This is a preliminary version of the slides that will be used for tutorials.

The slides will be revised to reflect recent studies and recommended improvements.

The final version may differ from this version.

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

Part I. Introduction

Geon Lee

Jaemin Yoo

Kijung Shin

Group Interactions are Everywhere

E.g., 1: Collaborations of researchers

Authors

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

Publications

(L, K, F) KDD'05

(L, H, K) WWW'10

(B, G, L) Science'16

(S, R, F) VLDB'87

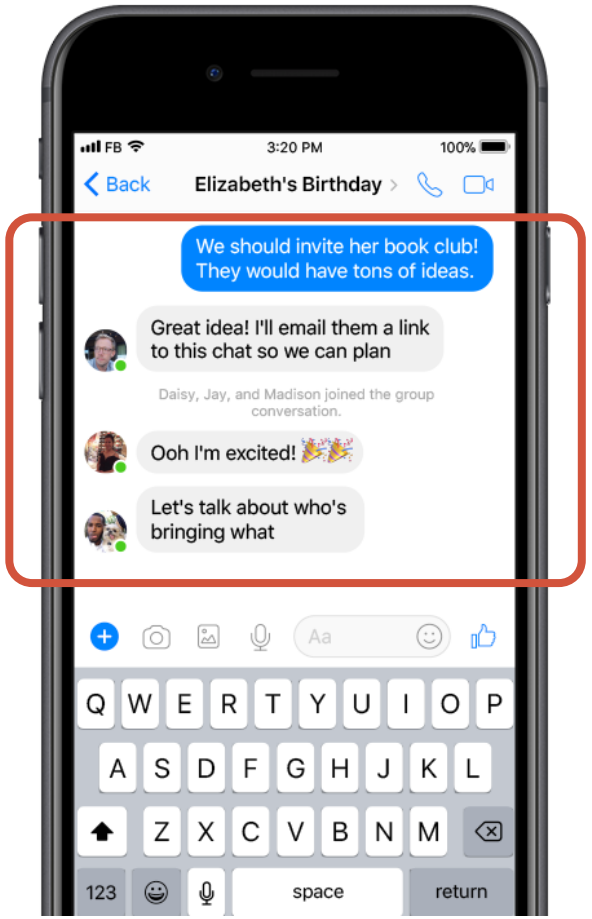
NETWORK SCIENCE

Higher-order organization of complex networks

Austin R. Benson,[1] David F. Gleich,[2] Jure Leskovec[3]*

Group Interactions are Everywhere (cont.)

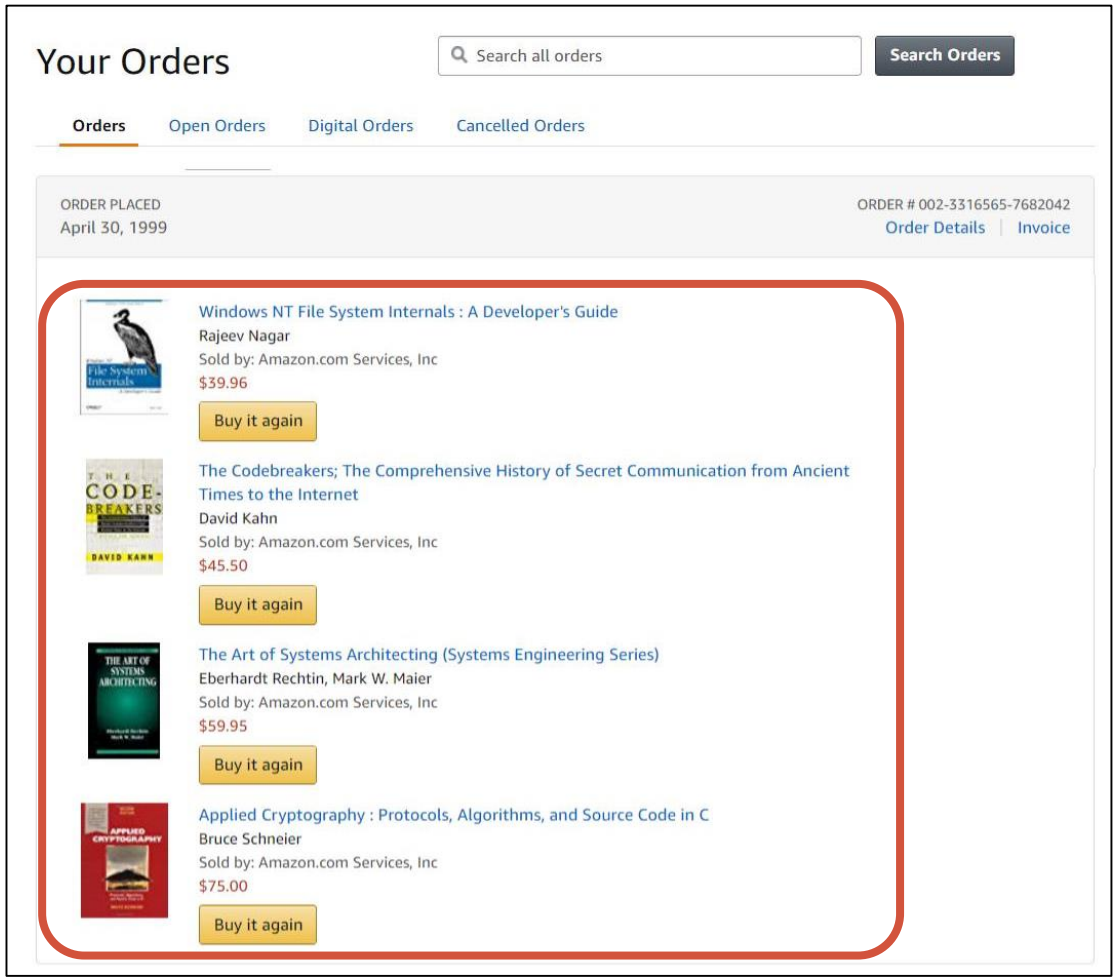
E.g., 2: Group chats on a messenger



Group Interactions are Everywhere (cont.)

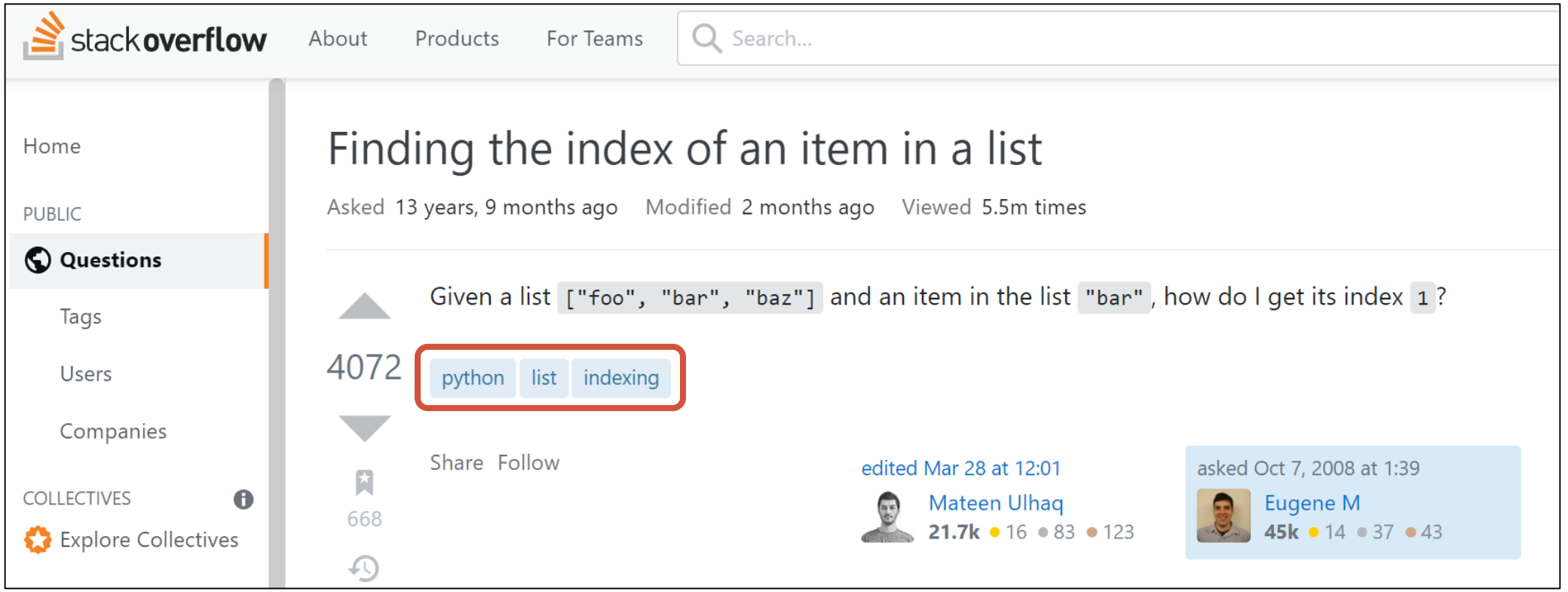
E.g., 3: Co-purchases of items

amazon



Group Interactions are Everywhere (cont.)

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

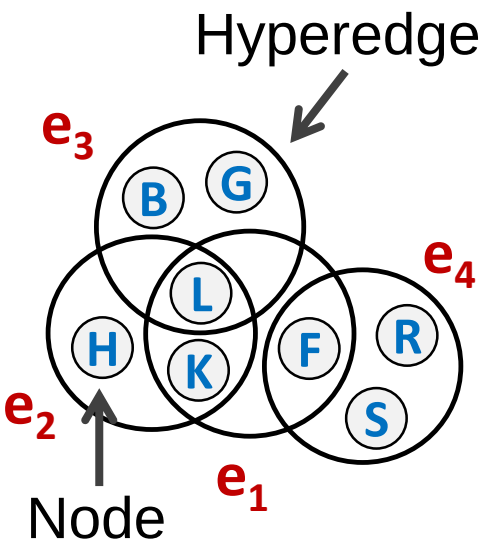


Hypergraphs Model Group Interactions

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

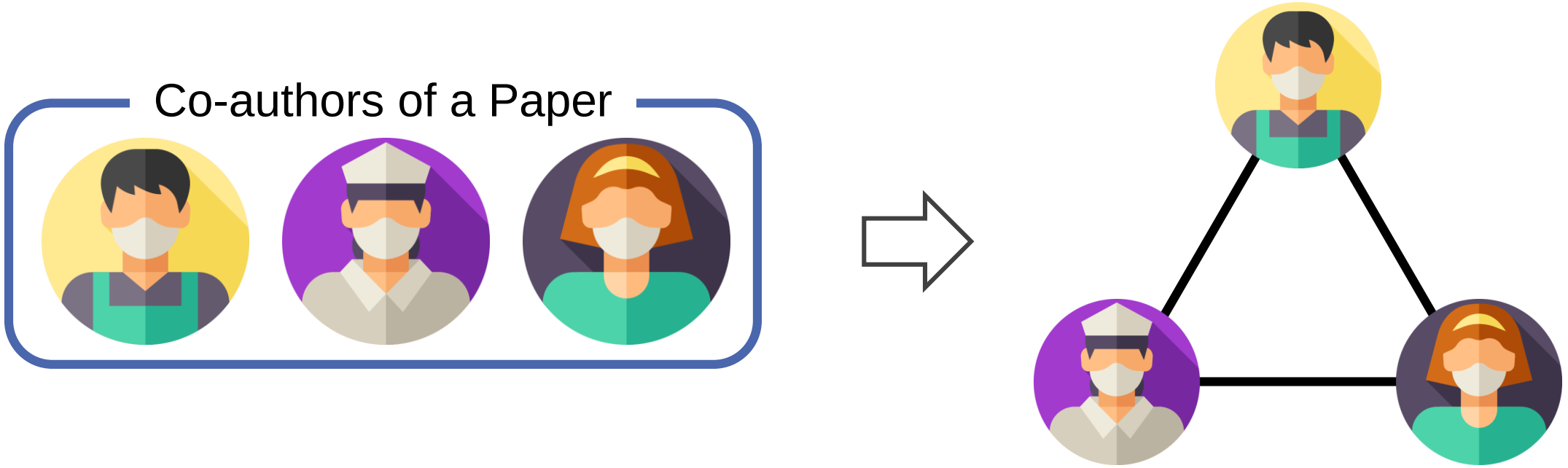
Authors (Nodes)	
Jure Leskovec (L)	Austin Benson (B)
Jon Kleinberg (K)	David Gleich (G)
Christos Faloutsos (F)	Timos Sellis (S)
Daniel Huttenlocher (H)	Nick Roussopoulos (R)

Publications (Hyperedges)
e_1: (L, K, F) KDD'05
e_2: (L, H, K) WWW'10
e_3: (B, G, L) Science'16
e_4: (S, R, F) VLDB'87



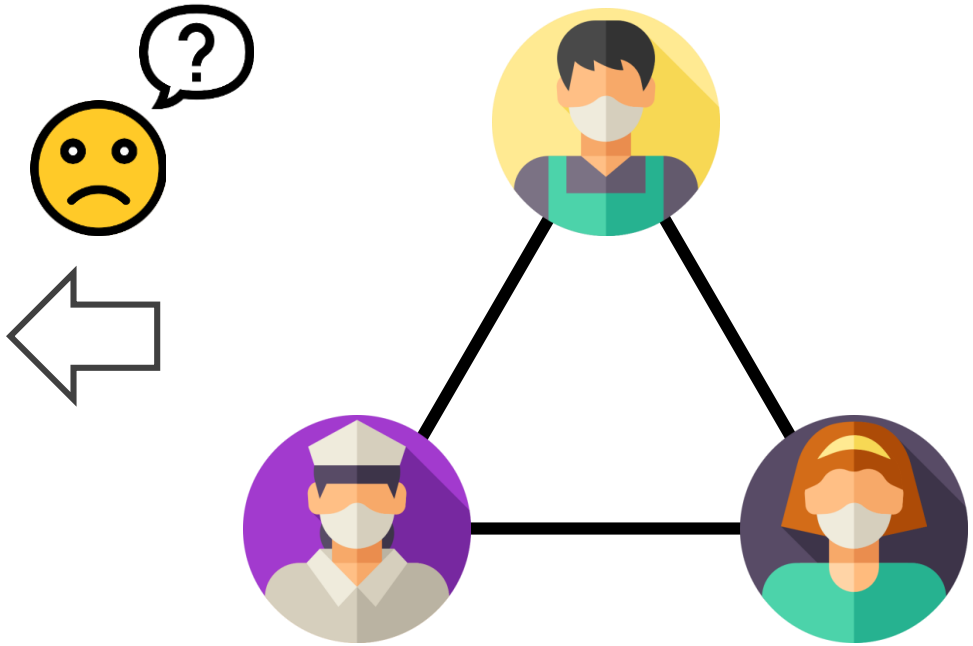
Limitations of Graphs

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



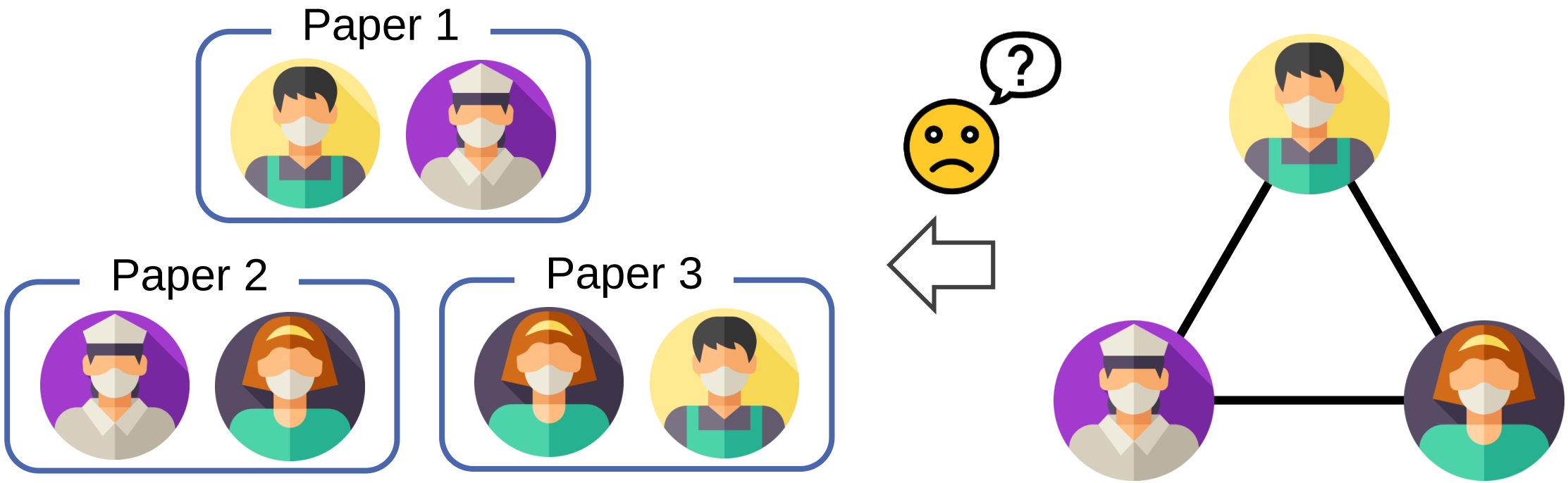
Limitations of Graphs (cont.)

- Simple reduction to pairwise relations loses group information.
- 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.

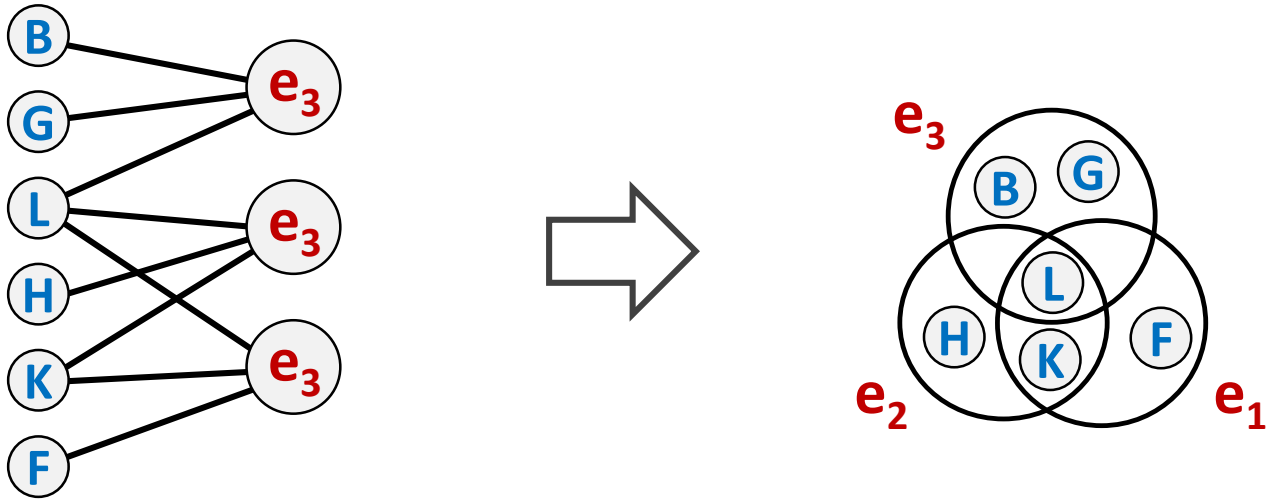


Power of Hypergraphs

- Hypergraph modeling is often fruitful, compared to graph modeling.
  - Classification [LKS20] [FYZJG19]
  - Ranking [CR19]
  - Link prediction [YSSY20]
  - Anomaly detection [LCS22]

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

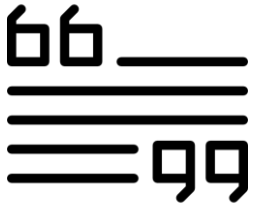


"set of edges" (memberships)

"set of sets"

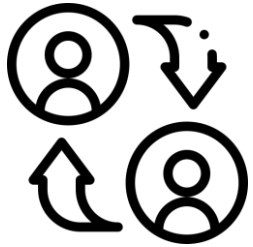
Public Hypergraph Datasets

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



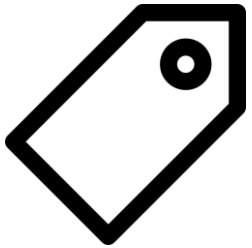
Co-authorship
(sets of authors of publications)

E-mail
(sets of email addresses on emails)



Contact
(groups of people in contact)

Threads
(sets of users asking and answering questions on threads)

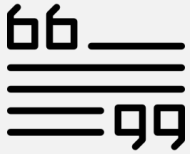
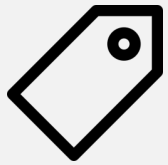


Tags
(sets of tags attached to questions)

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

Public Hypergraph Datasets (cont.)

- Statistics of 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 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

Open-source Software

- Open-source software for hypergraph mining & Generation

Name	Reference	URL	Language	License
Configuration Model	C20	https://github.com/PhilChodrow/hypergraph		MIT
CRU	BKT18	https://github.com/arbenson/Sequences-of-Sets		-
HyperFF	KKS20	https://github.com/yunbum-kook/icdm20-hyperff		GPL-3.0
Hypergraph	DYHS20	https://github.com/manhtuando97/KDD-20-Hypergraph		-
Hypergraph-Assembly	CK21	https://github.com/Cazamere/hypergraph-assembly		-
HyperLap	LCS20	https://github.com/young917/www21-hyperlap		GPL-3.0
MiDaS	CYLBKS22	https://github.com/young917/MiDaS		-
MoCHy	LKS20	https://github.com/geonlee0325/MoCHy		GPL-3.0
Motif Analysis	LMMB22	https://github.com/FraLotito/higher-order-motifs		MIT
Persistence	CS22	https://github.com/jin-choo/persistence		-
ScHoLP	BASJK18	https://github.com/arbenson/ScHoLP-Tutorial		-
THyMe	LS21	https://github.com/geonlee0325/THyMe		GPL-3.0

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

Outline of This Tutorial

Structural Patterns in Hypergraphs (Part 1 & Part 2)

		Part 1. Static Patterns	Part 2. Dynamic Patterns
Basic Patterns	Node-Level	DYHS20, KKS20, LCS21	BKT18, CS22
	Hyperedge-Level	KKS20, LCS21	BKT18, CBLK21, LS21
	Hypergraph-Level	BASJK18, DYHS20, KKS20	KKS20
Advanced Patterns	Sub-hypergraph-Level	BASJK18, LMMB22, LKK20, LCS21	BASJK18, CJ21, LS21

Outline of This Tutorial (cont.)

Generative Models of Hypergraphs (Part 3)

		Part 3. Generative Models 
Static Models	Full-Hypergraphs	C20, LCS21
	Sub-Hypergraphs	DYHS20, KKS20
Dynamic Models	Full-Hypergraphs	CYLBKS22
	Sub-Hypergraphs	BKT18, CK21

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 (2020): 2256-2269.
5. [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 (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. [DYHS20] Do, Manh Tuan, et al. “Structural Patterns and Generative Models of Real-world Hypergraphs.” KDD 2020.
8. [KKS20] Kook, Yunbum, Jihoon Ko, and Kijung Shin. “Evolution of Real-world Hypergraphs: Patterns and Models without Oracles.” ICDM 2020.
9. [LKS20] Lee, Geon, Jihoon Ko, and Kijung Shin. “Hypergraph Motifs: Concepts, Algorithms, and Discoveries.” PVLDB 13(12):2256-2269, 2020.
10. [LCS21] Lee, Geon, Minyoung Choe, and Kijung Shin. “How Do Hyperedges Overlap in Real-world Hypergraphs? – Patterns, Measures, and Generators.” WWW 2021.
11. [LS21] Lee, Geon, and Kijung Shin. “THyMe+: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting.” ICDM 2021.
12. [LMMB20] Lotito, Quintino Francesco, et al. “Higher-order Motif Analysis in Hypergraphs.” Communication Physics 5(1):1–8, 2022.