



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

Part 0. Introduction



Geon Lee



Jaemin Yoo



Kijung Shin

Group Interactions are Everywhere

E.g., 1: Collaborations of researchers

Hypergraph Motifs: Concepts, Algorithms, and Discoveries

Geon Lee
KAIST AI
geonlee0325@kaist.ac.kr

Jihoon Ko
KAIST AI
jihoonko@kaist.ac.kr

Kijung Shin
KAIST AI & EE
kijungs@kaist.ac.kr

On the Persistence of Higher-Order Interactions in Real-World Hypergraphs

Hyunjin Choo*

Kijung Shin[†]

MiDaS: Representative Sampling from Real-world Hypergraphs

Minyoung Choe
KAIST
minyoung.choe@kaist.ac.kr

Jaemin Yoo
Seoul National University
jaeminyoo@snu.ac.kr

Geon Lee
KAIST
geonlee0325@kaist.ac.kr

Woonsung Baek
KAIST
wbaek@kaist.ac.kr

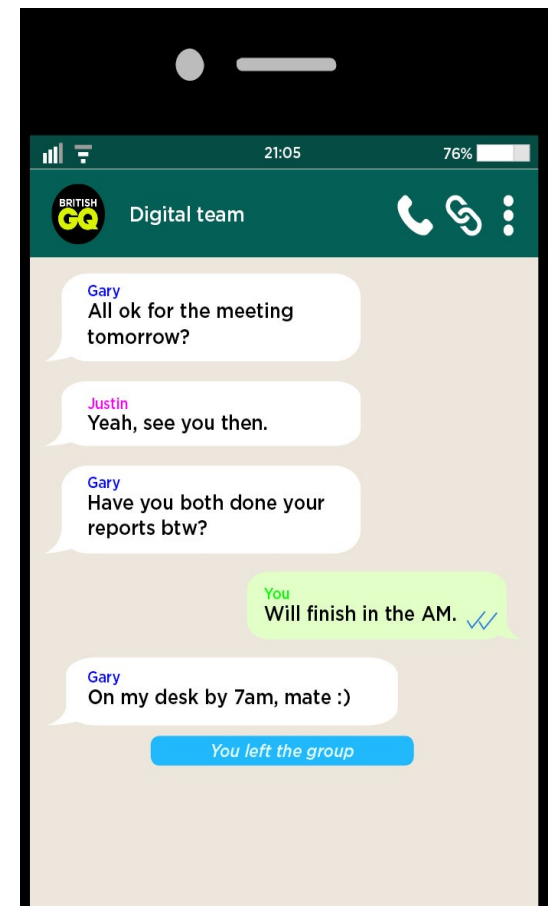
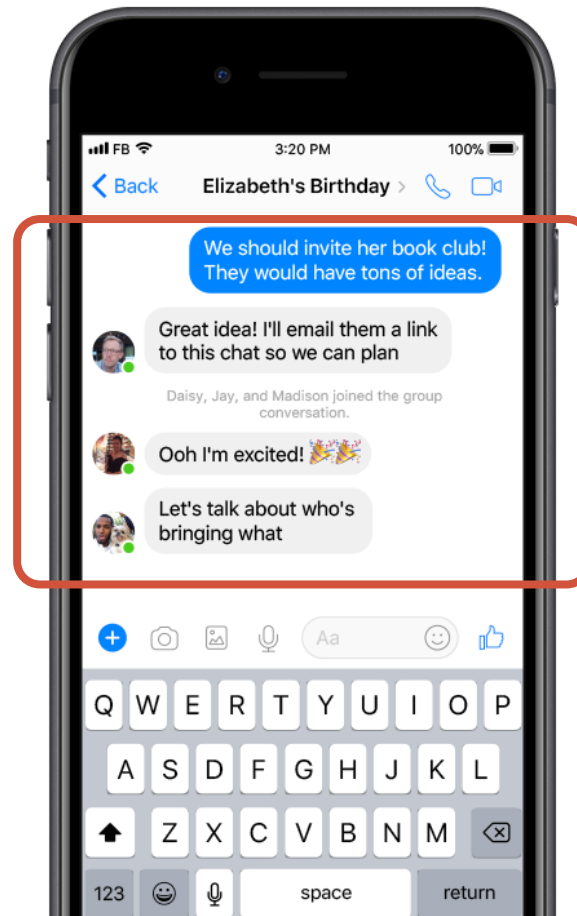
U Kang
Seoul National University
ukang@snu.ac.kr

Kijung Shin*
KAIST
kijungs@kaist.ac.kr



Group Interactions are Everywhere (cont.)

E.g., 2: Group chats on a messenger



Group Interactions are Everywhere (cont.)


E.g., 3: Co-purchases of items



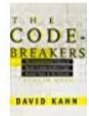
Your Orders [Search Orders](#)

[Orders](#) [Open Orders](#) [Digital Orders](#) [Cancelled Orders](#)


ORDER PLACED April 30, 1999 ORDER # 002-3316565-7682042 [Order Details](#) [Invoice](#)



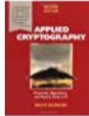
Windows NT File System Internals : A Developer's Guide
Rajeev Nagar
Sold by: Amazon.com Services, Inc
\$39.96
[Buy it again](#)



The Codebreakers; The Comprehensive History of Secret Communication from Ancient Times to the Internet
David Kahn
Sold by: Amazon.com Services, Inc
\$45.50
[Buy it again](#)



The Art of Systems Architecting (Systems Engineering Series)
Eberhardt Rechtin, Mark W. Maier
Sold by: Amazon.com Services, Inc
\$59.95
[Buy it again](#)



Applied Cryptography : Protocols, Algorithms, and Source Code in C
Bruce Schneier
Sold by: Amazon.com Services, Inc
\$75.00
[Buy it again](#)

Group Interactions are Everywhere (cont.)

E.g., 4: Tags in online Q&A sites

The screenshot shows a Stack Overflow question titled "Finding the index of an item in a list". The question is asked 13 years, 9 months ago, modified 2 months ago, and viewed 5.5m times. The question text is: "Given a list `["foo", "bar", "baz"]` and an item in the list `"bar"`, how do I get its index `1`?". Below the question, there are three tags: `python`, `list`, and `indexing`, which are highlighted with a red box. The question has 4072 votes and 668 answers. The user who asked the question is Eugene M, with a reputation of 45k. The user who edited the question is Mateen Ulhaq, with a reputation of 21.7k. The question was asked on Oct 7, 2008 at 1:39 and edited on Mar 28 at 12:01.

Hypergraphs Model Group Interactions

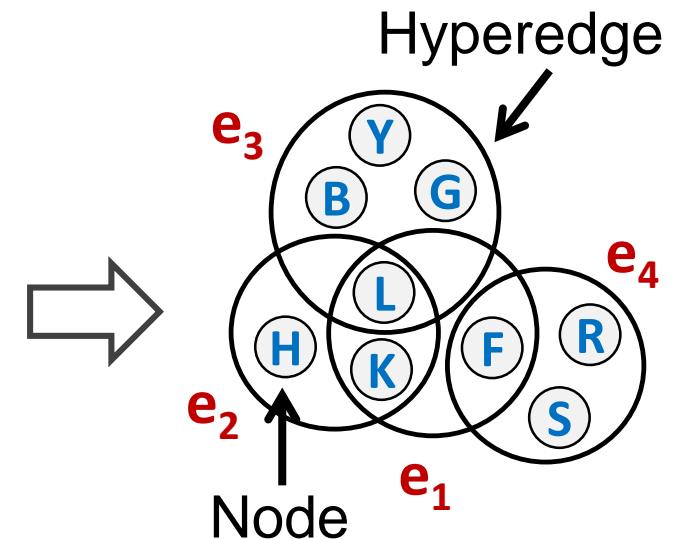
- **Hypergraphs** model group interactions among individuals or objects.
- Each **hyperedge** is a subset of any number of nodes.
- Each hyperedge indicates a **group interaction** among its members.

Authors (Nodes)

Jure Leskovec (L)	Austin Benson (B)
Jon Kleinberg (K)	David Gleich (G)
Hao Yin (Y)	Timos Sellis (S)
Christos Faloutsos (F)	Nick Roussopoulos (R)
Daniel Huttenlocher (H)	

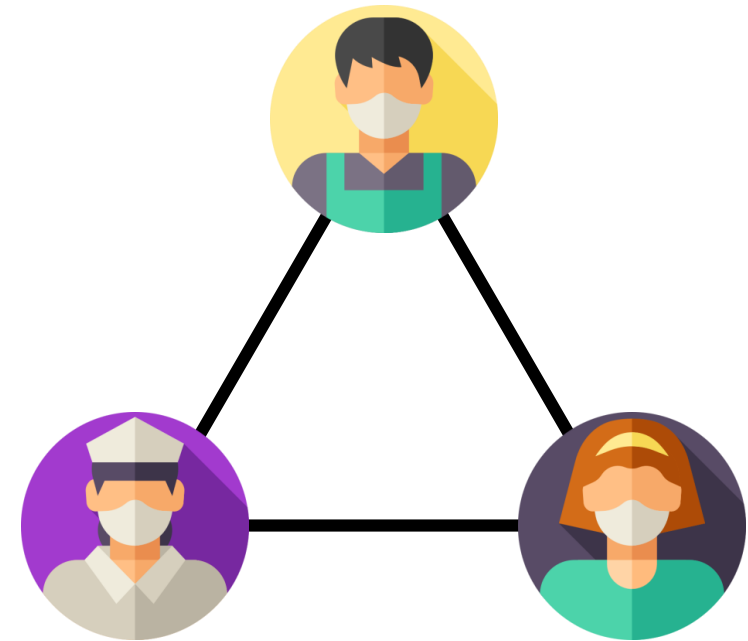
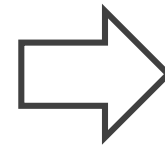
Publications (Hyperedges)

e_1 : (L, K, F) KDD'05
e_2 : (L, H, K) WWW'10
e_3 : (Y, B, G, L) KDD'17
e_4 : (S, R, F) VLDB'87



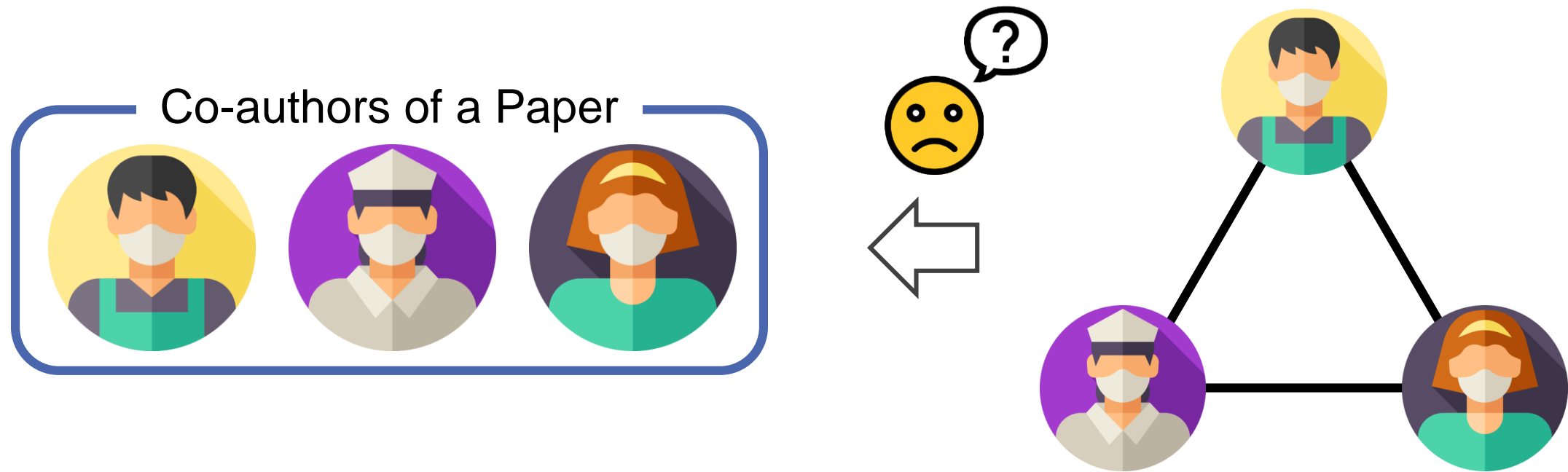
Limitations of Graphs

- **Graphs** can only model pairwise relations by **edges**.
- **Example:** Co-authorship



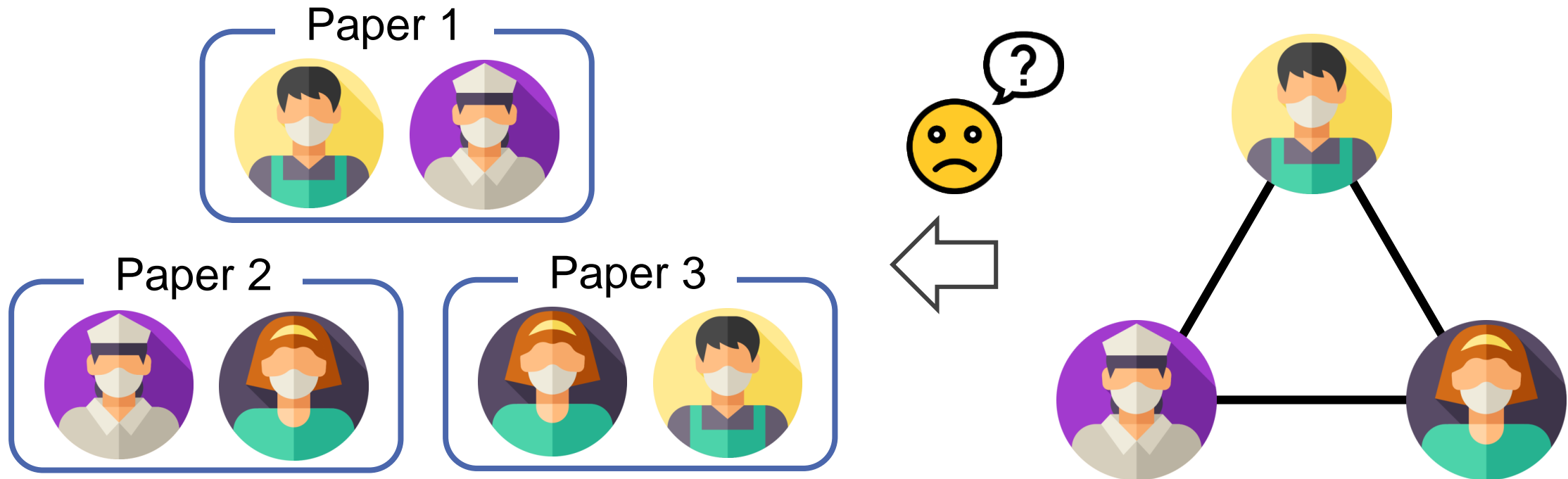
Limitations of Graphs (cont.)

- Simple reduction to pairwise relations causes information loss.
- **Example:** Did the three authors co-work as a group?



Limitations of Graphs (cont.)

- **Example:** The three authors may have never co-worked in the past.



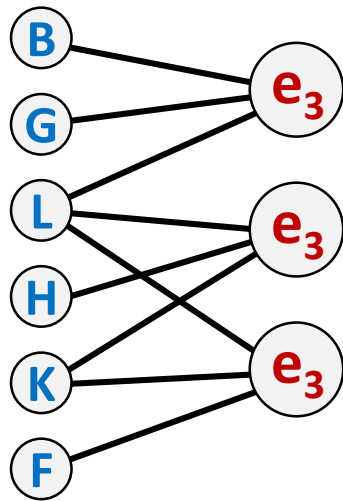
Power of Hypergraphs

- Hypergraph modeling is often fruitful, compared to graph modeling.
 - Classification [LKS20] [FYZJG19]
 - Ranking [CR19]
 - Link prediction [YSSY20]
 - Anomaly detection [LCS22]
- Refer to [TBBE21] for a comprehensive comparison of modeling methods

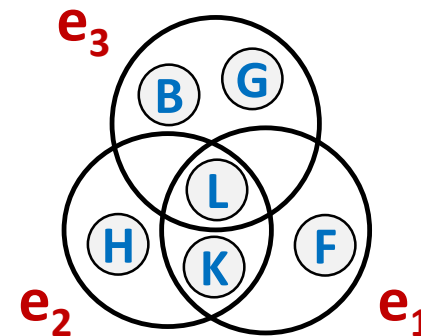
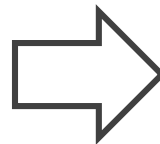


Power of Hypergraphs (cont.)

- Hypergraph can be transformed into bipartite graphs (“set of edges”)
- However, hypergraph modeling provides a new perspective (“set of sets”) and poses new questions overlooked for graphs



“set of edges” (memberships)



“set of sets”

Public Hypergraph Datasets

Dr. Austin Benson has publicly released real-world hypergraph datasets at <https://www.cs.cornell.edu/~arb/data/>.



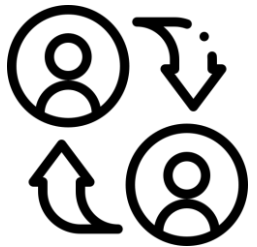
Co-authorship

(sets of authors of publications)



E-mail

(sets of email addresses on emails)



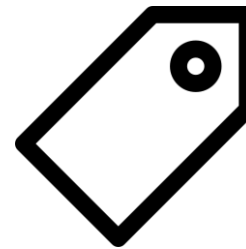
Contact

(groups of people in contact)



Threads

(sets of users asking and answering questions on threads)



Tags

(sets of tags attached to questions)






Drugs

(sets of substances making up drugs /
sets of classifications applied to drugs)




Public Hypergraph Datasets (cont.)

- Statistics of Benson's datasets

Domain	Dataset	Number of nodes	Number of static hyperedges	Number of temporal hyperedges
	coauth-DBLP	1,924,991	2,599,087	3,700,067
	coauth-MAG-Geology	1,256,385	1,207,390	1,590,335
	coauth-MAG-History	1,014,734	895,668	1,812,511
	tags-stack-overflow	49,998	5,675,497	14,458,875
	tags-math-sx	1,629	174,933	822,059
	tags-ask-ubuntu	3,029	151,441	271,233
	threads-stack-overflow	2,675,955	9,705,709	11,305,343
	threads-math-sx	176,445	595,778	719,792
	threads-ask-ubuntu	125,602	167,001	192,947

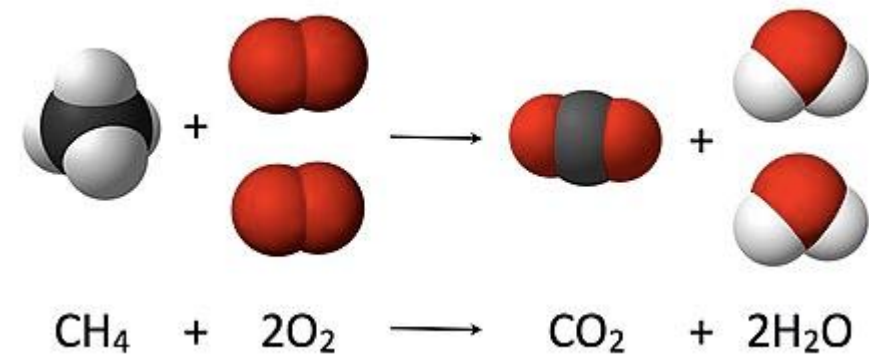
Public Hypergraph Datasets (cont.)

- Statistics of Benson's datasets

Domain	Dataset	Number of nodes	Number of static hyperedges	Number of temporal hyperedges
	NDC-substances	5,311	10,025	112,405
	NDC-classes	1,161	1,222	49,724
	email-Eu	998	25,791	234,760
	email-Enron	143	1,542	10,883
	contact-high-school	327	7,937	172,035
	contact-primary-school	242	12,799	106,879
Others	congress-bills	1,718	85,082	260,851
	DAWN	2,558	143,523	2,272,433

Additional Hypergraph Datasets

- **Large-scale hypergraph** datasets [KLKPHS23]
 - About 30 million hyperedges, node features, and node labels
 - <https://github.com/kswoo97/pcl>
- **Directed hypergraph** datasets [KCY22]
 - Directed group interactions (e.g., chemical reactions)
 - 11 hypergraphs from 6 domains
 - <https://github.com/kswoo97/hyprec>



Open-source Software

- Open-source software for hypergraph mining & generation

Reference	URL	License
C20	https://github.com/PhilChodrow/hypergraph	MIT
BKT18	https://github.com/arbenson/Sequences-of-Sets	-
KKS20	https://github.com/yunbum-kook/icdm20-hyperff	GPL-3.0
DYHS20	https://github.com/manhtuando97/KDD-20-Hypergraph	-
CK21	https://github.com/Cazamere/hypergraph-assembly	-
LCS20	https://github.com/young917/www21-hyperlap	GPL-3.0
KBCYS23	https://github.com/kswoo97/hypertrans	-
CYLBKS22	https://github.com/young917/MiDaS	-
LKS20	https://github.com/geonlee0325/MoCHy	GPL-3.0
LL23	https://github.com/tlarock/encapsulation-dynamics/	MIT
LMMB22	https://github.com/FraLotito/higher-order-motifs	MIT
CS22	https://github.com/jin-choo/persistence	-
BASJK18	https://github.com/arbenson/ScHoLP-Tutorial	-
LS21	https://github.com/geonlee0325/THyMe	GPL-3.0

Our Focus: Hypergraph Mining

To better understand group interactions in the real world
through hypergraph modeling and analysis



Our Focus: Hypergraph Mining (cont.)

To better understand group interactions in the real world through hypergraph modeling and analysis

- **Part 1. Static structural patterns**
 - *What do real-world hypergraphs look like?*
- **Part 2. Dynamic structural patterns**
 - *How do real-world hypergraphs evolve over time?*
- **Part 3. Generative models**
 - *How can we generate realistic hypergraphs?*
 - *What are underlying mechanisms that lead to the patterns?*

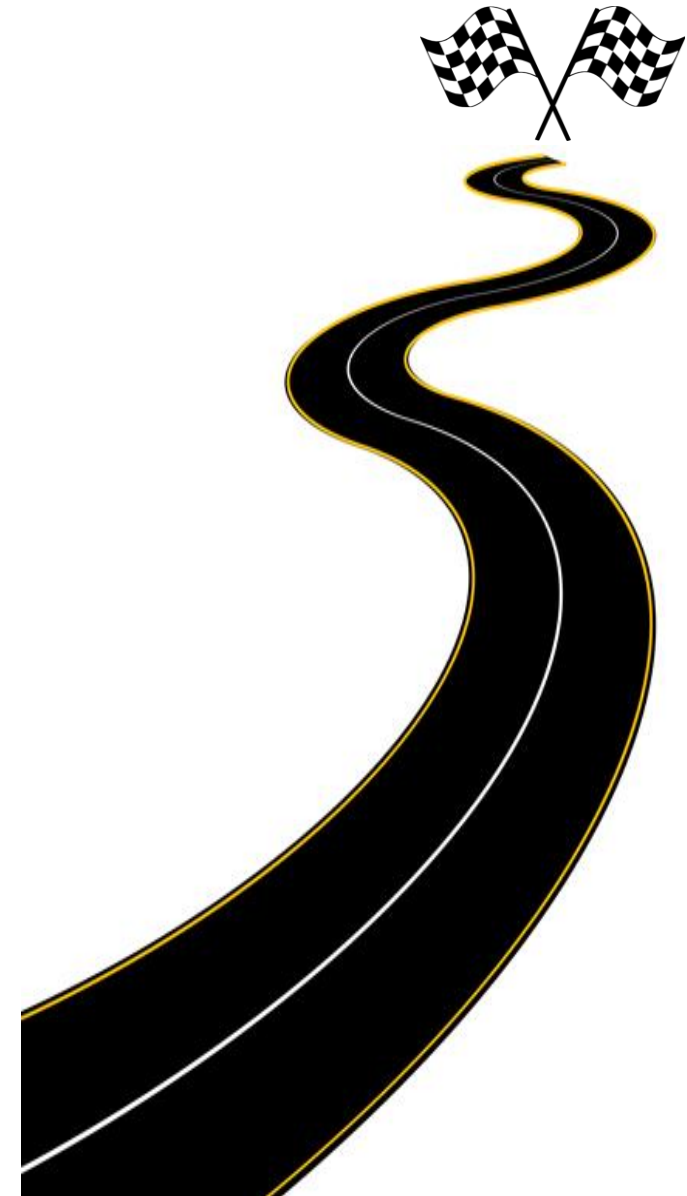
Why are These Important?

Structural patterns and **generative models** are crucial for understanding and utilizing hypergraph-structured data

- **Prediction about complex systems**
 - *How do group interactions evolve over time?*
- **Anomaly detection**
 - *Are nodes and hyperedges naturally structured?*
- **Algorithm design**
 - *Which structural properties can be exploited for fast algorithm design?*

Roadmap

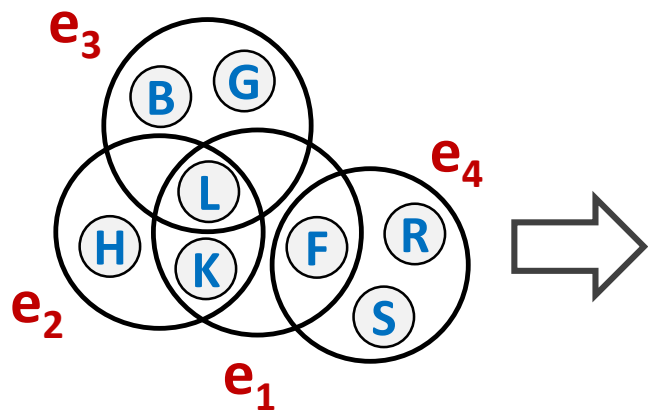
- **Part 1. Static Structural Patterns**
 - Basic Patterns
 - Advanced Patterns
- **Part 2. Dynamic Structural Patterns**
 - Basic Patterns
 - Advanced Patterns
- **Part 3. Generative Models**
 - Static hypergraph Generator
 - Dynamic hypergraph Generator



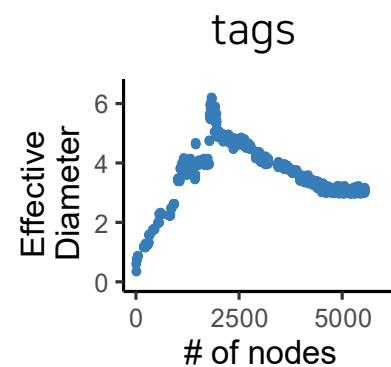
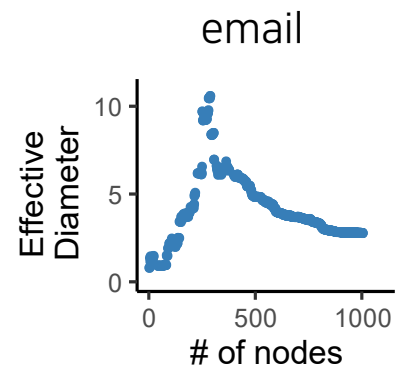
Part 1. Static Structural Patterns

“What do real-world hypergraphs look like?”

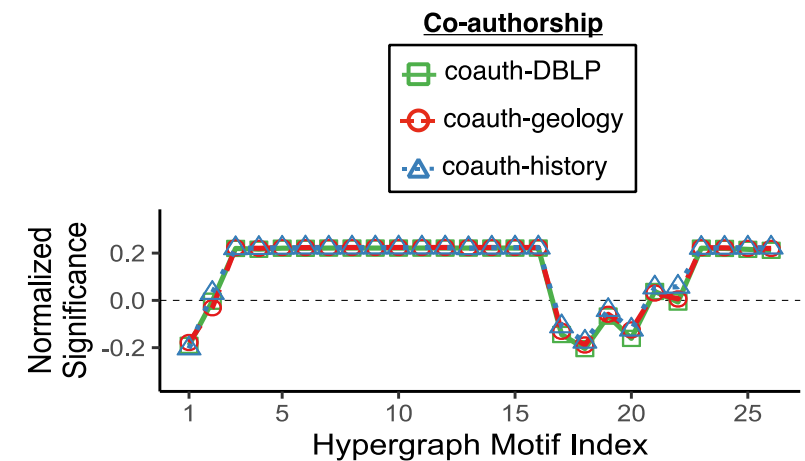
*“Given a **static** hypergraph, how can we analyze its structure?”*



Input Hypergraph



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



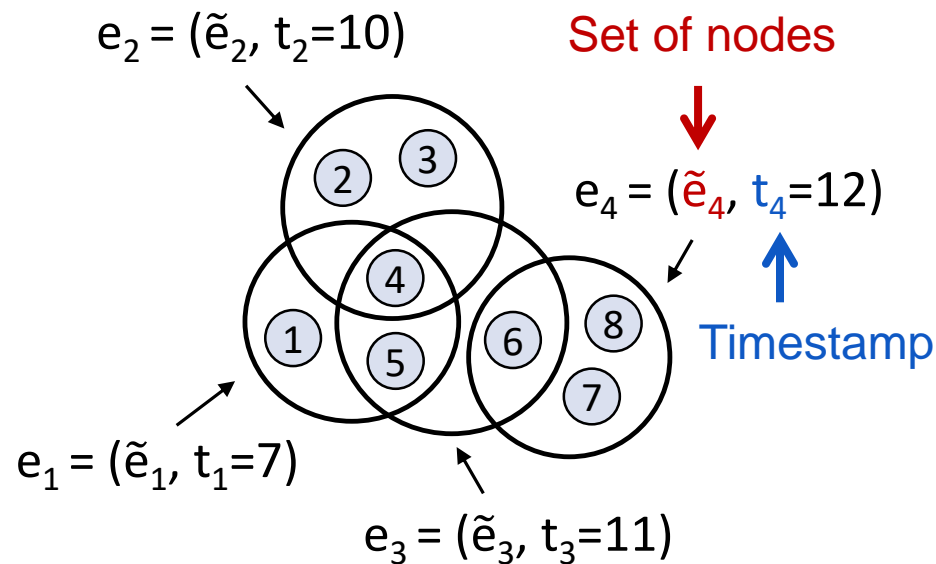
Advanced Patterns (**Part 1-2**)



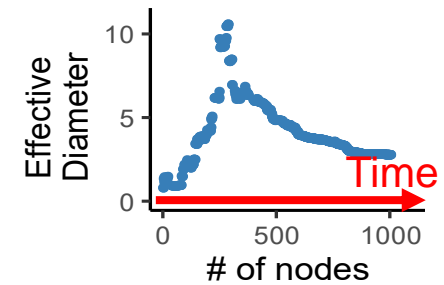
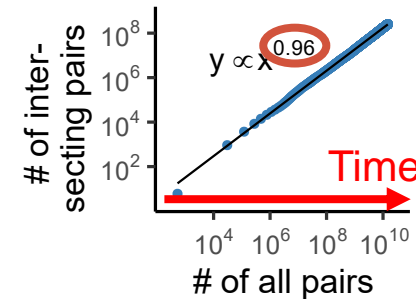
Part 2. Dynamic Structural Patterns

“How do real-world hypergraphs evolve over time?”

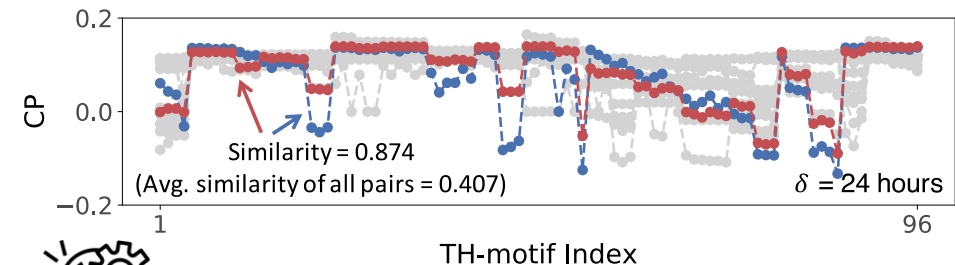
*“Given a **dynamic** hypergraph, how can we analyze its structure?”*



Input Temporal Hypergraph



Basic Patterns (Part 2-1)

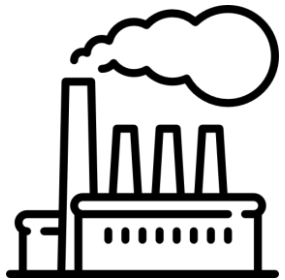


Advanced Patterns (Part 2-2)

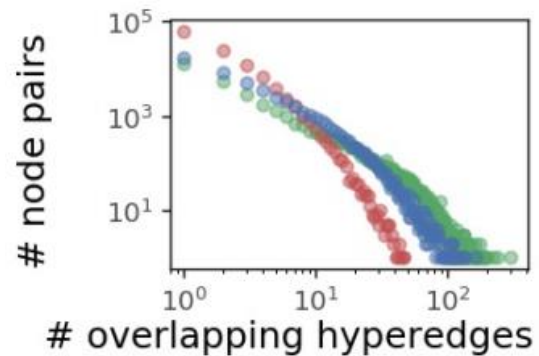
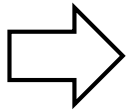
Part 3. Generative Models

*“How can we generate **realistic hypergraphs**?”*

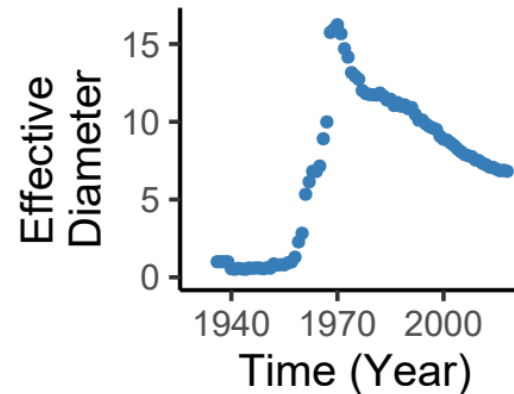
*“What are **underlying mechanisms** that lead to the observed patterns?”*



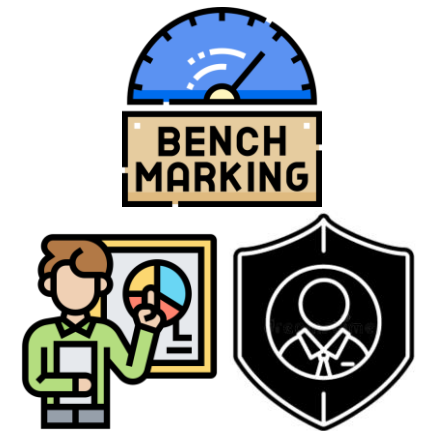
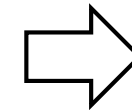
Generative
models



Static Hypergraphs
(Part 3-1)



Dynamic Hypergraphs
(Part 3-2)



Applications

Tutorial Materials

- <https://sites.google.com/view/hypergraph-tutorial>
 - Slides  
 - Videos 
 - Code and Datasets  

References

1. [CR19] Chitra, Uthsav, and Benjamin Raphael. "Random walks on hypergraphs with edge-dependent vertex weights." ICML 2019.
2. [FYZJG19] Feng, Yifan, et al. "Hypergraph neural networks." AAAI 2019
3. [LCS22] Lee, Geon, Minyoung Choe, and Kijung Shin, "HashNWalk: Hash and Random Walk Based Anomaly Detection in Hyperedge Streams." IJCAI 2022.
4. [LKS20] Lee, Geon, Jihoon Ko, and Kijung Shin. "Hypergraph motifs: concepts, algorithms, and discoveries." PVLDB 13(12):2256-2269, 2020.
5. [TBBE21] Torres, Leo, et al. "The why, how, and when of representations for complex systems." SIAM Review 63(3):435-485, 2021.
6. [YSSY20] Yoon, Se-eun, et al. "How much and when do we need higher-order information in hypergraphs? a case study on hyperedge prediction." WWW 2020.

References (Datasets)

1. [BASJK18] Benson, Austin R., et al. "Simplicial Closure and Higher-order Link Prediction." PNAS 115(48):E11221–E11230, 2018.
2. [KCYS22] Kim, Sunwoo, et al. "Reciprocity in Directed Hypergraphs: Measures, Findings, and Generators." ICDM 2022.
3. [KLKPHS23] Kim, Sunwoo et al. " Datasets, Tasks, and Training Methods for Large-Scale Hypergraph Learning." Data Mining and Knowledge Discovery, 2023.

References (Open-source Software)

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. [C20] Philip S. Chodrow, “Configuration Models of Random Hypergraphs.” Complex Networks, 2020.
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. [CYLBKS22] Choe, Minyoung, et al. “MiDaS: Representative Sampling from Real-world Hypergraphs.” WWW 2022.

References (Open-source Software) (cont.)

7. [KBCYS23] Kim, Sunwoo, et al. “How Transitive Are Real-World Group Interactions? - Measurement and Reproduction.” KDD 2023.
8. [DYHS20] Do, Manh Tuan, et al. “Structural Patterns and Generative Models of Real-world Hypergraphs.” KDD 2020.
9. [KKS20] Kook, Yunbum, Jihoon Ko, and Kijung Shin. “Evolution of Real-world Hypergraphs: Patterns and Models without Oracles.” ICDM 2020.
10. [LKS20] Lee, Geon, Jihoon Ko, and Kijung Shin. “Hypergraph Motifs: Concepts, Algorithms, and Discoveries.” PVLDB 13(12):2256-2269, 2020.
11. [LCS21] Lee, Geon, Minyoung Choe, and Kijung Shin. “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators.” WWW 2021.

References (Open-source Software) (cont.)

12. [LL23] LaRock, Timothy, and Renaud Lambiotte. “Encapsulation Structure and Dynamics in Hypergraphs.” arXiv, 2023.
13. [LMMB20] Lotito, Quintino Francesco, et al. “Higher-order Motif Analysis in Hypergraphs.” Communication Physics 5(1):1–8, 2022.
8. [LS21] Lee, Geon, and Kijung Shin. “THyMe+: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting.” ICDM 2021.