence matrix**

# KKS20: Three Basic Static Patterns

- **P1.** Degree distribution
- **P2.** Hyperedge size distribution
- **P3.** Intersection size distribution

| | | ⚙ | ⚙ |
|---|---|---|---|
| 📖 | **Node** | P1 | |
| | **Hyperedge** | P2, P3 | |
| | **Hypergraph** | | |
| 💡 | **Sub-hypergraph** | | |

Y. Kook, J. Ko, K. Shin, “Evolution of Real-world Hypergraphs: Patterns and Models without Oracles”, **ICDM 2020**

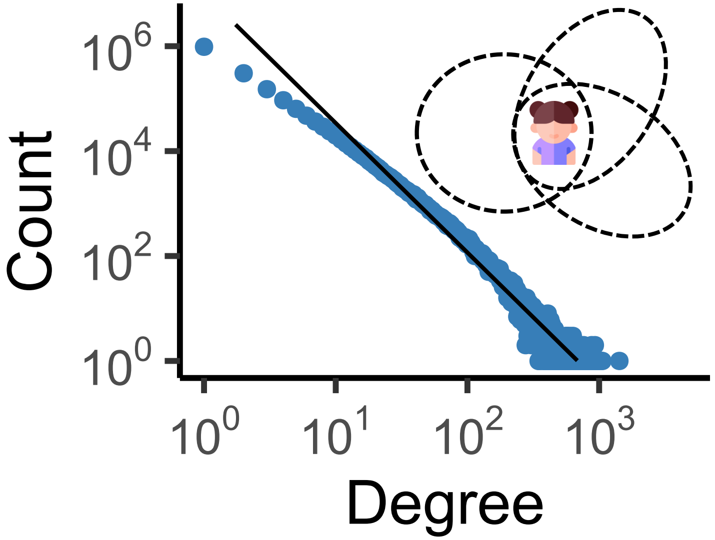
# Simple Questions

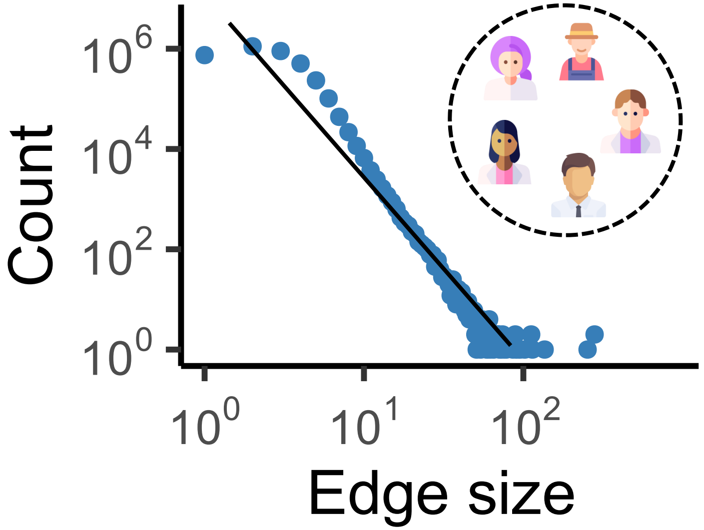
- How many groups does a person belong to?
- How many people are in each group?
- How many people belong to two groups at the same time?

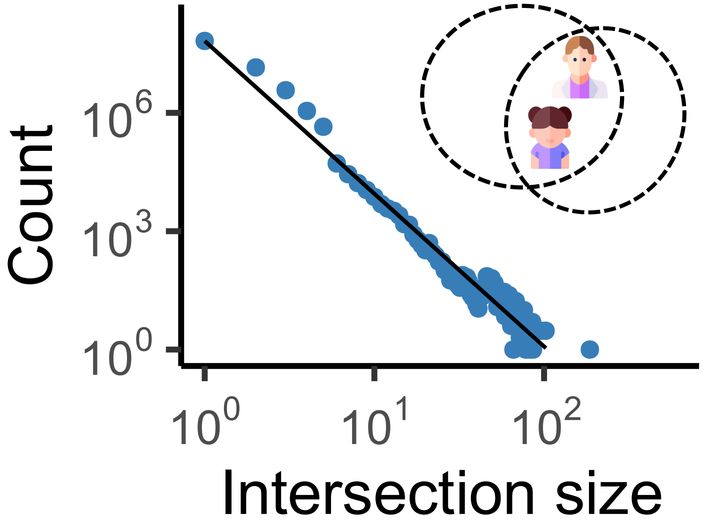
Y. Kook, J. Ko, K. Shin, “Evolution of Real-world Hypergraphs: Patterns and Models without Oracles”, **ICDM 2020**

# Answers to the Simple Questions

- Degree distributions of real-world hypergraphs are **heavy-tailed**.
- Size distributions of real-world hypergraphs are **heavy-tailed**.
- Intersection size distributions of real-world hypergraphs are **heavy-tailed**.



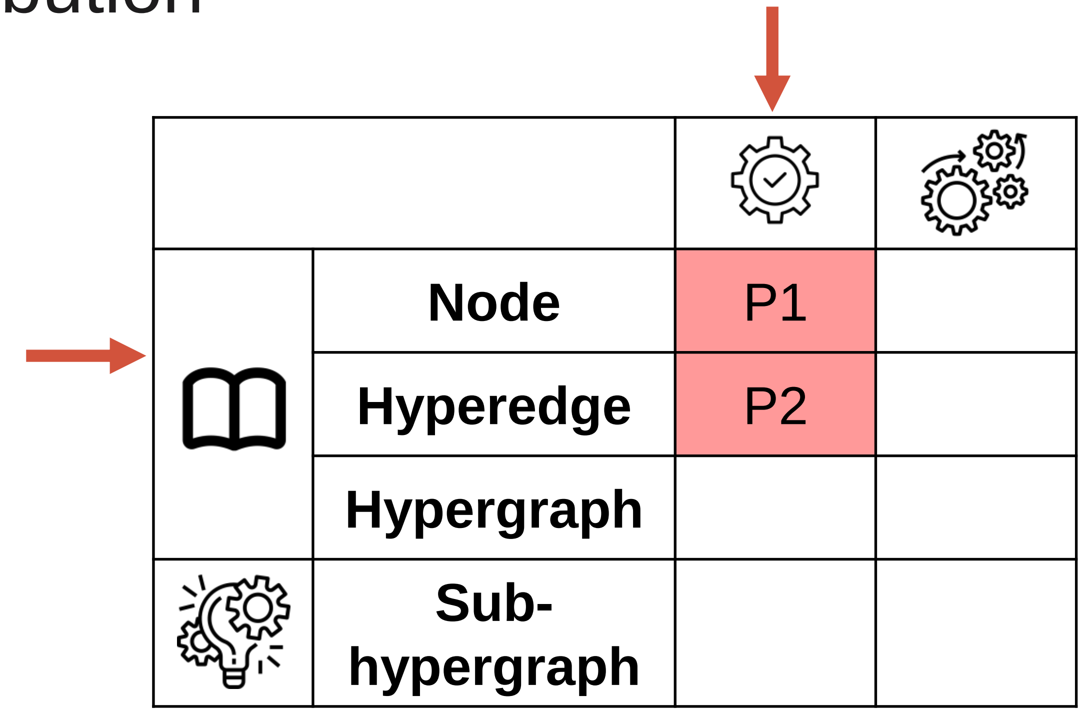




Y. Kook, J. Ko, K. Shin, “Evolution of Real-world Hypergraphs: Patterns and Models without Oracles”, **ICDM 2020**

# LCS21: Two Basic Static Patterns

- **P1.** Pair/triple-of-nodes degree distribution
- **P2.** Hyperedge homogeneity distribution

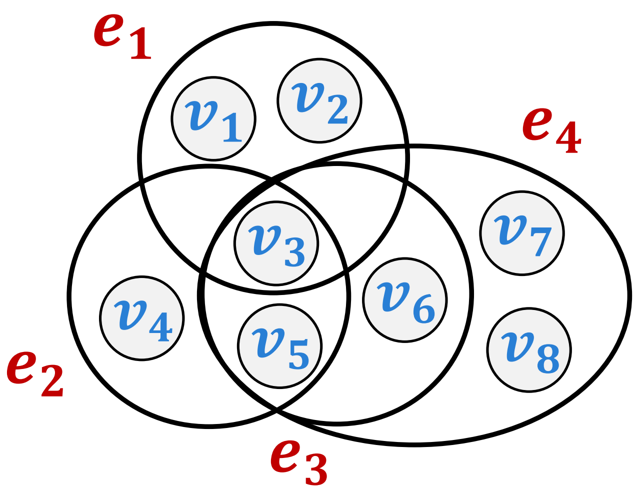


| | | ⚙ | ⚙ |
|---|---|---|---|
| 📖 | **Node** | 🟥 P1 | |
| | **Hyperedge** | 🟥 P2 | |
| | **Hypergraph** | | |
| 💡 | **Sub-hypergraph** | | |

G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Pair/Triple Degree Distribution

- **Degree of pair/triple of nodes** is the number of hyperedges overlapping the nodes.



**Degree** of $\{v_3, v_5\}$ is 3.

**Degree** of $\{v_3, v_5, v_6\}$ is 2.

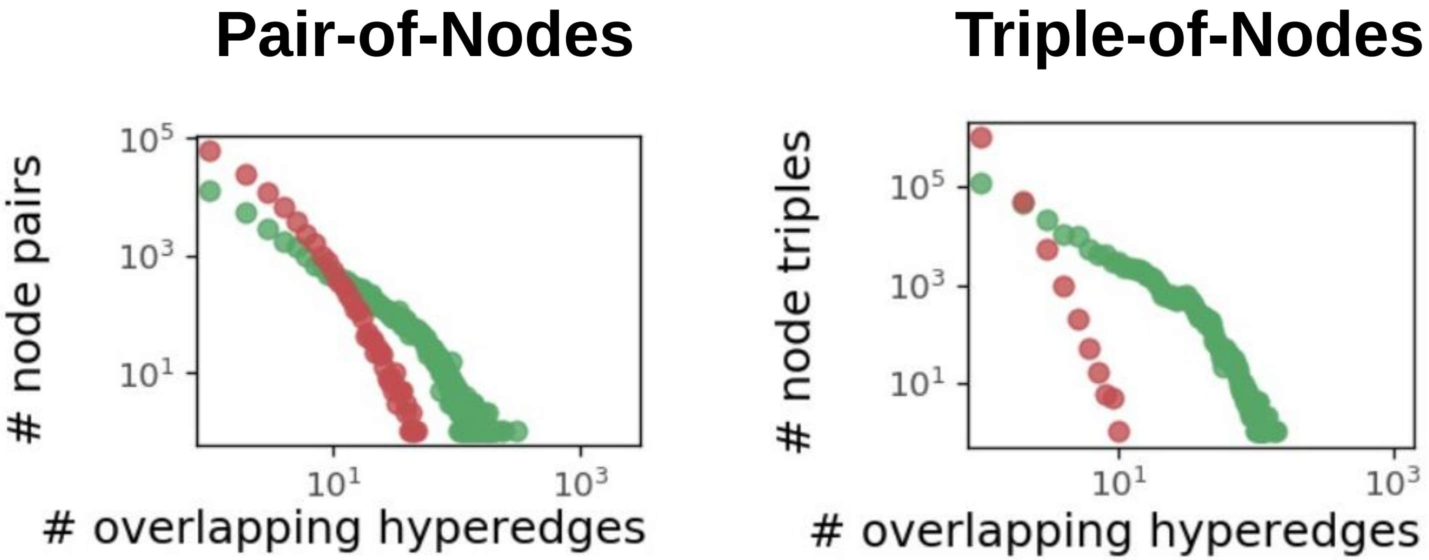
**Hypergraph**

**Example**

G. Lee, M. Choe, K. Shin, "How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators", **WWW 2021**

# Pair/Triple Degree Distribution (cont.)

- Degree distributions of pair/triple of nodes in real-world hypergraphs are **more skewed with a heavier tail** than those in randomized ones.



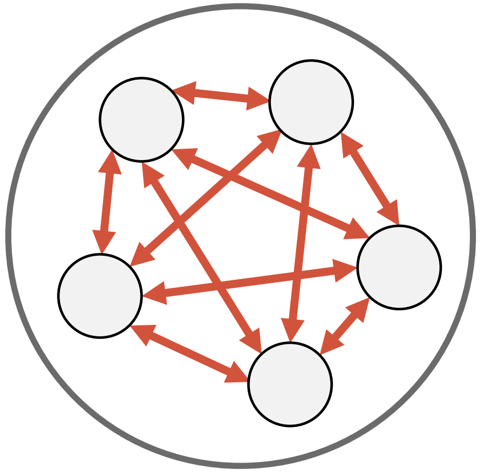
● **Real-world hypergraph**

● **Randomized hypergraph**

G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Hyperedge Homogeneity

- **Homogeneity** of a hyperedge $\boldsymbol{e}$ is the average number of hyperedges overlapping at all the pairs of nodes in the hyperedge.



**Hyperedge $\boldsymbol{e}$**

Number of hyperedges overlapping nodes $u$ and $v$.

$$\textbf{homogeneity}(\boldsymbol{e}) := \begin{cases} \dfrac{\sum_{\{u,v\}\in\binom{e}{2}} |E_{\{u,v\}}|}{\binom{|e|}{2}}, & \text{if } |\boldsymbol{e}| > \mathbf{1} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

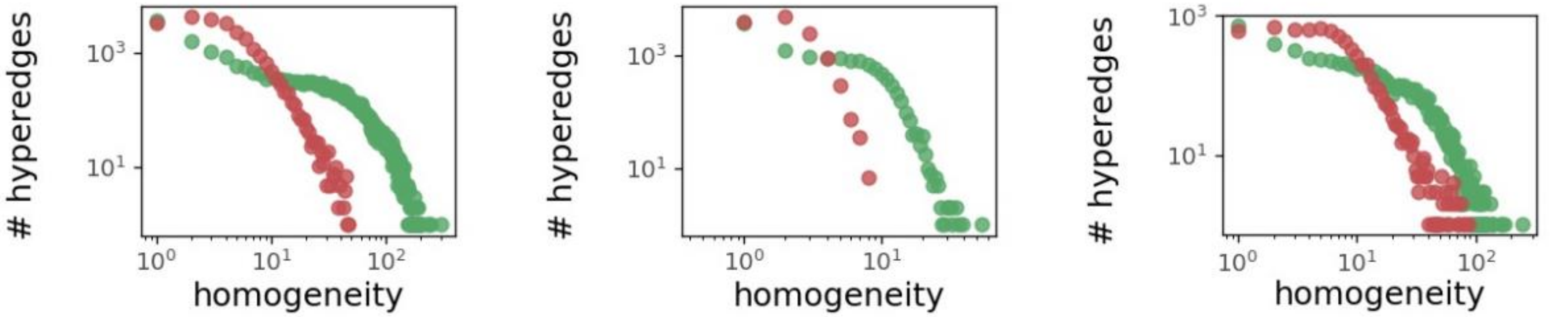
G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Hyperedge Homogeneity (cont.)

- Hyperedges in real-world hypergraphs tend to have **higher homogeneity** than those in randomized ones.

● Real-world hypergraph ● Randomized hypergraph

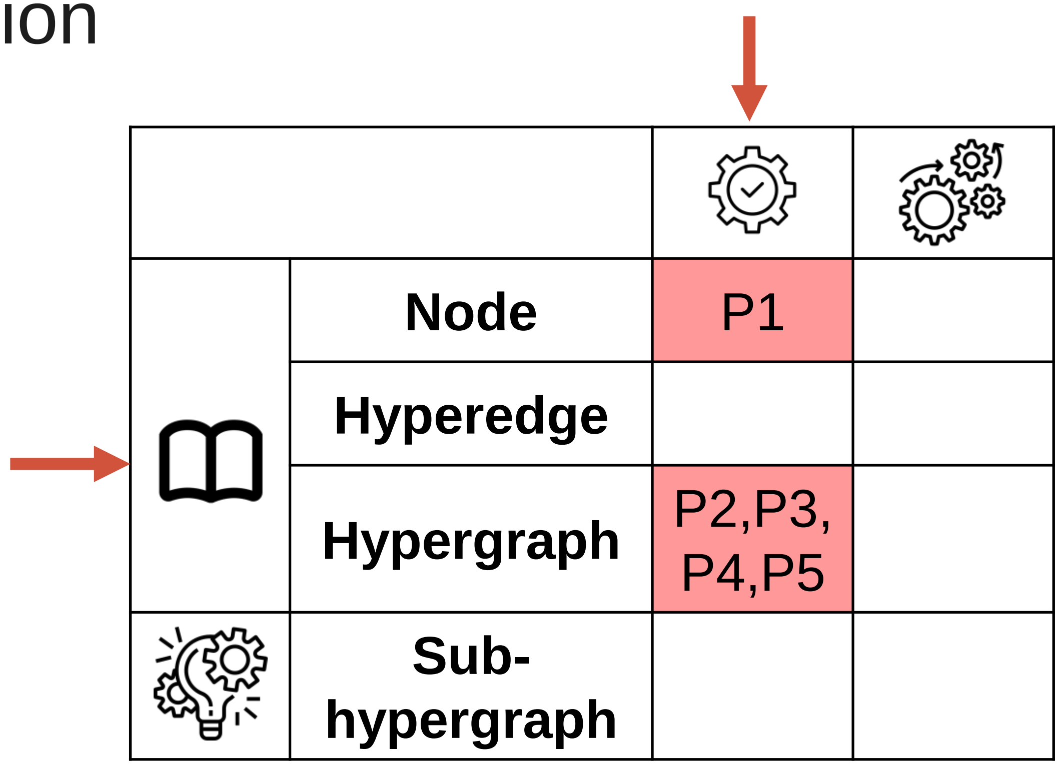
**email-Eu** **contact-primary** **NDC-substances**



G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# DYHS20: Five Basic Static Patterns

- **P1.** Heavy-tailed degree distribution
- **P2.** Skewed singular values distribution
- **P3.** Giant connected component
- **P4.** High clustering coefficient
- **P5.** Small effective diameter

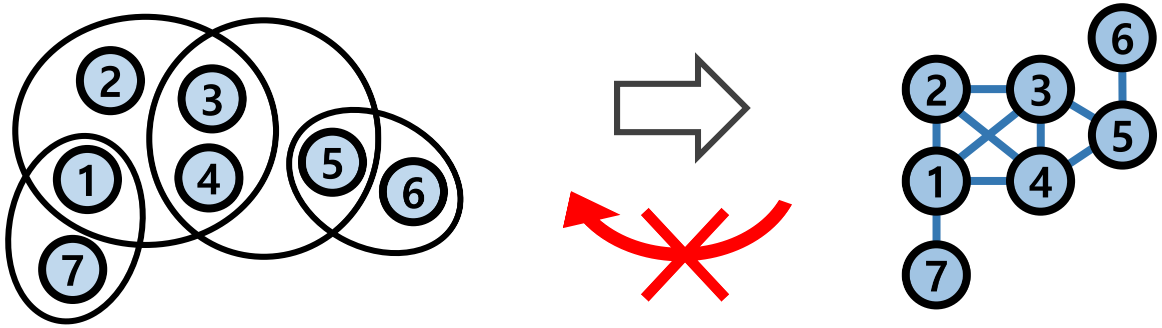


| | | | |
|---|---|---|---|
| | **Node** | P1 | |
| | **Hyperedge** | | |
| | **Hypergraph** | P2,P3,<br>P4,P5 | |
| | **Sub-hypergraph** | | |

M. T. Do, S. Yoon, B. Hooi, K. Shin, “Structural Patterns and Generative Models of Real-world Hypergraphs”, **KDD 2020**

# Multi-level Decomposition

- **Hypergraphs**: not straightforward to analyze.
  - Complex representation
  - Lack of tools
- Projection (a.k.a., **clique expansion**)
  - Information loss
  - No higher-order information

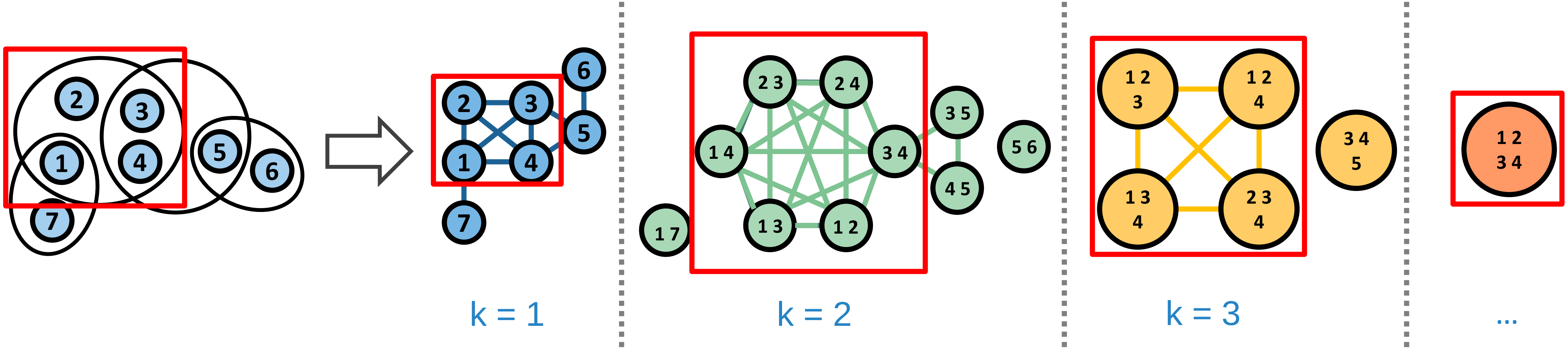


Only interactions at the level of nodes

M. T. Do, S. Yoon, B. Hooi, K. Shin, “Structural Patterns and Generative Models of Real-world Hypergraphs”, **KDD 2020**

# Multi-level Decomposition (cont.)

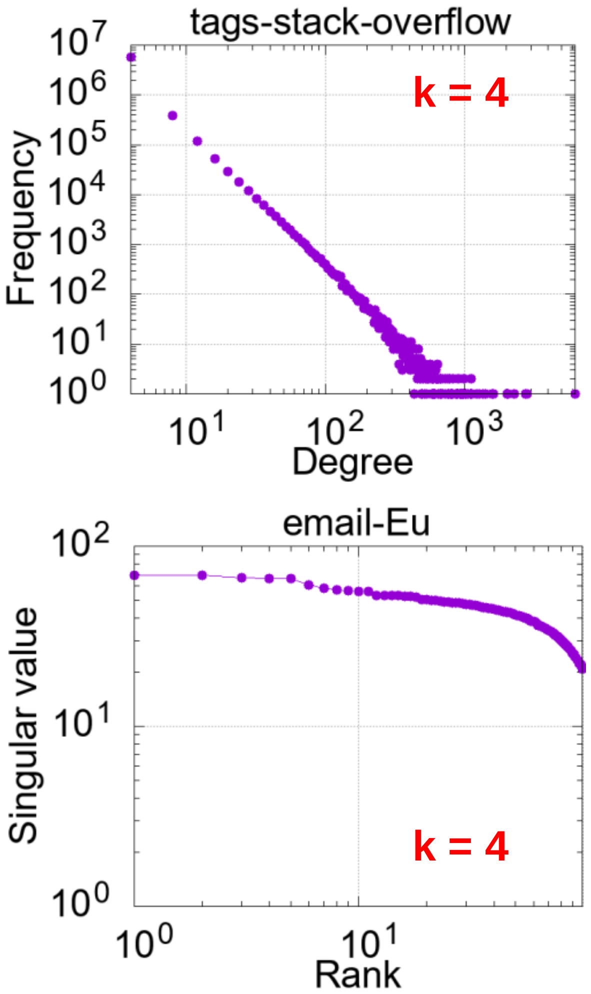
- **Multi-level decomposition**
  - Representation by pairwise unipartite graphs
  - Leveraging existing tools & measurements
  - No information loss: Original hypergraph is reconstructible



M. T. Do, S. Yoon, B. Hooi, K. Shin, “Structural Patterns and Generative Models of Real-world Hypergraphs”, **KDD 2020**

# Multi-level Decomposition (cont.)

- At every decomposition level,
  - degree distributions are **heavy-tailed**
  - singular value distributions are **heavy-tailed**
  - a large proportion of nodes are **connected**
    - *Most nodes are in a single component.*
  - clustering coefficient is **high**
    - *Real-world hypergraphs are clustered.*
  - diameter is **small**
    - *Most pairs of nodes are reachable in a few steps.*



M. T. Do, S. Yoon, B. Hooi, K. Shin, “Structural Patterns and Generative Models of Real-world Hypergraphs”, **KDD 2020**

# Roadmap

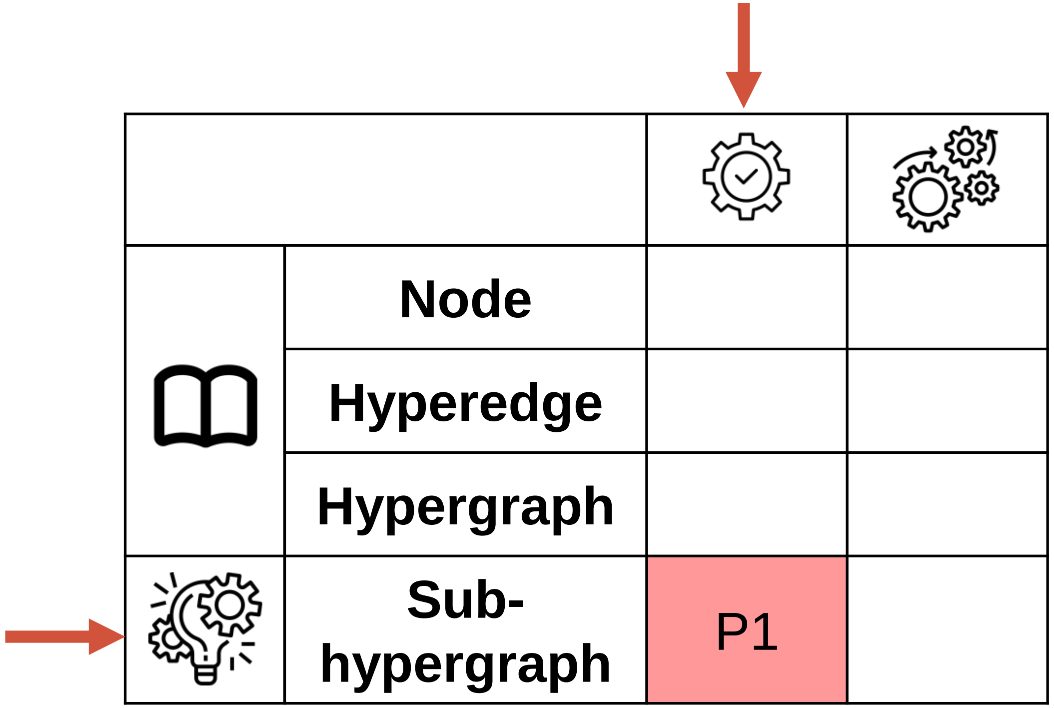
- **Part 1. Static Structural Patterns**
  - Basic Patterns
  - Advanced Patterns **<<**
- **Part 2. Dynamic Structural Patterns**
  - Basic Patterns
  - Advanced Patterns
- **Part 3. Generative Models**
  - Static hypergraph Generator
  - Dynamic hypergraph Generator

# Part 1-2. Advanced Static 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** | KBCYS23, BASJK18, LMMB22, LKS20, LCS21, LL23, BLS23 | BASJK18, CJ21, LS21 |

# BASJK18: One Advanced Static Pattern

- **P1.** Open and closed triangles

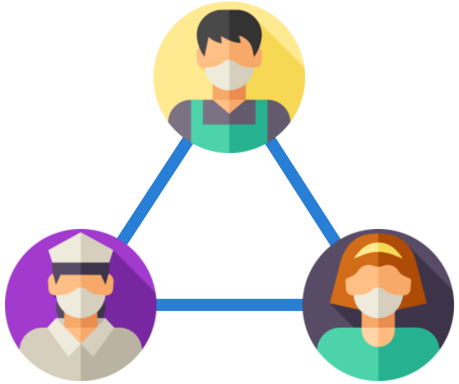


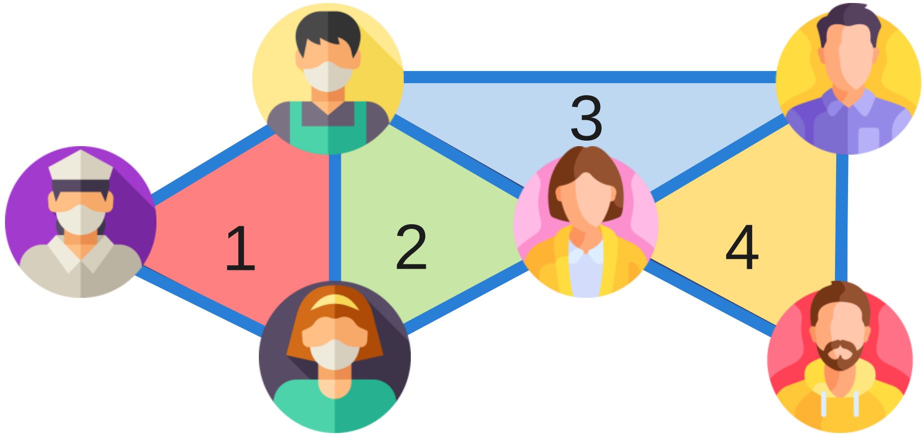
| | | | |
|---|---|---|---|
| | | | |
| | Node | | |
| | Hyperedge | | |
| | Hypergraph | | |
| | Sub-hypergraph | P1 | |

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

# Background

- A **triangle** is a clique (complete subgraph) of 3 nodes
- The **count** of triangles is an important primitive.
  - E.g., Community detection, spam detection, link prediction





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

# Triangles in Hypergraphs

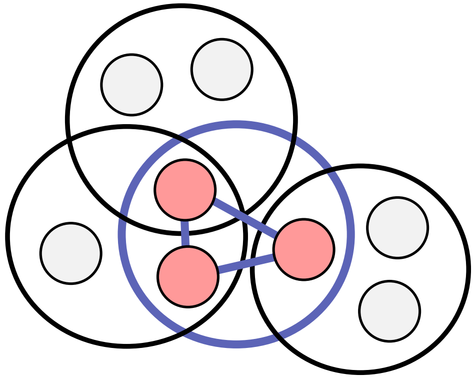
?

**Question:**

How can we define **triangles** in hypergraphs?

**Answer:**

Tri-wise relations (i.e., group interactions of three nodes) should be taken into account.

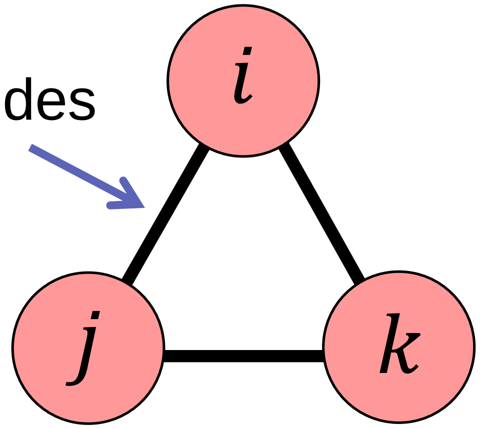


!

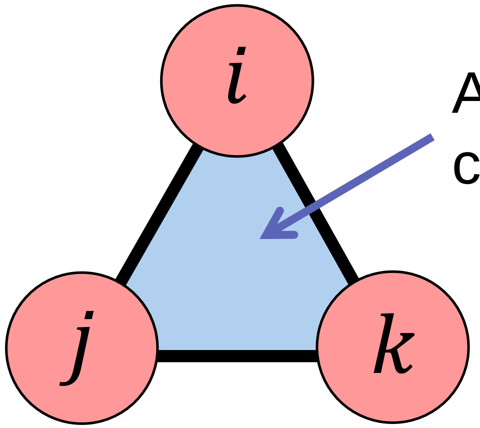
A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, “Simplicial Closure and Higher-order Link Prediction”, **PNAS (2018)**

# Open and Closed Triangles: Definition

- There are **two types of triangles** in hypergraphs.
  - **Closed triangles** cannot be captured by pairwise graphs.



**Open Triangle**

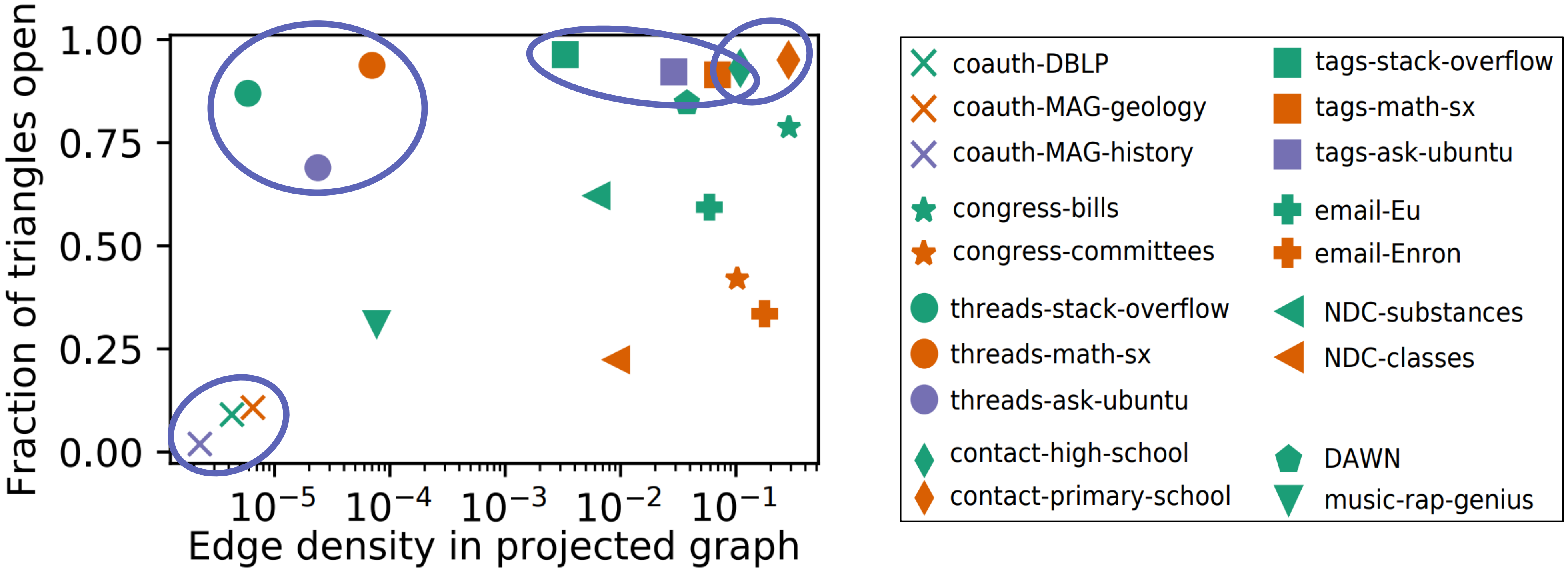


**Closed Triangle**

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

# Triangles across Domains

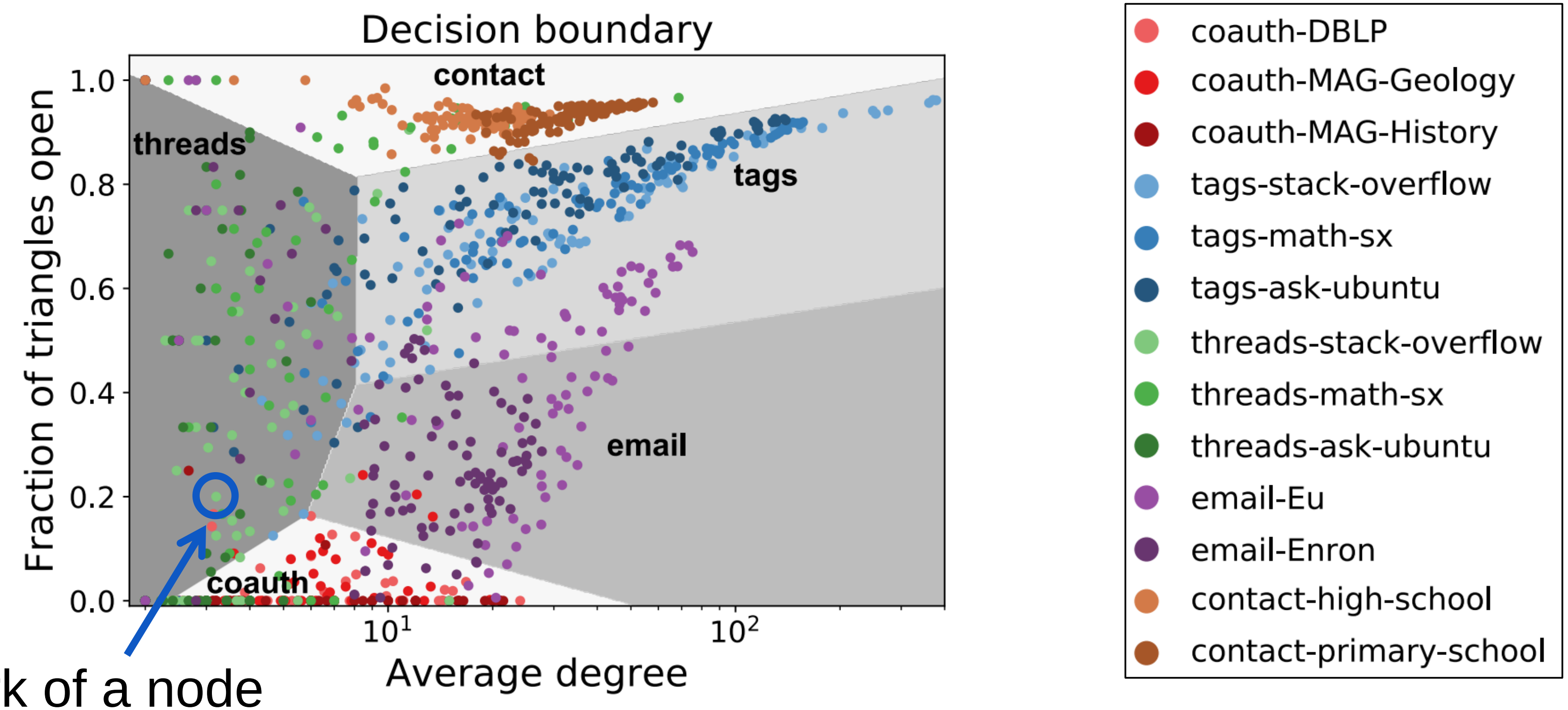
- **Fractions of open triangles** are similar within domains.



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

# Triangles across Domains (cont.)

- **Fractions of open triangles** are similar within domains.

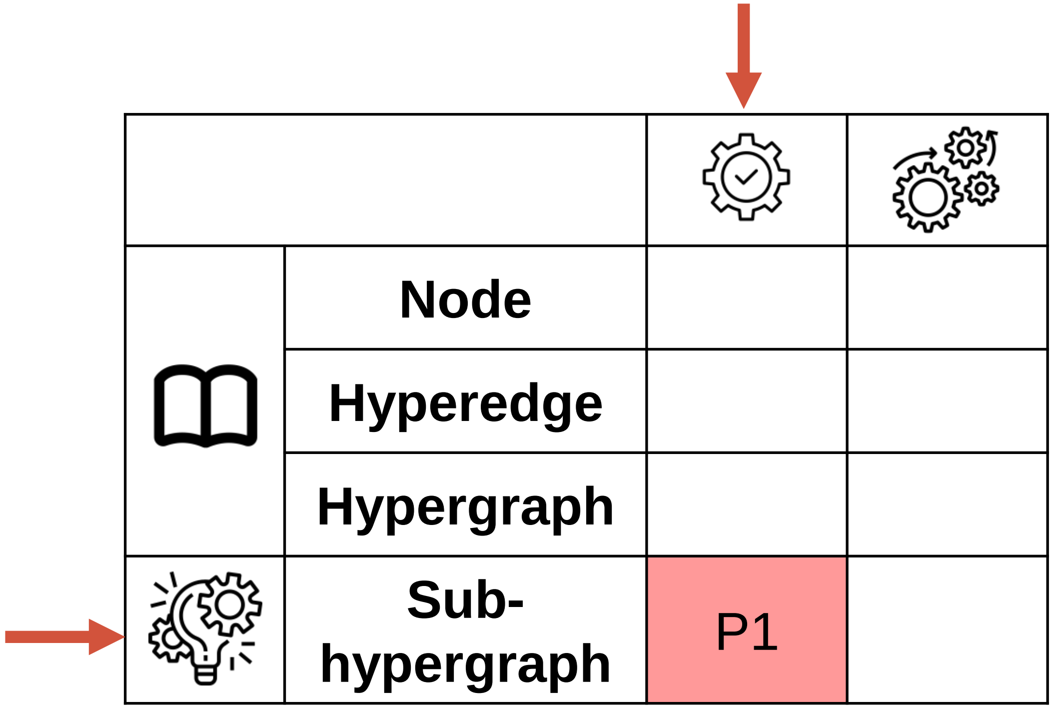


Ego-network of a node

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

# LMMB20: One Advanced Static Pattern

- **P1.** Higher-order network motifs

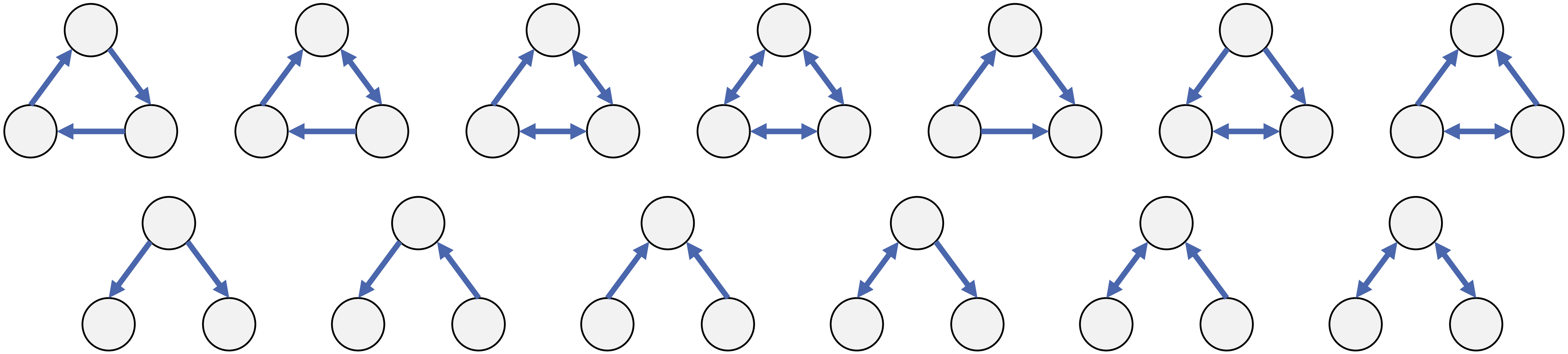


| | | | |
|---|---|---|---|
| | | | |
| | Node | | |
| | Hyperedge | | |
| | Hypergraph | | |
| | Sub-hypergraph | P1 | |

Q. F. Lotito, F. Musciotto, A. Montresor, F. Battiston, “Higher-order Motif Analysis in Hypergraphs”, **Communication Physics (2022)**

# Background

- **Network motifs** are fundamental building blocks of complex networks.
  - They appear in real-world graphs at a frequency that is **significantly higher** than randomized graphs.



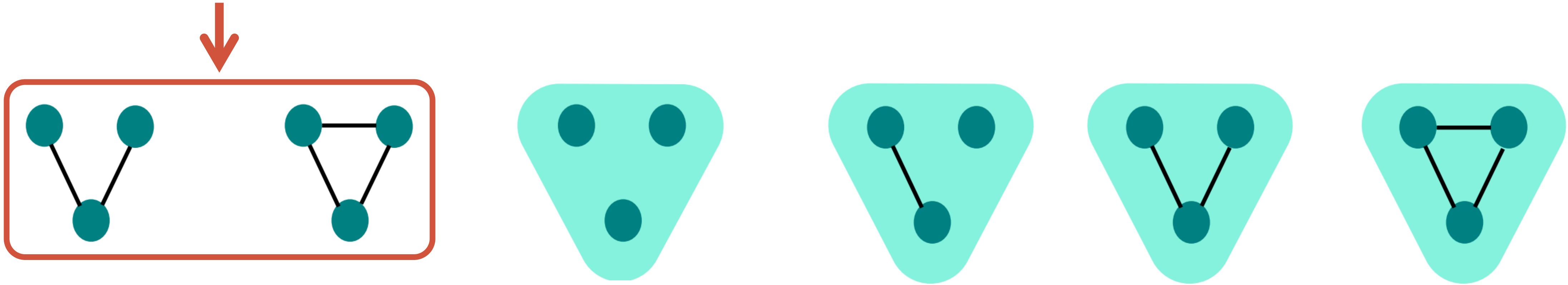
**13 different types of 3-node network motifs**

Q. F. Lotito, F. Musciotto, A. Montresor, F. Battiston, "Higher-order Motif Analysis in Hypergraphs", **Communication Physics (2022)**

# Higher-order Network Motifs: Definition

- **Higher-order network motifs** are a generalization of network motifs.
- They additionally consider **group interactions** between the nodes.

Network motifs can only describe 2 patterns.

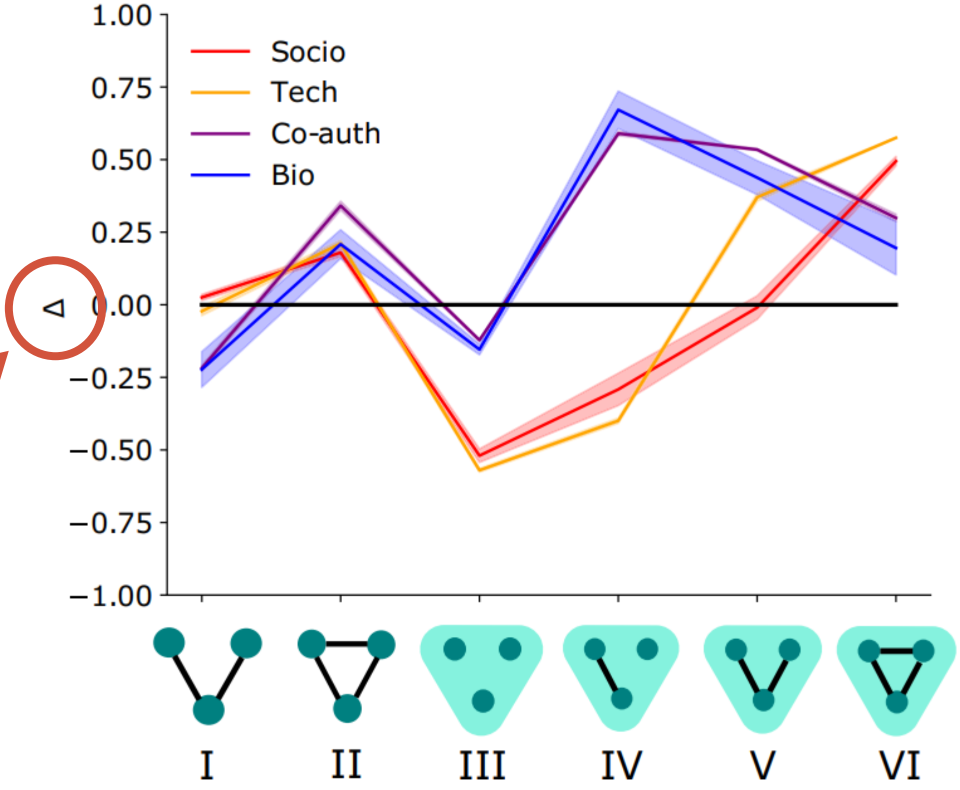


**6 different types of 3-node higher-order motifs**

Q. F. Lotito, F. Musciotto, A. Montresor, F. Battiston, “Higher-order Motif Analysis in Hypergraphs”, **Communication Physics (2022)**

# Comparison across Domains

- Different **higher-order motifs** are highlighted in each domain.



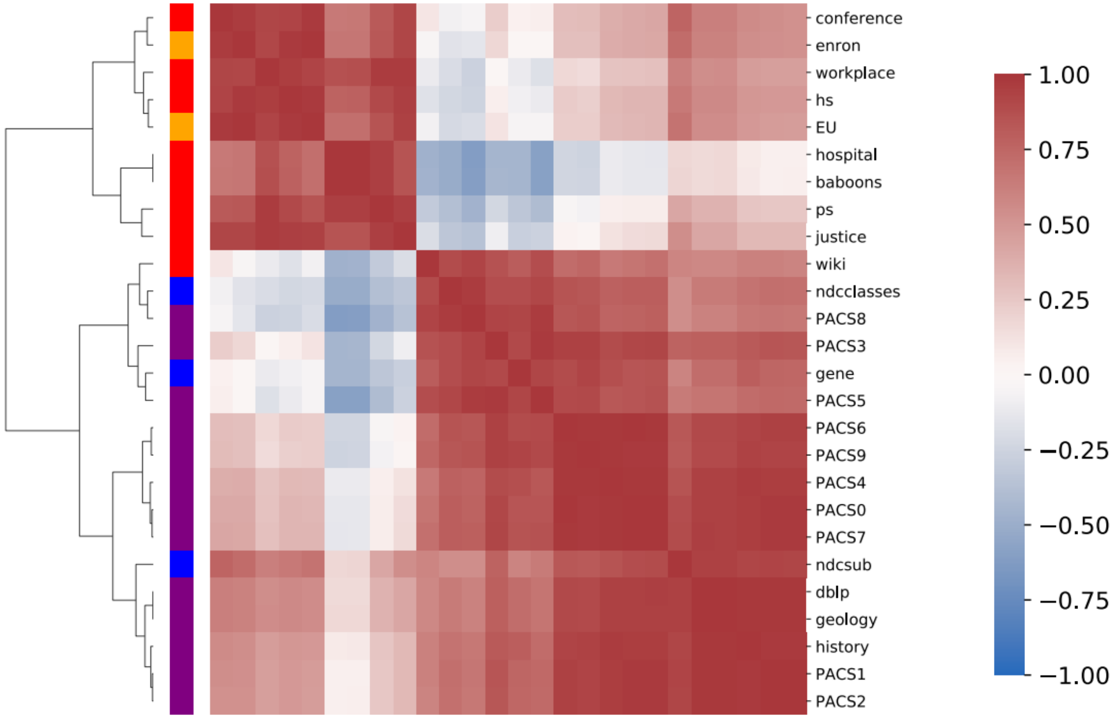
Abundance of higher-order motifs relative to random networks

Q. F. Lotito, F. Musciotto, A. Montresor, F. Battiston, “Higher-order Motif Analysis in Hypergraphs”, **Communication Physics (2022)**

# Comparison across Domains (cont.)

- Distributions of **higher-order motifs** are similar within domains and different across domains.

🟥 Social domain
🟧 E-mail domain
🟪 Co-authorship domain
🟦 Biological domain



Q. F. Lotito, F. Musciotto, A. Montresor, F. Battiston, “Higher-order Motif Analysis in Hypergraphs”, **Communication Physics (2022)**

# LKS20: One Advanced Static Pattern

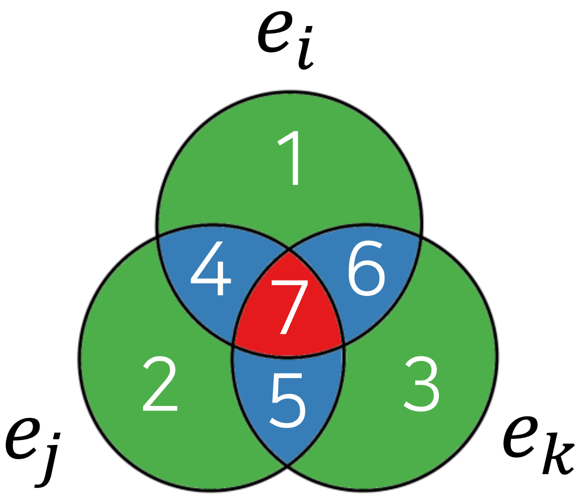
- **P1.** Hypergraph motifs (h-motifs)

| | | ⚙ | ⚙ |
|---|---|---|---|
| 📖 | **Node** | | |
| | **Hyperedge** | | |
| | **Hypergraph** | | |
| 💡 | **Sub-hypergraph** | P1 | |

G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

# 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) – (7).

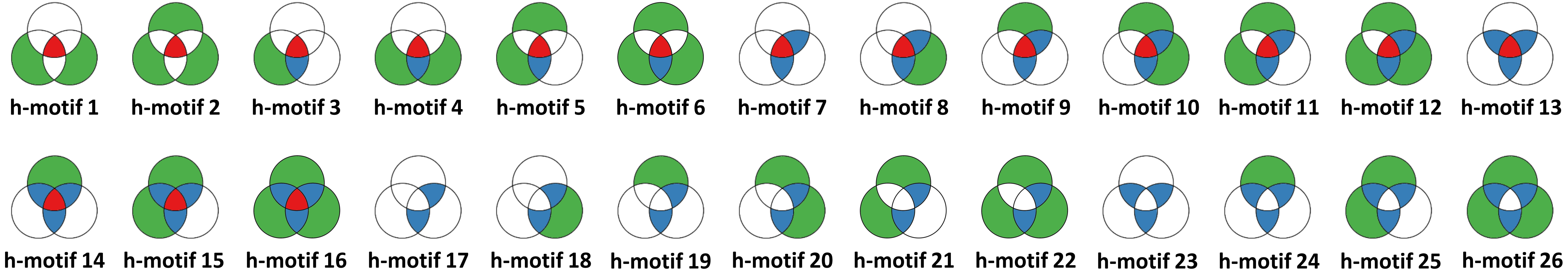


(1) $e_i \backslash e_j \backslash e_k$
(2) $e_j \backslash e_k \backslash e_i$
(3) $e_k \backslash e_i \backslash e_j$
(4) $e_i \cap e_j \backslash e_k$
(5) $e_j \cap e_k \backslash e_i$
(6) $e_k \cap e_i \backslash e_j$
(7) $e_i \cap e_j \cap e_k$

G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

# Hypergraph Motifs: Definition (cont.)

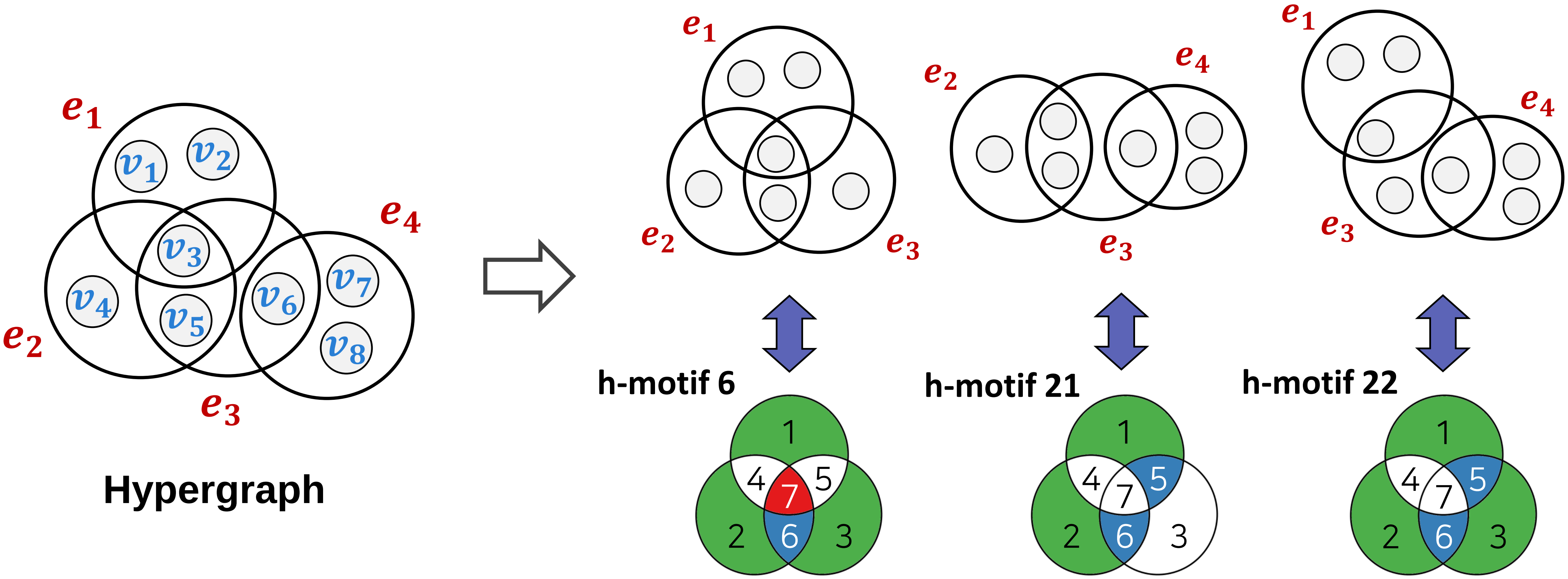
- While there can exist $2^7$ h-motifs, **26 h-motifs** remain once we exclude:
  1. symmetric ones
  2. those cannot be obtained from distinct hyperedges
  3. those cannot be obtained from connected hyperedges



G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

# Hypergraph Motifs: Example

- Example: A hypergraph with 8 nodes and 4 hyperedges



G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

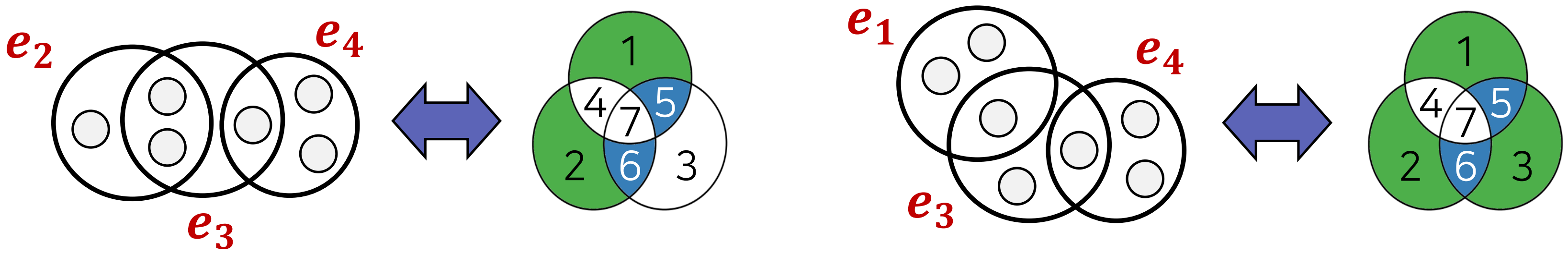
# Hypergraph Motifs: Properties (cont.)

**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.

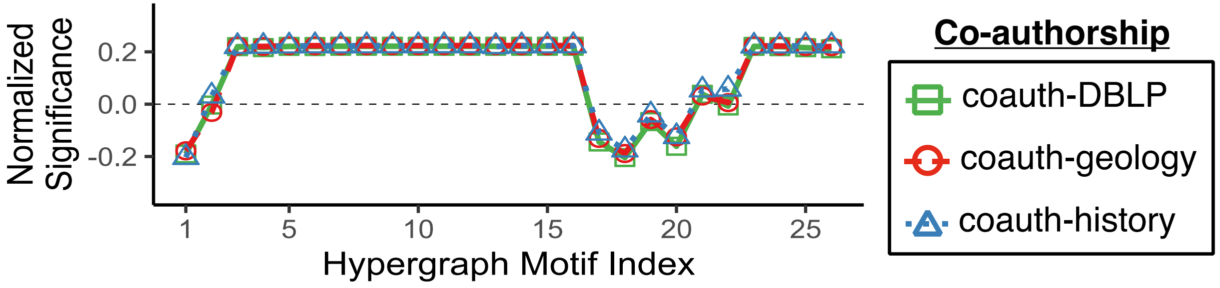


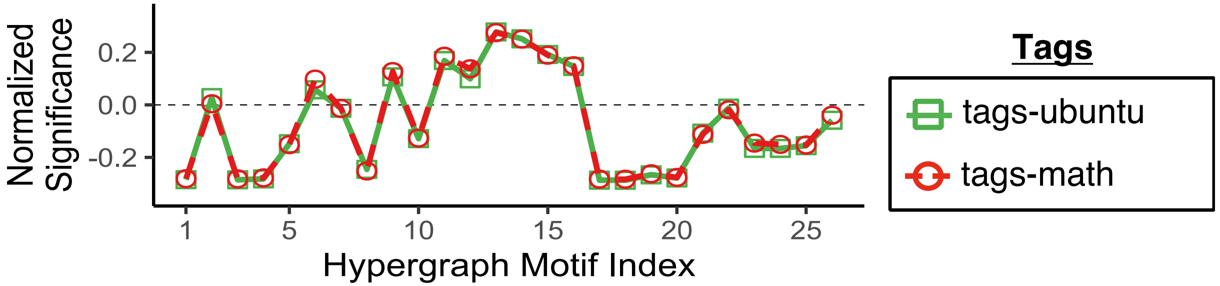
For example, $\{e_2, e_3, e_4\}$ and $\{e_1, e_3, e_4\}$ have same pairwise relations, while their connectivity patterns are distinguished by h-motifs.

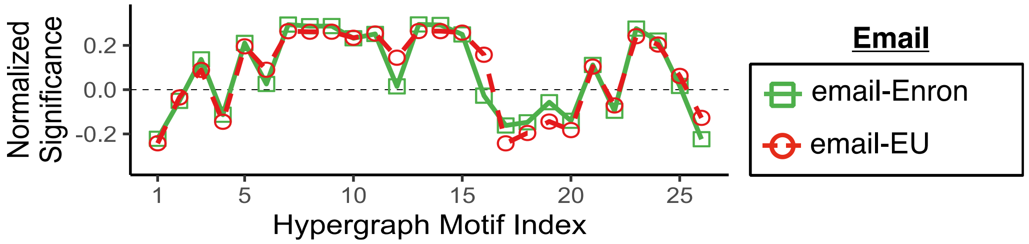
G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

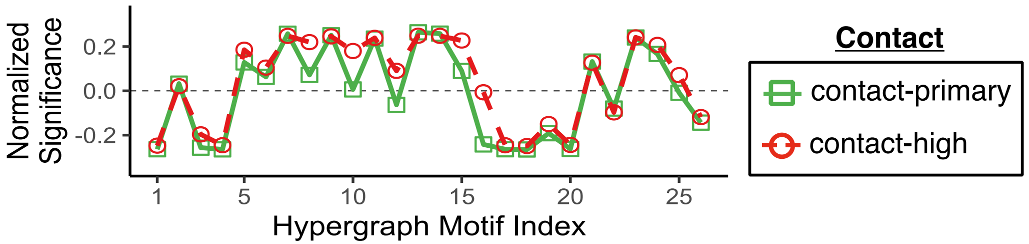
# Comparison across Domains

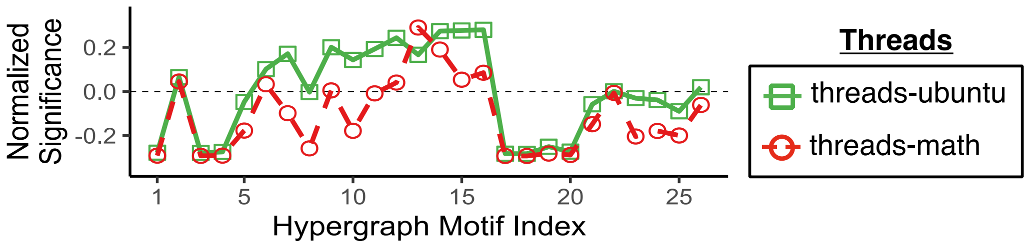
- CPs are **similar within domains** but **different across domains**.







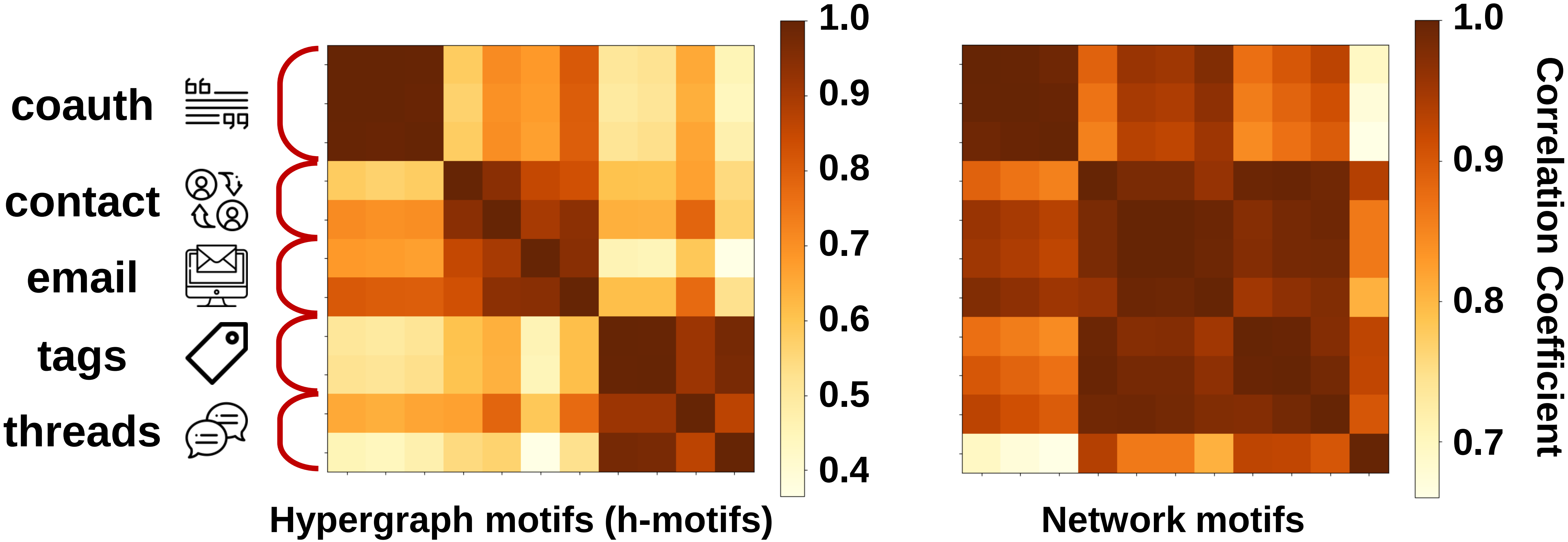




G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

# Comparison across Domains (cont.)

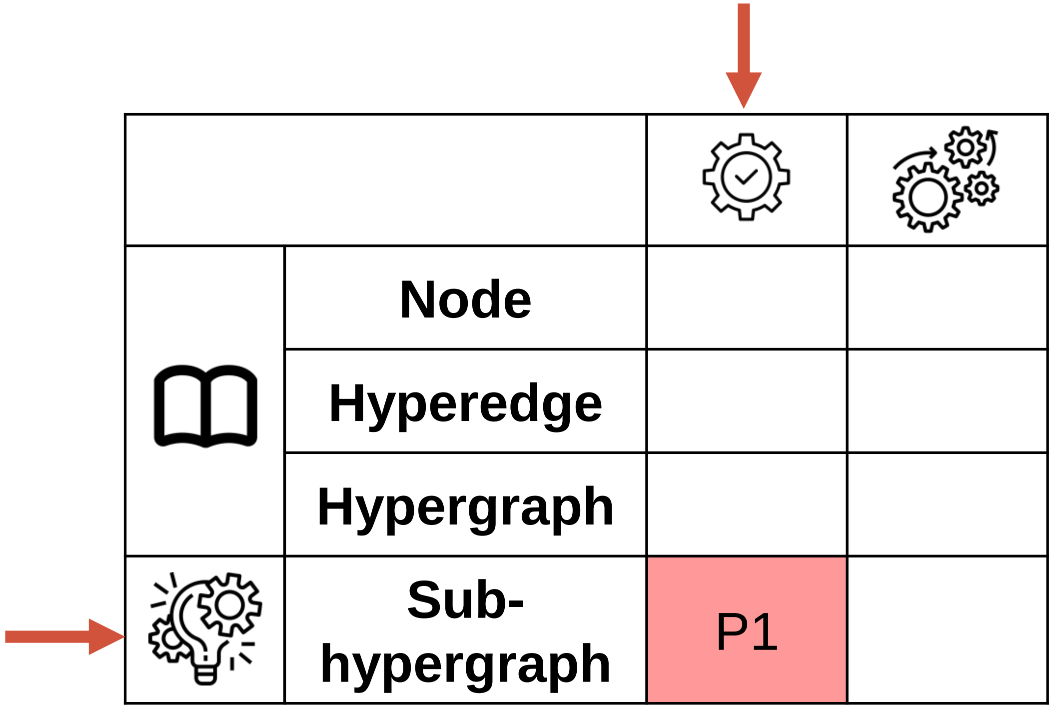
- **CPs based on h-motifs** capture local structural patterns more accurately than **CPs based on network motifs**.



G. Lee, J. Ko, K. Shin, “Hypergraph Motifs: Concepts, Algorithms, and Discoveries”, **VLDB 2020**

# LCS21: One Advanced Static Pattern

- **P1.** Density & overlapness of ego-network



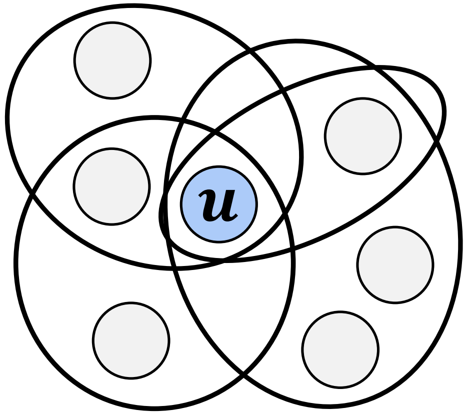
| | | | |
|---|---|---|---|
| | | | |
| | **Node** | | |
| | **Hyperedge** | | |
| | **Hypergraph** | | |
| | **Sub-hypergraph** | P1 | |

G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Hypergraph Ego-network

- An **ego-network** $\mathcal{E}$ of node $u$ is the set of hyperedges that contains $u$.

$$\mathcal{E}(u) \coloneqq \{e \in E : u \in e\}$$



G. Lee, M. Choe, K. Shin, "How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators", **WWW 2021**

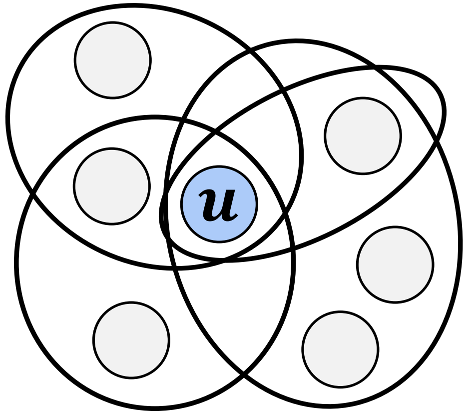
# Density of Ego-networks

- **Density** measures how densely hyperedges overlap.

$$\rho(\mathcal{E}) := \frac{|\mathcal{E}|}{\left|\bigcup_{e\in\mathcal{E}} e\right|}$$

← # of hyperedges

← # of nodes



**Hypergraph**

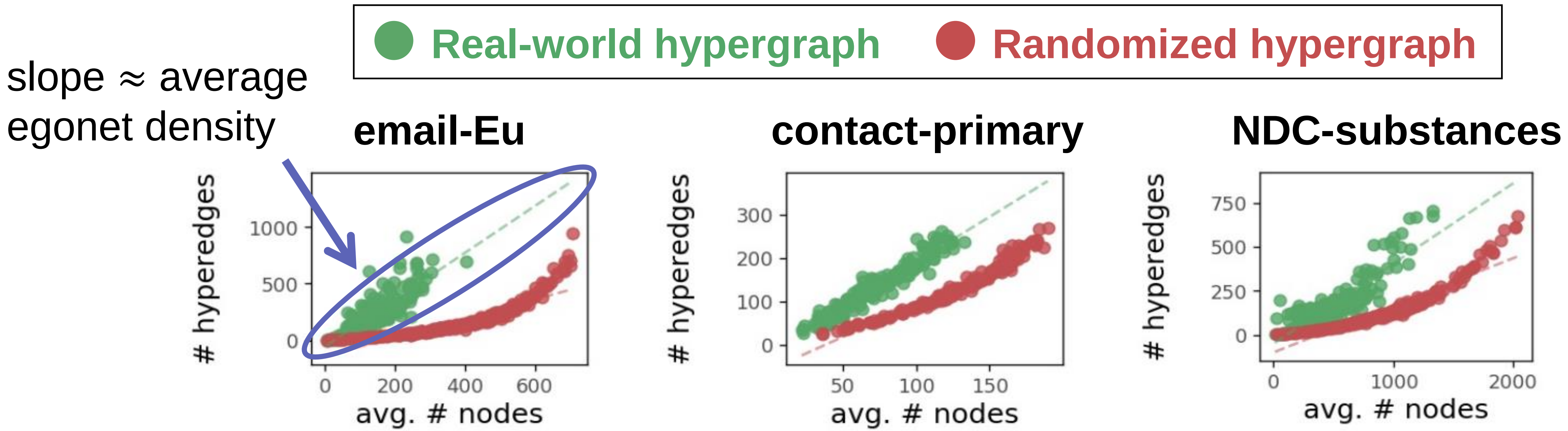
**Density** of egonet $\mathcal{E}(u)$ is $\frac{4}{7}$.

**Example**

G. Lee, M. Choe, K. Shin, "How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators", **WWW 2021**

# Density of Ego-networks (cont.)

- Ego-networks in real-world hypergraphs tend to have **higher density** than those in randomized ones.

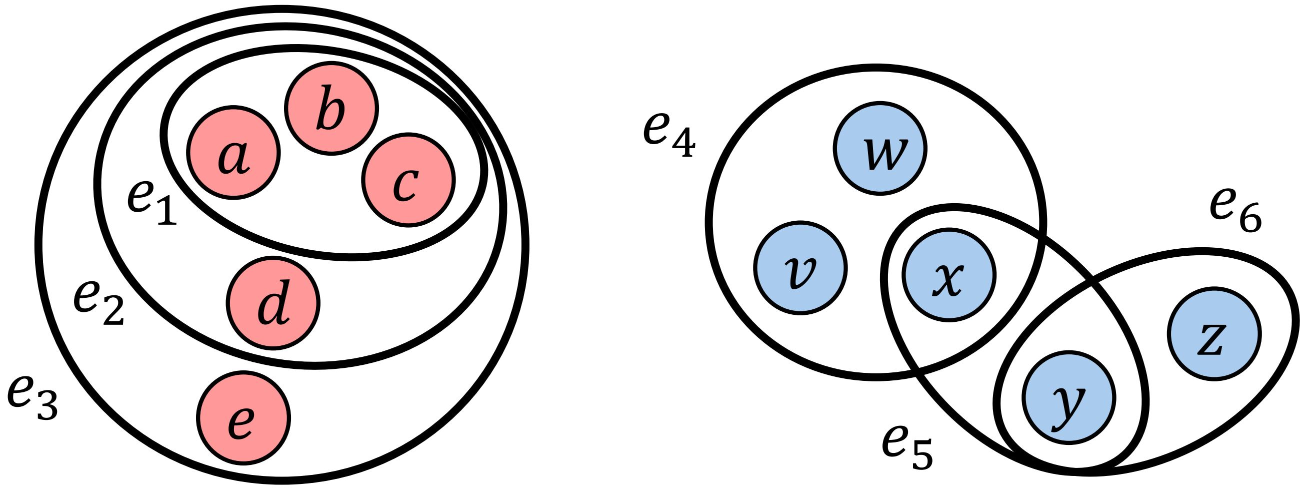


slope ≈ average egonet density

G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Density of Ego-networks (cont.)

- Does **density** fully capture the degree of overlaps of a set of hyperedges?



$\mathcal{E}_1 = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$

$\mathcal{E}_2 = \{\boldsymbol{e}_4, \boldsymbol{e}_5, \boldsymbol{e}_6\}$

**Our intuition**

$\mathcal{E}_1$ is more overlapped than $\mathcal{E}_2$.

**Density**

$$\rho(\mathcal{E}_1) = \rho(\mathcal{E}_2) = \frac{3}{5}$$

G. Lee, M. Choe, K. Shin, "How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators", **WWW 2021**

# Degree of Hyperedge Overlaps

?

**Question:**

What is the principled measure for the degree of overlaps of a set of hyperedges?

**Answer:**

- We present **three axioms** that any reasonable measure of the hyperedge overlaps should satisfy.
- Then, we propose **overlapness**, a new measure that satisfies all the axioms.

!

G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

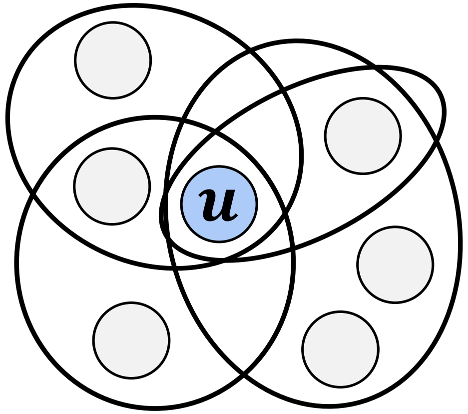
# Overlapness of Ego-networks

- **Overlapness** measures how densely hyperedges overlap.

$$o(\mathcal{E}) \coloneqq \frac{\sum_{e\in\mathcal{E}} |e|}{\left|\bigcup_{e\in\mathcal{E}} e\right|}$$

← sum of the hyperedge sizes

← # of nodes



**Hypergraph**

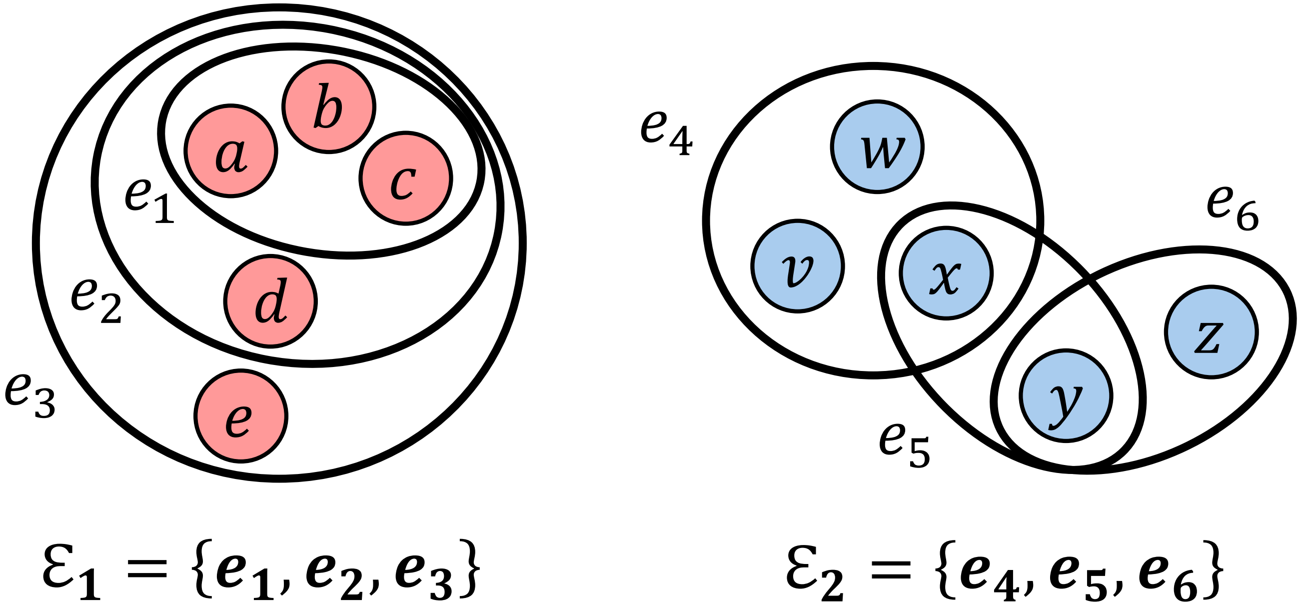
**Overlapness** of egonet $\mathcal{E}(u)$ is $\frac{12}{7}$.

**Example**

G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Overlapness of Ego-networks (cont.)

- Does **overlapness** capture the degree of overlaps of a set of hyperedges?



$\mathcal{E}_1 = \{e_1, e_2, e_3\}$

$\mathcal{E}_2 = \{e_4, e_5, e_6\}$

**Our intuition**

$\mathcal{E}_1$ is more overlapped than $\mathcal{E}_2$.

**Density**

$$\rho(\mathcal{E}_1) = \rho(\mathcal{E}_2) = \frac{3}{5}$$

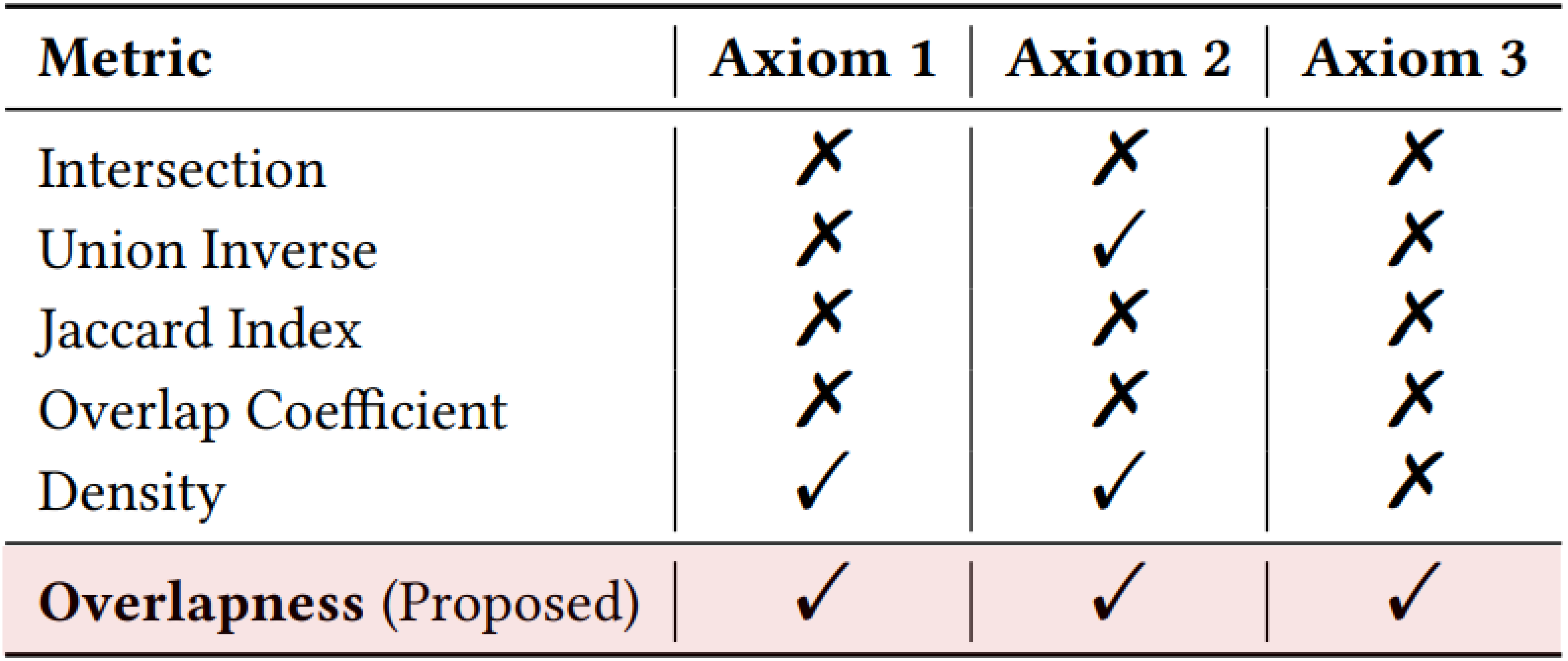
**Overlapness**

$$o(\mathcal{E}_1) = \frac{12}{5} \quad > \quad o(\mathcal{E}_2) = \frac{7}{5}$$

G. Lee, M. Choe, K. Shin, "How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators", **WWW 2021**

# Overlapness of Ego-networks (cont.)

- **Overlapness** satisfies all the axioms while others do not.



| **Metric** | **Axiom 1** | **Axiom 2** | **Axiom 3** |
|---|---|---|---|
| Intersection | ✗ | ✗ | ✗ |
| Union Inverse | ✗ | ✓ | ✗ |
| Jaccard Index | ✗ | ✗ | ✗ |
| Overlap Coefficient | ✗ | ✗ | ✗ |
| Density | ✓ | ✓ | ✗ |
| **Overlapness** (Proposed) | ✓ | ✓ | ✓ |

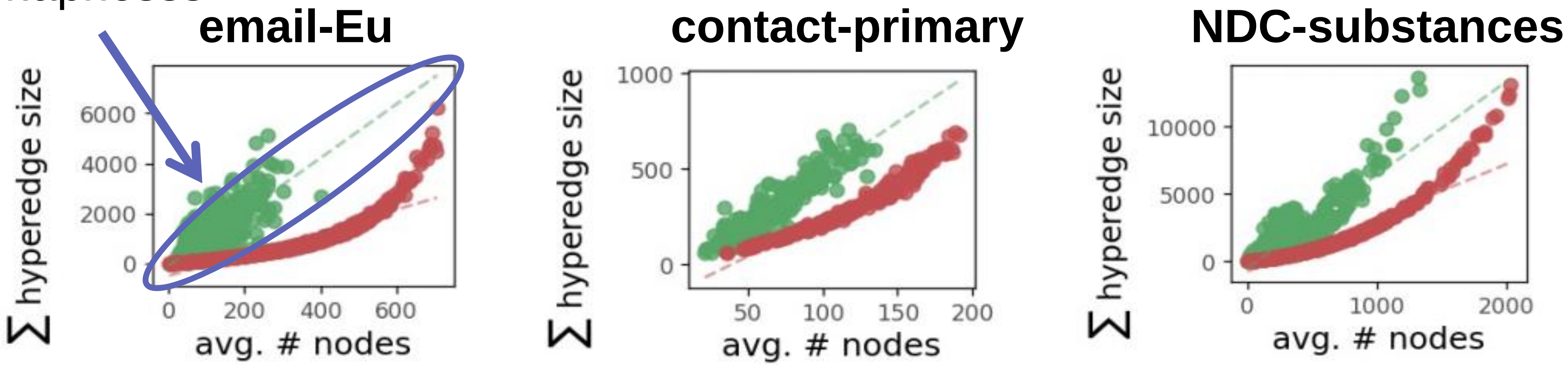
G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# Overlapness of Ego-networks (cont.)

- Ego-networks in real-world hypergraphs tend to have **higher overlapness** than those in randomized ones.

slope ≈ average egonet overlapnesss

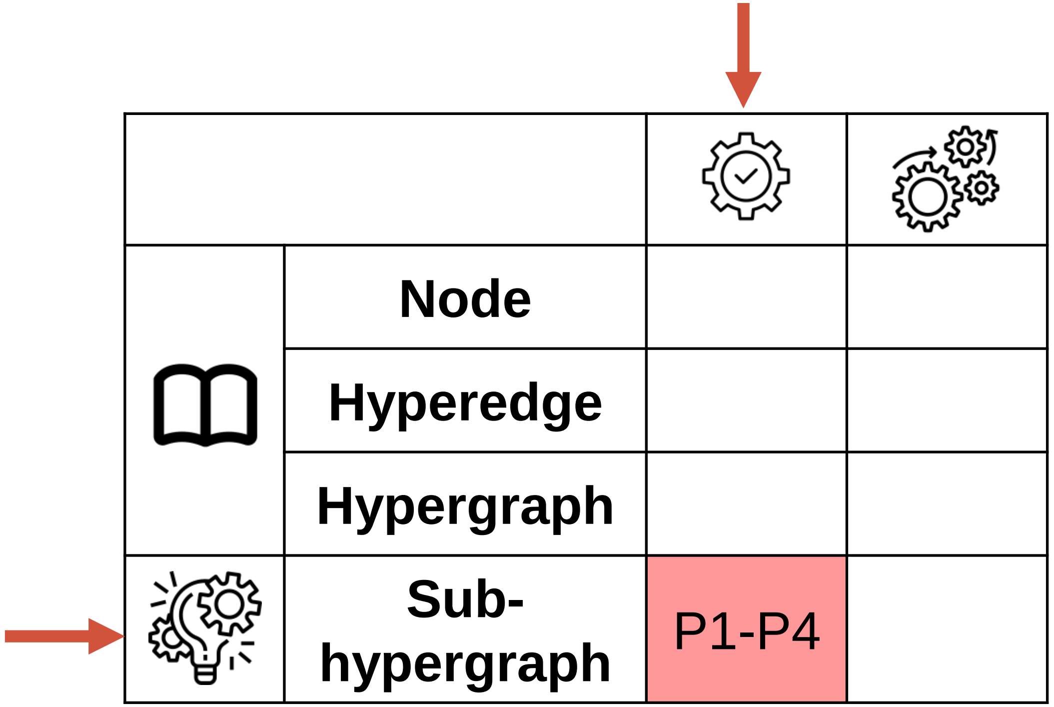
● **Real-world hypergraph** ● **Randomized hypergraph**



G. Lee, M. Choe, K. Shin, “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators”, **WWW 2021**

# KBCYS23: Five Advanced Static Patterns

- **P1.** Transitivity of hypergraphs
- **P2.** Transitivity of hyperwedges
- **P3.** Transitivity of nodes
- **P4.** Transitivity of hyperedges

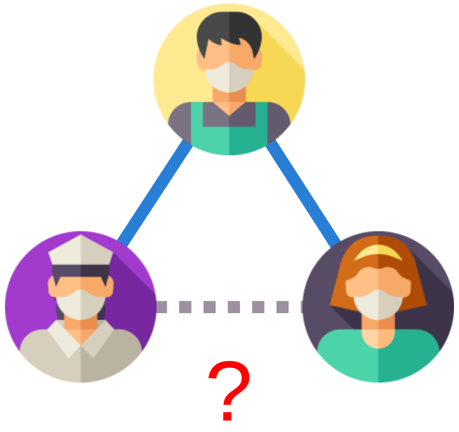


| | | | |
|---|---|---|---|
| | **Node** | | |
| | **Hyperedge** | | |
| | **Hypergraph** | | |
| | **Sub-hypergraph** | P1-P4 | |

S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Background

- **Transitivity** (a.k.a., **clustering coefficient**) measures the likelihood of two neighbors of a node in a graph being adjacent.
- It has been used in diverse fields, e.g., neuroscience and finance.



Transitivity of node **u** =

$$\frac{\text{\# of pairs of neighbors of } \mathbf{u} \text{ that are connected}}{\text{\# of pairs of neighbors of } \mathbf{u}}$$

S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Background

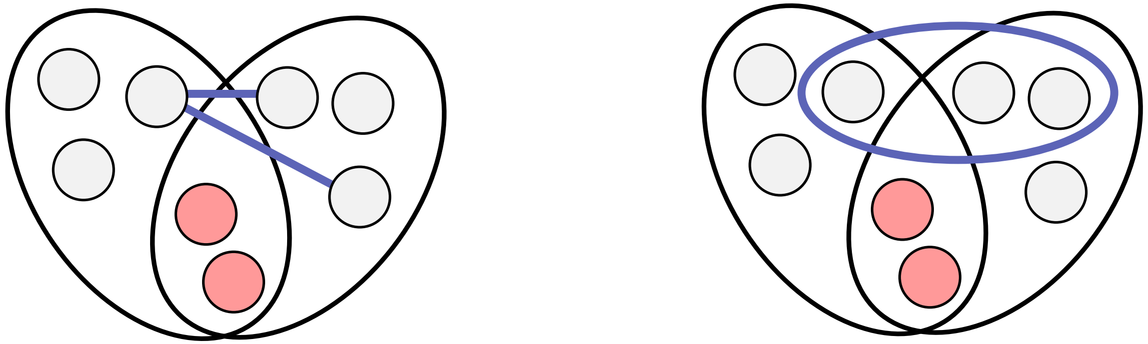
?

**Question:**

How can we define **transitivity** of group interactions?

**Answer:**

Higher-order interactions between the two groups of neighbors should be taken into account.

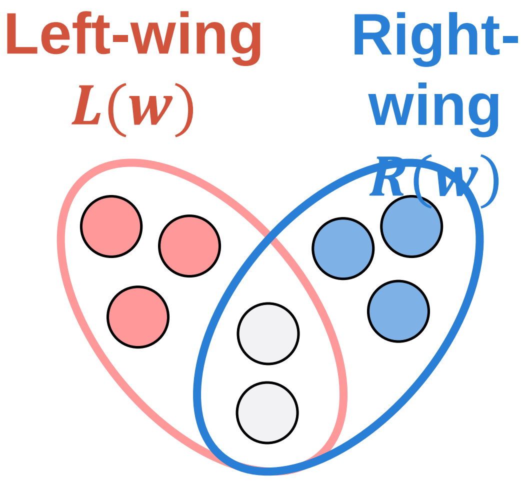


!

S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Hypergraph Transitivity: Definition

- **HyperTrans** is a principled hypergraph transitivity measure.
  - It quantifies the group interactions between left and right wings.



**Hyperwedge** $w$

Hyperedge score

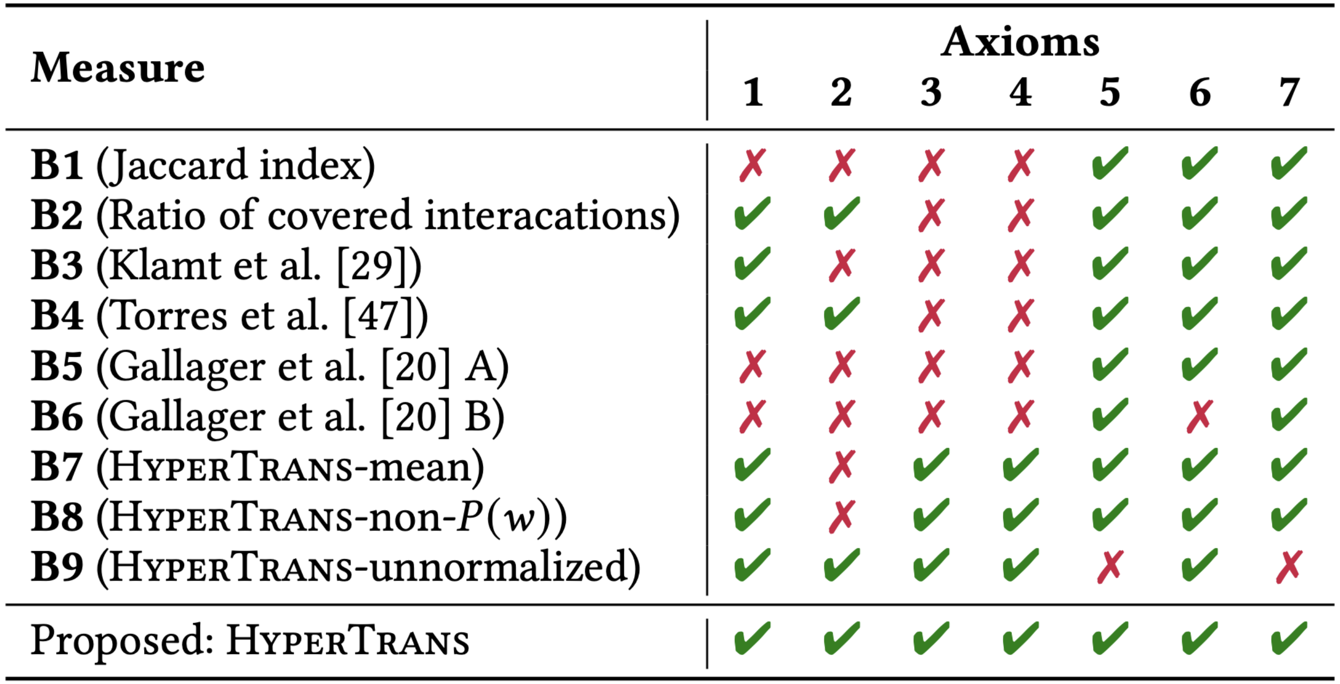
$$\mathcal{T}(w) = \sum_{v\in L(w)} \sum_{v'\in R(w)} \frac{\max_{e\in E} \boxed{f(w,e)} \cdot \boxed{\mathbb{1}[v,v' \in e]}}{|L(w)| \times |R(w)|}$$

Consider hyperedges that include each pair of nodes from $L(w)$ and $R(w)$

S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Hypergraph Transitivity: Definition (cont.)

- **HyperTrans** satisfies all axioms that a proper hypergraph transitivity measure should satisfy.

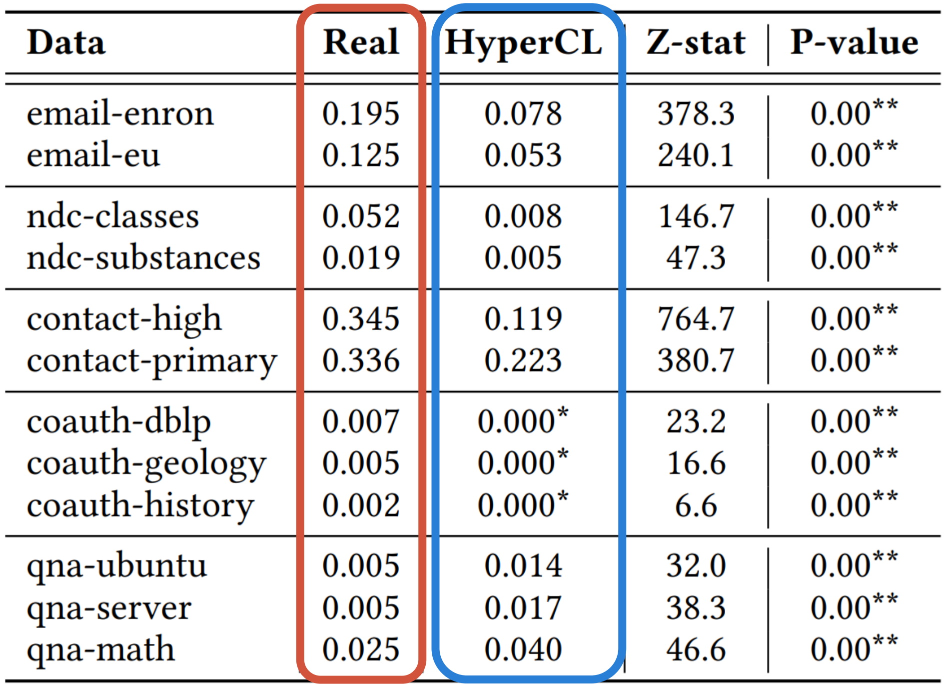


| **Measure** | **Axioms** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **B1** (Jaccard index) | ✗ | ✗ | ✗ | ✗ | ✔ | ✔ | ✔ |
| **B2** (Ratio of covered interacations) | ✔ | ✔ | ✗ | ✗ | ✔ | ✔ | ✔ |
| **B3** (Klamt et al. [29]) | ✔ | ✗ | ✗ | ✗ | ✔ | ✔ | ✔ |
| **B4** (Torres et al. [47]) | ✔ | ✔ | ✗ | ✗ | ✔ | ✔ | ✔ |
| **B5** (Gallager et al. [20] A) | ✗ | ✗ | ✗ | ✗ | ✔ | ✔ | ✔ |
| **B6** (Gallager et al. [20] B) | ✗ | ✗ | ✗ | ✗ | ✔ | ✗ | ✔ |
| **B7** (HyperTrans-mean) | ✔ | ✗ | ✔ | ✔ | ✔ | ✔ | ✔ |
| **B8** (HyperTrans-non-$P(w)$) | ✔ | ✗ | ✔ | ✔ | ✔ | ✔ | ✔ |
| **B9** (HyperTrans-unnormalized) | ✔ | ✔ | ✔ | ✔ | ✗ | ✔ | ✗ |
| Proposed: HyperTrans | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |

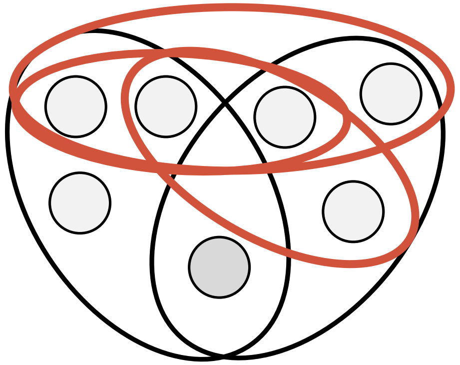
S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Transitivity of Hypergraphs

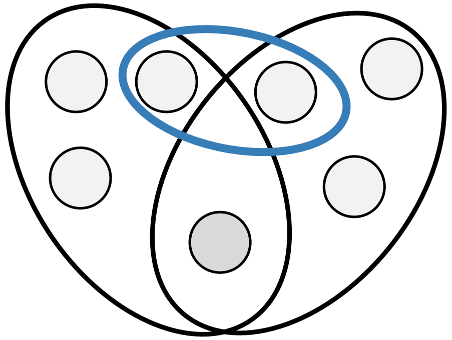
- Real-world hypergraphs are **more transitive** than randomized ones.
- Transitivity of **a hypergraph** is the average transitivity of hyperwedges.



| **Data** | **Real** | **HyperCL** | **Z-stat** | **P-value** |
|---|---|---|---|---|
| email-enron | 0.195 | 0.078 | 378.3 | 0.00** |
| email-eu | 0.125 | 0.053 | 240.1 | 0.00** |
| ndc-classes | 0.052 | 0.008 | 146.7 | 0.00** |
| ndc-substances | 0.019 | 0.005 | 47.3 | 0.00** |
| contact-high | 0.345 | 0.119 | 764.7 | 0.00** |
| contact-primary | 0.336 | 0.223 | 380.7 | 0.00** |
| coauth-dblp | 0.007 | 0.000* | 23.2 | 0.00** |
| coauth-geology | 0.005 | 0.000* | 16.6 | 0.00** |
| coauth-history | 0.002 | 0.000* | 6.6 | 0.00** |
| qna-ubuntu | 0.005 | 0.014 | 32.0 | 0.00** |
| qna-server | 0.005 | 0.017 | 38.3 | 0.00** |
| qna-math | 0.025 | 0.040 | 46.6 | 0.00** |



Real-world Hypergraph



Randomized Hypergraph

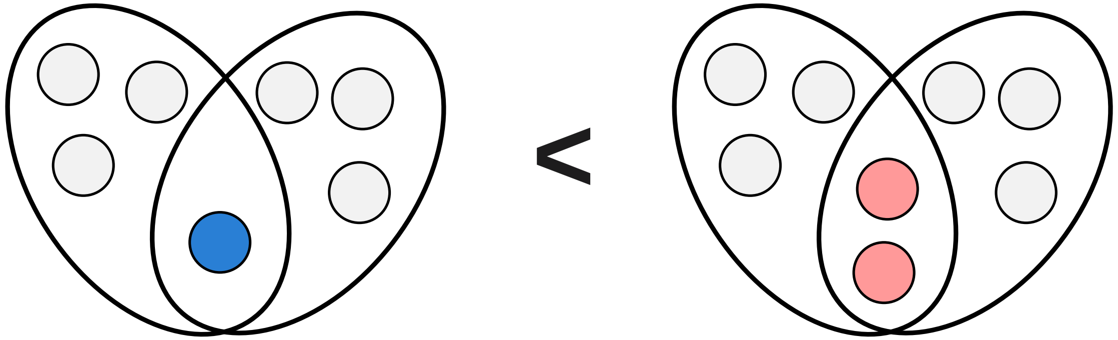
S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Transitivity of Hyperwedges

- The **larger** the body-group size of hyperwedges, the more likely they are to exhibit **high** transitivity.

| Data | Real | HyperCL |
|---|---|---|
| email-enron | 0.09 | -0.09 |
| email-eu | 0.12 | -0.14 |
| ndc-classes | 0.32 | -0.10 |
| ndc-substances | 0.14 | -0.10 |
| contact-high | 0.13 | 0.00* |
| contact-primary | 0.13 | 0.00* |
| coauth-dblp | 0.12 | 0.00* |
| coauth-geology | 0.14 | 0.00* |
| coauth-history | 0.12 | 0.05 |
| qna-ubuntu | 0.04 | 0.00* |
| qna-server | 0.04 | 0.00* |
| qna-math | 0.04 | 0.01 |

Correlations between **body-group size** and **transitivity**

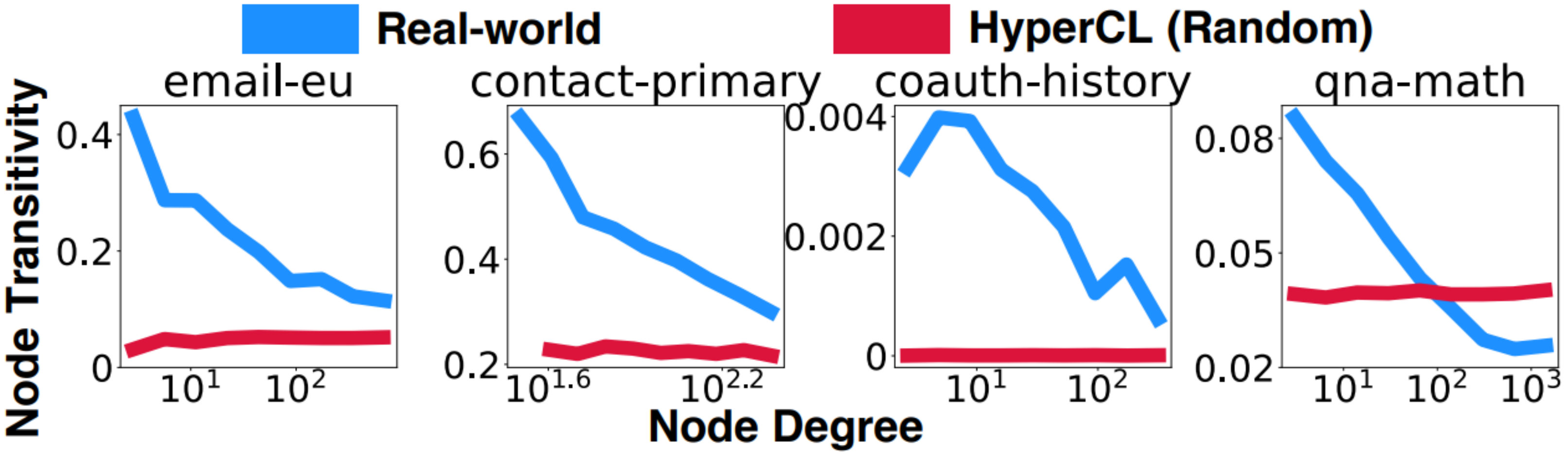


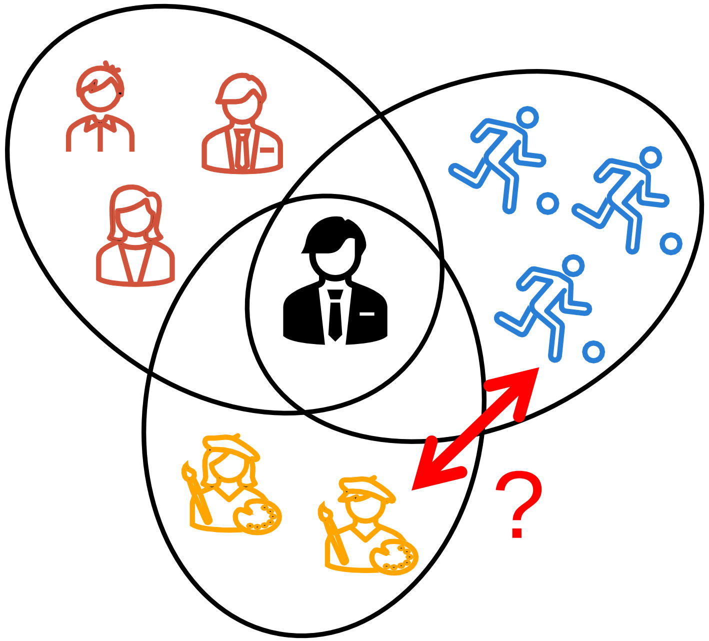
**More Transitive**

S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Transitivity of Nodes

- **High-degree** nodes are likely to have **low** transitivity.



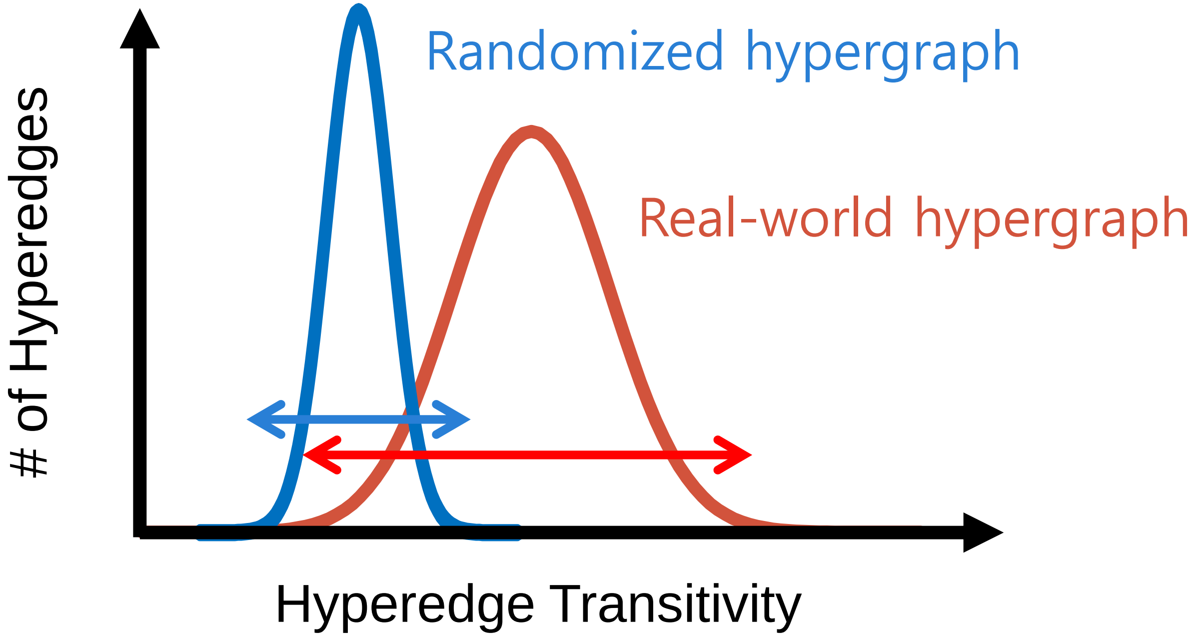


S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, “How Transitive Are Real-World Group Interactions? Measurement and Reproduction”, **KDD 2023**

# Transitivity of Hyperedges

- Hyperedges in real-world hypergraphs have **wider ranges of transitivity** compared to those in randomized hypergraphs.

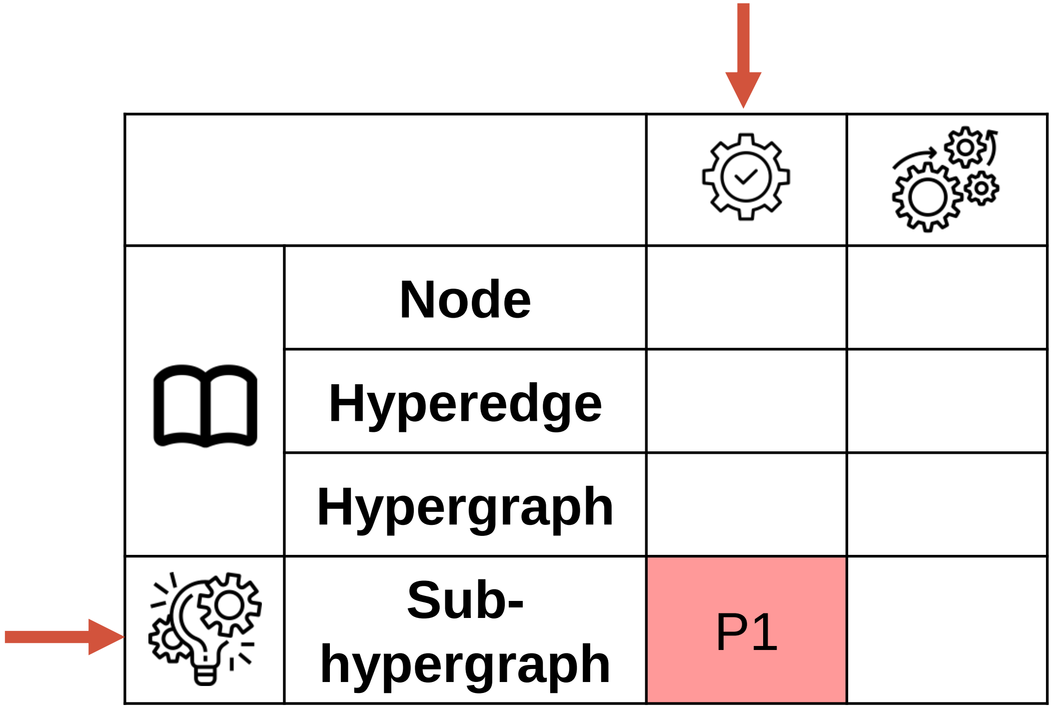
| Data | Real | HyperCL |
|---|---|---|
| email-enron | 0.725 | 0.279 |
| email-eu | 0.809 | 0.248 |
| ndc-classes | 0.600 | 0.075 |
| ndc-substances | 1.0 | 0.032 |
| contact-high | 0.794 | 0.316 |
| contact-primary | 0.693 | 0.395 |
| coauth-dblp | 1.0 | 0.105 |
| coauth-geology | 1.0 | 0.069 |
| coauth-history | 1.0 | 0.333 |
| qna-ubuntu | 0.667 | 0.5 |
| qna-server | 0.667 | 0.333 |
| qna-math | 0.667 | 1.00 |



S. Kim, F. Bu, M. Choe, J. Yoo, and K. Shin, "How Transitive Are Real-World Group Interactions? Measurement and Reproduction", **KDD 2023**

# LL23: One Advanced Static Pattern

- **P1.** Degree of hyperedge encapsulation

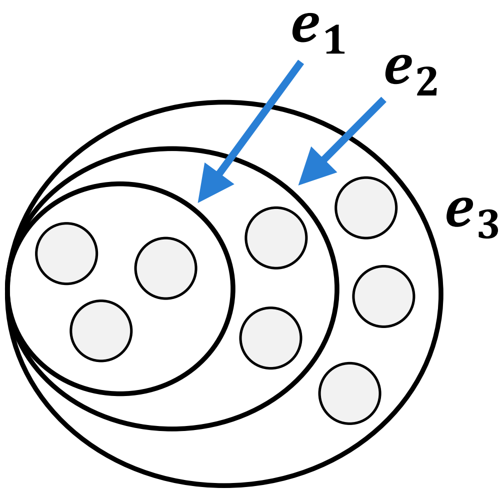


| | | ⚙ | ⚙ |
|---|---|---|---|
| 📖 | **Node** | | |
| | **Hyperedge** | | |
| | **Hypergraph** | | |
| 💡 | **Sub-hypergraph** | P1 | |

T. LaRock and R. Lambiotte, "Encapsulation Structure and Dynamics in Hypergraphs", **arXiv (2023)**

# Encapsulation of Hyperedges

- Hyperedges can contain smaller hyperedges.
- However, they do not contain all possible sub-hyperedges.

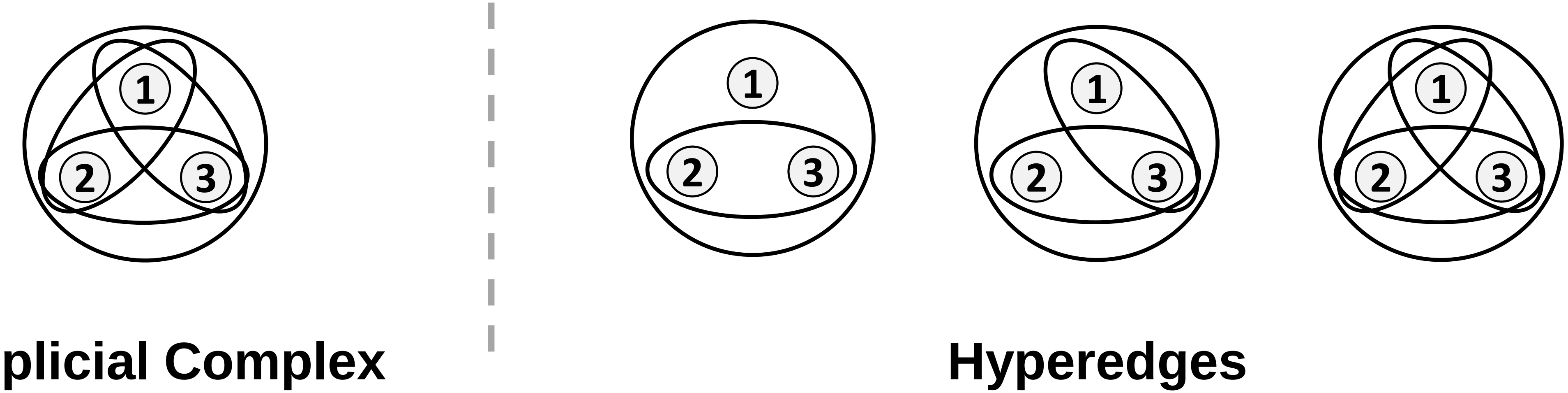


$$e_1 \subset e_2 \subset e_3$$

T. LaRock and R. Lambiotte, "Encapsulation Structure and Dynamics in Hypergraphs", **arXiv (2023)**

# Difference from Simplicial Complexes

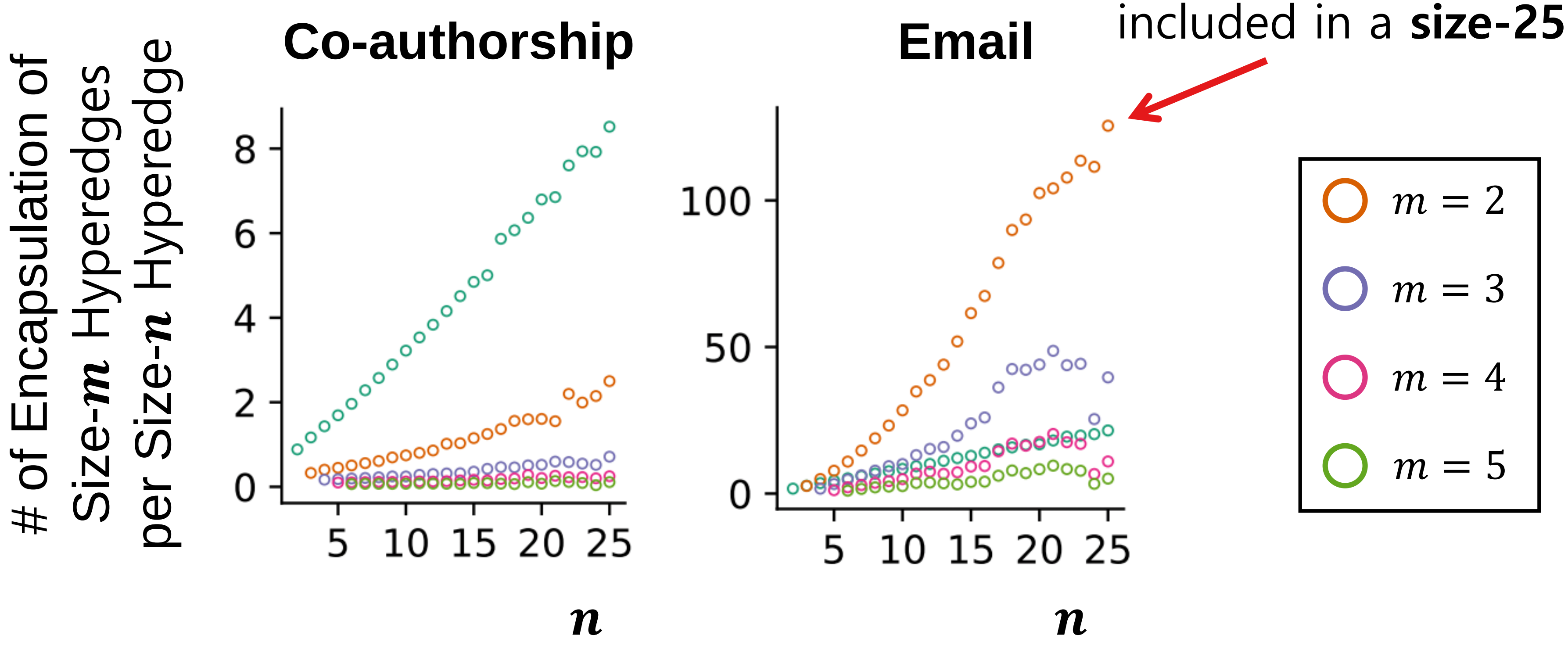
- A **simplicial complex** includes all subsets of the complex.
- A **hyperedge** can include various subsets in different ways.



T. LaRock and R. Lambiotte, "Encapsulation Structure and Dynamics in Hypergraphs", **arXiv (2023)**

# Degree of Encapsulation

- Larger hyperedges encapsulate smaller hyperedges.

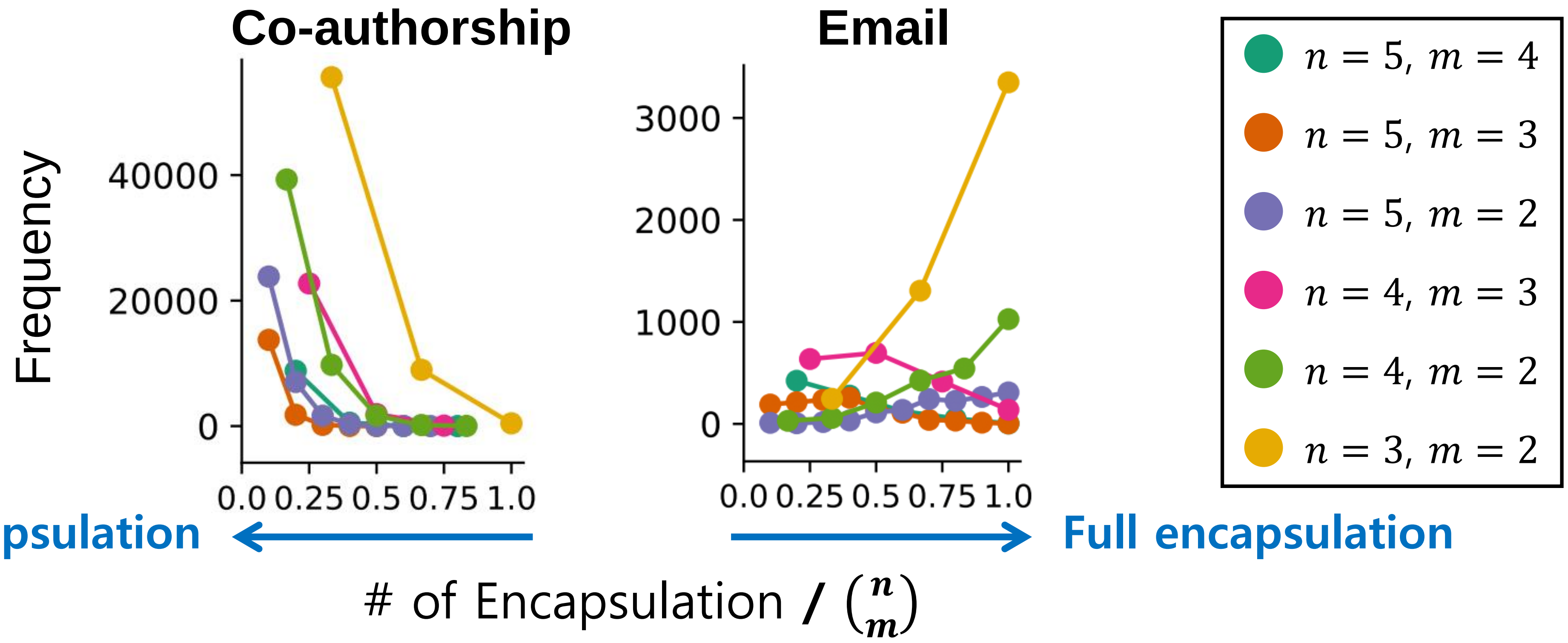


100+ **size-2** hyperedges are included in a **size-25** hyperedge.

T. LaRock and R. Lambiotte, “Encapsulation Structure and Dynamics in Hypergraphs”, **arXiv (2023)**

# Degree of Encapsulation (cont.)

- Hyperedges exhibit different encapsulation patterns across domains.



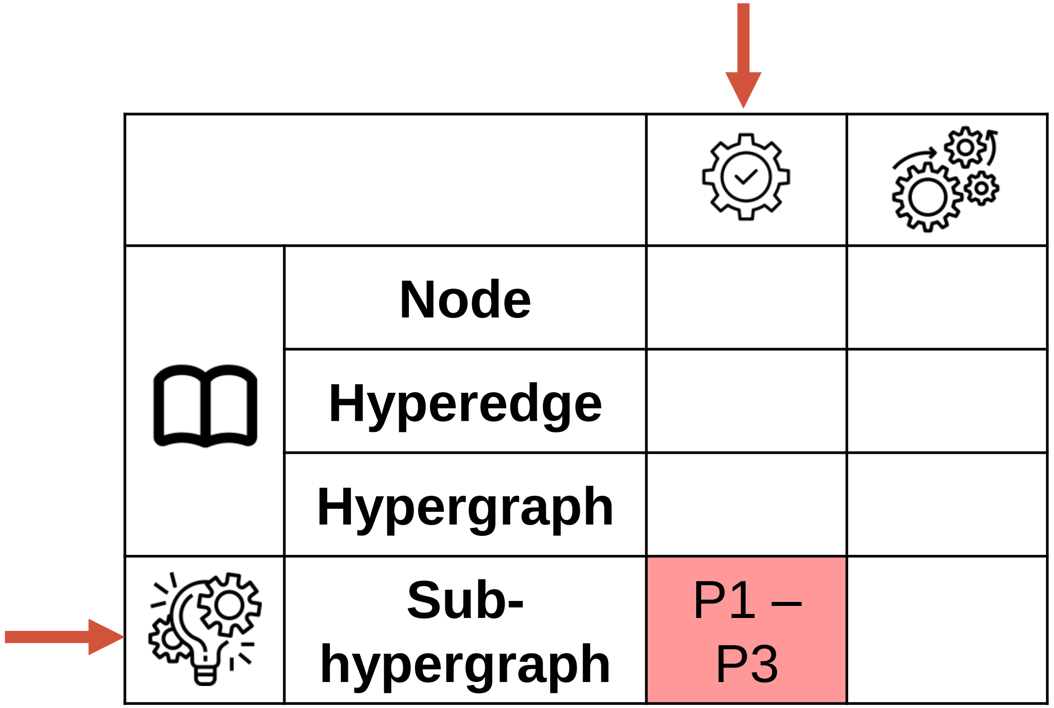
**Sparse encapsulation** ←

→ **Full encapsulation**

T. LaRock and R. Lambiotte, "Encapsulation Structure and Dynamics in Hypergraphs", **arXiv (2023)**

# BLS23: Three Advanced Static Patterns

- **P1.** Hypercore sizes of hypergraphs
- **P2.** Distributions of hypercoreness
- **P3.** Heterogeneity of hypercoreness

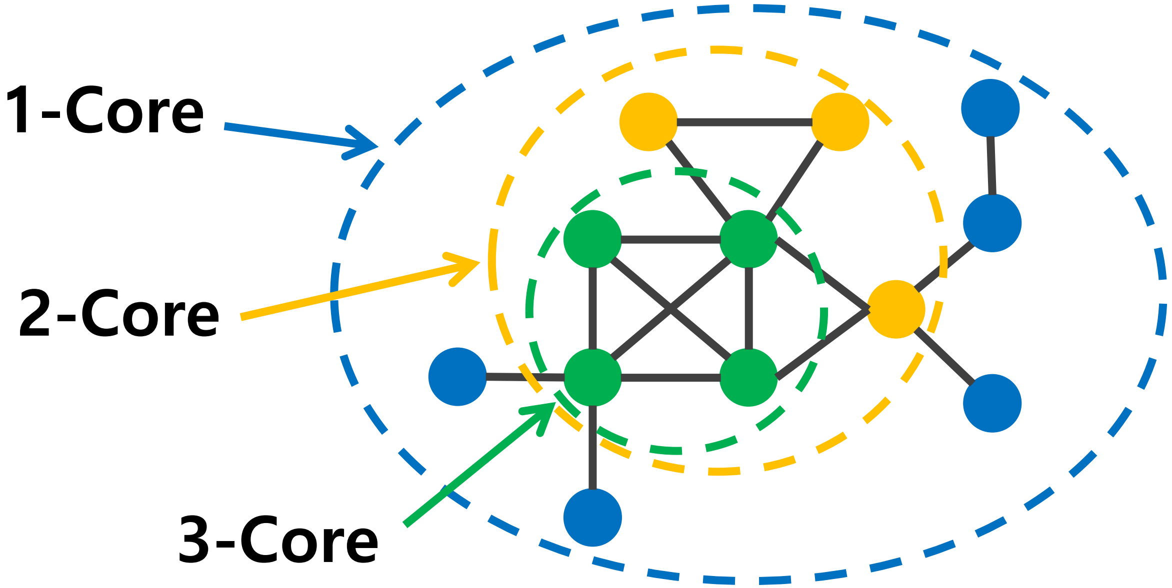


| | | ⚙ | ⚙⚙ |
|---|---|---|---|
| 📖 | **Node** | | |
| | **Hyperedge** | | |
| | **Hypergraph** | | |
| ⚙ | **Sub-hypergraph** | P1 – P3 | |

F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Background

- A ***k*-core** is a maximal subgraph of nodes with degree at least ***k***.
- It is useful in community detection, anomaly detection, etc.



🔵 Nodes with **coreness 1**

🟡 Nodes with **coreness 2**

🟢 Nodes with **coreness 3**

F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Hypercores

?

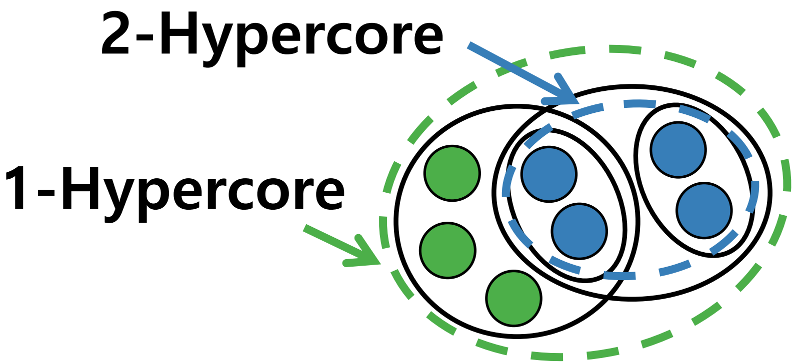
**Question:**

How can we define ***k*-cores** in hypergraphs?

**Answer:**

We can easily generalize them to hypergraphs:

!



🟢 Nodes with **hypercoreness 1**
🔵 Nodes with **hypercoreness 2**

F. Bu, G. Lee, K. Shin, "Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications", **Data Mining and Knowledge Discovery (2023)**

# Hypercores (cont.)

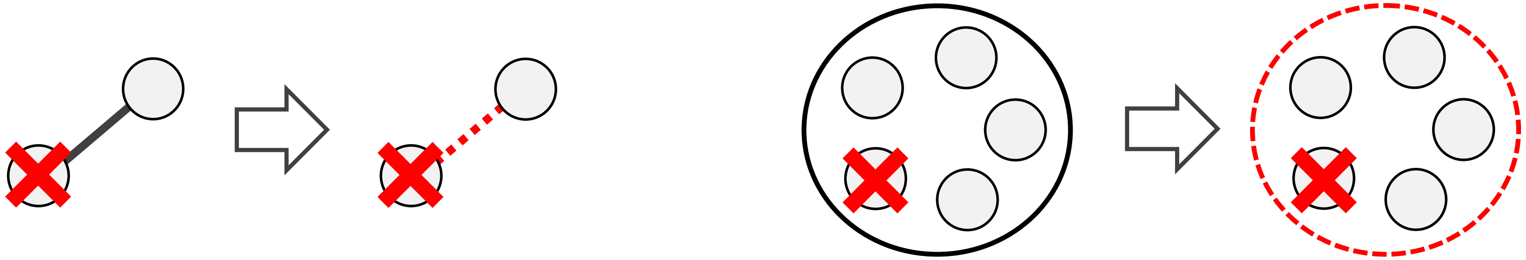
?

**Question:**

How can we define ***k*-cores** in hypergraphs?

**Answer (cont.):**

Consider the **fragility** of hyperedges.



Removing a single node breaks all of its edges.

Does a single node break the entire group interaction?

!

F. Bu, G. Lee, K. Shin, "Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications", **Data Mining and Knowledge Discovery (2023)**

# Hypercores (cont.)

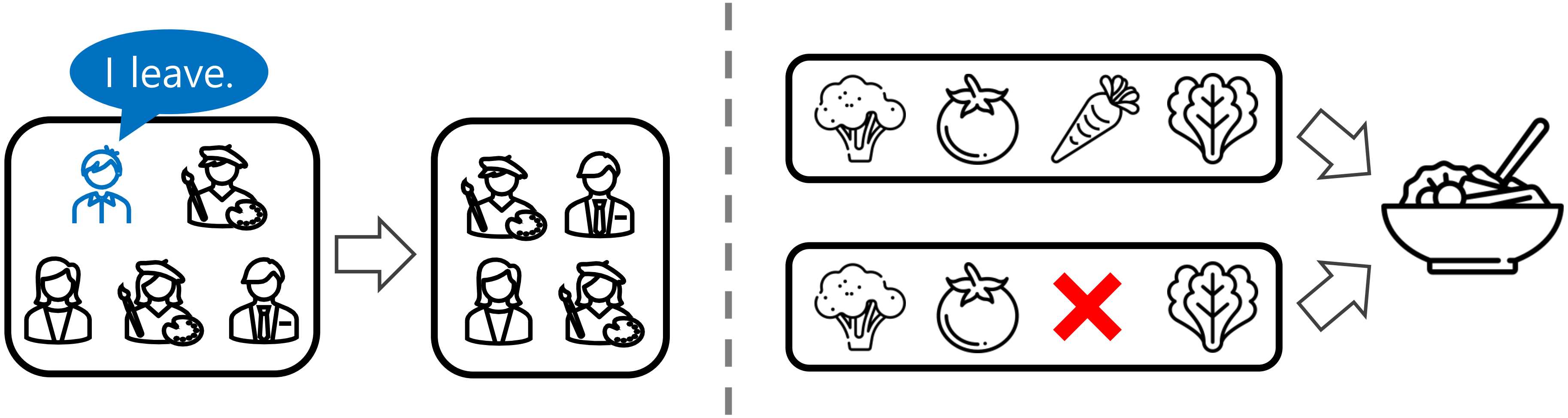
?

**Question:**

How can we define ***k*-cores** in hypergraphs?

**Answer (cont.):**

In many cases, group interactions are **non-fragile**.

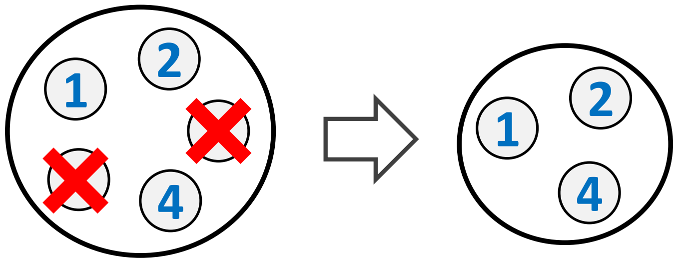


!

F. Bu, G. Lee, K. Shin, "Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications", **Data Mining and Knowledge Discovery (2023)**

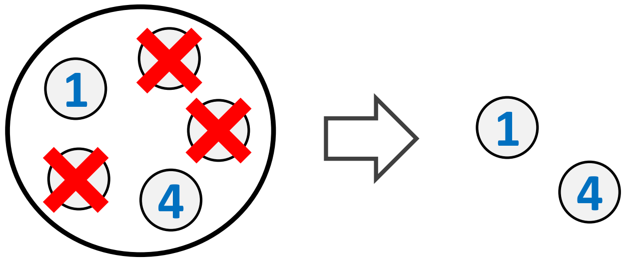
# Non-Fragile Hyperedges

- **Non-fragile hyperedges** break if at least $t$ fraction of the nodes remain.
  - The larger the value of $t$ is, the more fragile the hyperedges are.
- Example: $t = 0.5$



**60%** of the nodes remain

Degrees of nodes 1, 2, 4
are **unchanged**.



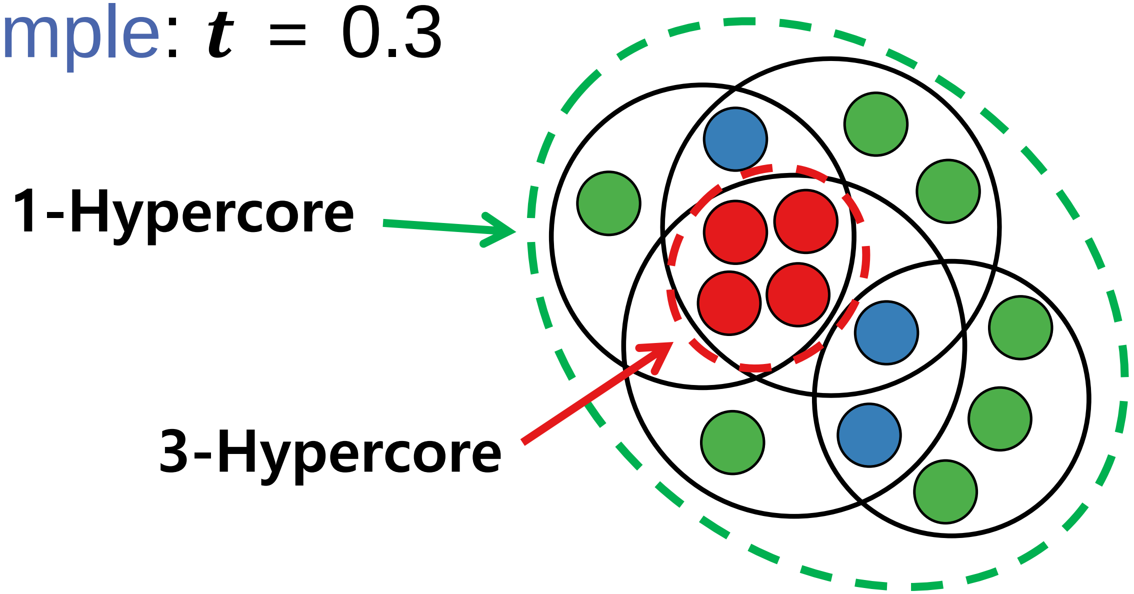
**40%** of the nodes remain

Degrees of nodes 1 and 4
are **decreased by 1**.

F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Hypercores for Non-Fragile Hyperedges

- $(\boldsymbol{k}, \boldsymbol{t})$**-hypercore** is the maximal sub-hypergraph of:
  - Nodes with degree at least $\boldsymbol{k}$
  - Hyperedges with at least $\boldsymbol{t}$ proportion of its nodes remaining
- Example: $\boldsymbol{t} = 0.3$



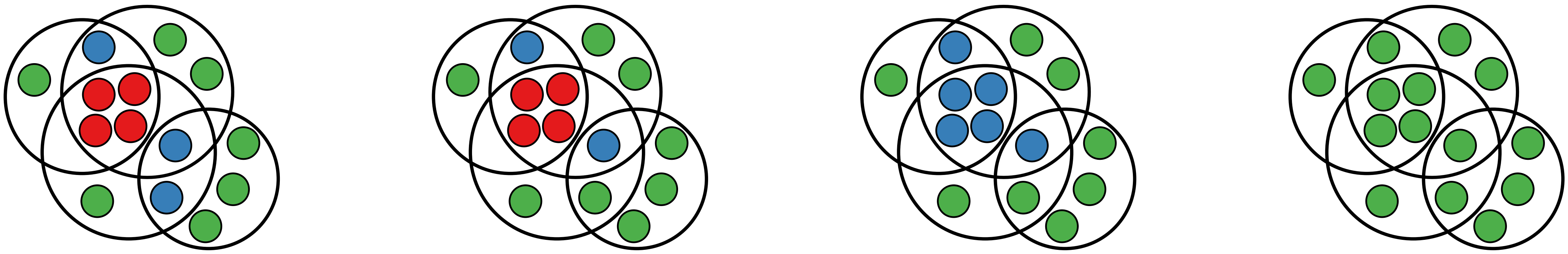
- 🟢 Nodes with **hypercoreness 1**
- 🔵 Nodes with **hypercoreness 2**
- 🔴 Nodes with **hypercoreness 3**

F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Hypercores for Non-Fragile Hyperedges (cont.)

- Different $t$ values give us different insights into cohesive structures.

🟢 hypercoreness 1 🔵 hypercoreness 2 🔴 hypercoreness 3

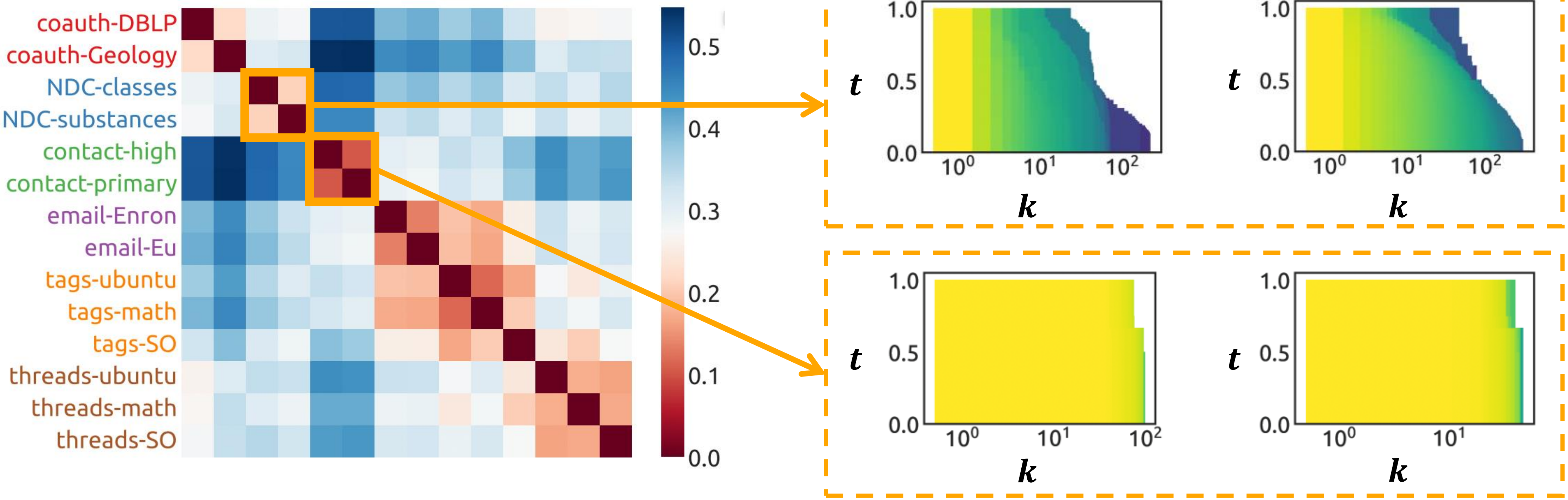


$t \leq \frac{2}{5}$ $\frac{2}{5} < t \leq \frac{4}{7}$ $\frac{4}{7} < t \leq \frac{5}{7}$ $t > \frac{5}{7}$

F. Bu, G. Lee, K. Shin, "Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications", **Data Mining and Knowledge Discovery (2023)**

# Hypercore Sizes of Hypergraphs

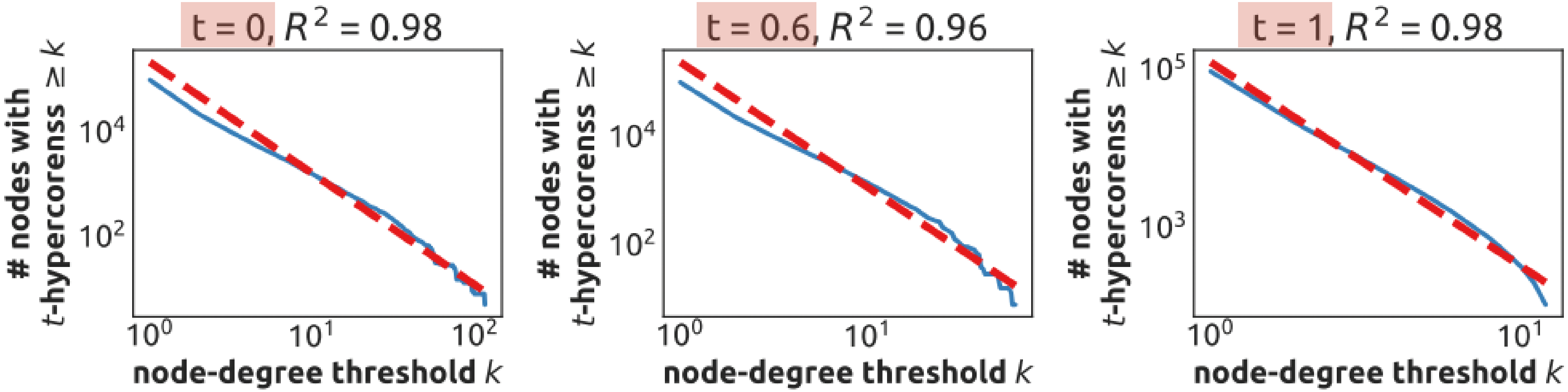
- Hypergraphs from the same domain tend to have similar **hypercore size distribution**, while they vary across domains.



F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Distributions of Hypercoreness

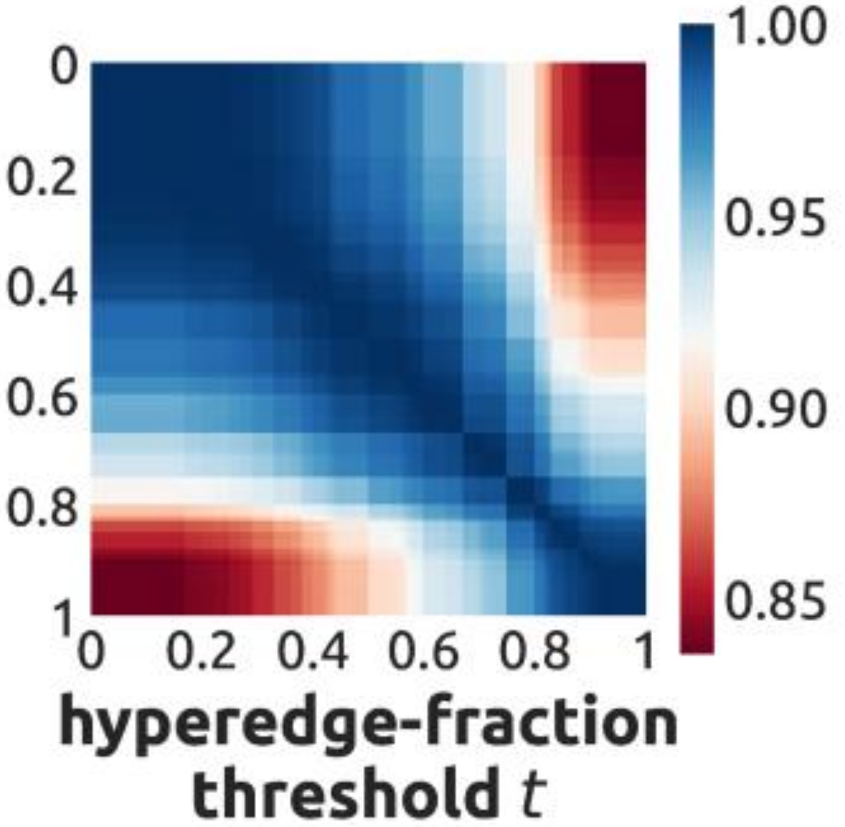
- ***t*-hypercoreness** of nodes in real-world hypergraphs follows heavy-tailed distributions regardless of ***t***.



F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Heterogeneity of Hypercoreness

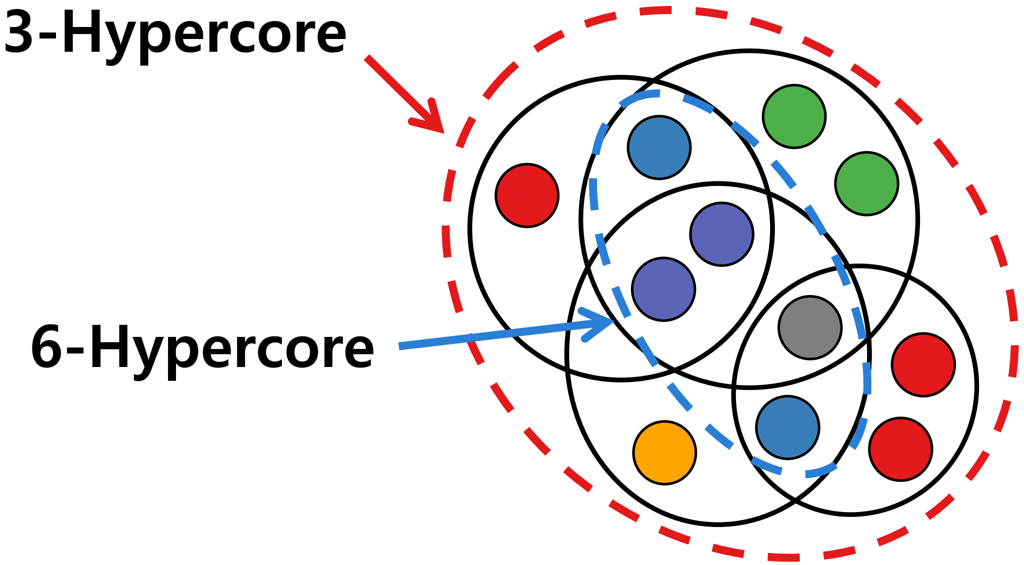
- ***t*-hypercoreness** of nodes gives distinct information depending on ***t***.



F. Bu, G. Lee, K. Shin, “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications”, **Data Mining and Knowledge Discovery (2023)**

# Other Definition of Hypercores

- **Neighborhood-based hypercore** is the maximal sub-hypergraph of every node having at least a certain number of neighbors.



● Nodes with **hypercoreness 3**
● Nodes with **hypercoreness 4**
● Nodes with **hypercoreness 5**
● Nodes with **hypercoreness 6**
● Nodes with **hypercoreness 7**
● Nodes with **hypercoreness 8**
● Nodes with **hypercoreness 9**

N. A. Arafat, A. Khan, A. K. Rai, and B. Ghosh, “Neighborhood-based Hypergraph Core Decomposition”, **VLDB 2023**

# References

1. [KKS20] Kook, Yunbum, Jihoon Ko, and Kijung Shin. “Evolution of Real-world Hypergraphs: Patterns and Models without Oracles.” ICDM 2020.
2. [LCK21] Lee, Geon, Minyoung Choe, and Kijung Shin. “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators.” WWW 2021.
3. [DYHS20] Do, Manh Tuan, et al. “Structural Patterns and Generative Models of Real-world Hypergraphs.” KDD 2020.
4. [KBCYS23] Kim, Sunwoo, Fanchen Bu, Minyoung Choe, Jaemin Yoo, and Kijung Shin. “How Transitive Are Real-World Group Interactions? Measurement and Reproduction.” KDD 2023.
5. [BASJK18] Benson, Austin R., et al. “Simplicial Closure and Higher-order Link Prediction.” PNAS 115(48):E11221–E11230, 2018.

# References (cont.)

6. [LMMB20] Lotito, Quintino Francesco, et al. “Higher-order Motif Analysis in Hypergraphs.” Communication Physics 5(1):1–8, 2022
7. [LKS20] Lee, Geon, Jihoon Ko, and Kijung Shin. “Hypergraph Motifs: Concepts, Algorithms, and Discoveries.” PVLDB 13(12):2256-2269, 2020.
8. [LL23] LaRock, Timothy, and Renaud Lambiotte. “Encapsulation Structure and Dynamics in Hypergraphs.” arXiv (2023).
9. [BLS23] Bu, Fanchen, Geon Lee, and Kijung Shin. “Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications.” Data Mining and Knowledge Discovery (2023).
10. [AKRG23] Arafat, Naheed, et al. “Neighborhood-Based Hypergraph Core Decomposition.” VLDB 2023.