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

Geon Lee
KAIST
Seoul, South Korea
geonlee0325@kaist.ac.kr

Jaemin Yoo
Carnegie Mellon University
Pittsburgh, PA, USA
jaeminyoo@cmu.edu

Kijung Shin
KAIST
Seoul, South Korea
kijungs@kaist.ac.kr

ABSTRACT

Group interactions are prevalent in various complex systems (e.g., collaborations of researchers and group discussions on online Q&A sites), and they are commonly modeled as hypergraphs. Hyperedges, which compose a hypergraph, are non-empty subsets of any number of nodes, and thus each hyperedge naturally represents a group interaction among entities. The higher-order nature of hypergraphs brings about unique structural properties that have not been considered in ordinary pairwise graphs.

In this tutorial, we offer a comprehensive overview of a new research topic called *hypergraph mining*. We first present recently revealed structural properties of real-world hypergraphs, including (a) static and dynamic patterns, (b) global and local patterns, and (c) connectivity and overlapping patterns. Together with the patterns, we describe advanced data mining tools used for their discovery. Lastly, we introduce simple yet realistic hypergraph generative models that provide an explanation of the structural properties.

CCS CONCEPTS

• **Information systems → Social networks**; **Data mining**.

KEYWORDS

hypergraphs, social networks, structure mining, graph generators

ACM Reference Format:
Geon Lee, Jaemin Yoo, and Kijung Shin. 2022. Mining of Real-world Hypergraphs: Patterns, Tools, and Generators. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3511808.3557505

1 BASIC INFORMATION

- The tutorial is half-day long, i.e., 3 hours with breaks.
- The slides for this tutorial are available at **https://sites.google.com/view/hypergraph-tutorial-cikm**.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557505

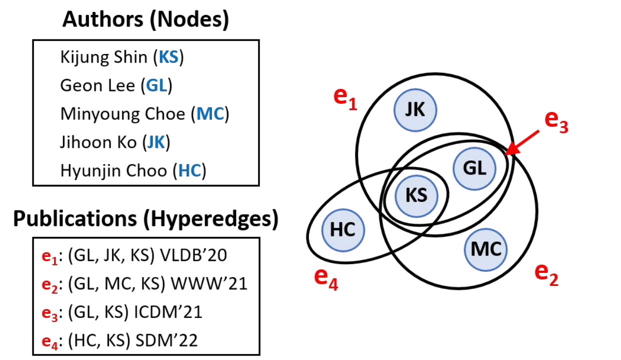

Figure 1: The co-authorship among five authors in four publications [7, 19–21] are represented as a hypergraph with five nodes and four hyperedges.

2 IMPORTANCE AND RELEVANCE

Group interactions are omnipresent in real-world complex systems: collaborations of researchers, joint interactions of proteins, co-purchases of items, to name a few. Such group interactions among entities are commonly modeled as a hypergraph, which consists of nodes and hyperedges (see Figure 1 for an example). A hyperedge, which is a non-empty subset of nodes, naturally models a group interaction among any number of entities. Thanks to the powerful expressiveness of hypergraphs, they have been used in a wide range of fields, including recommender systems [22, 30], computer vision [17, 27], natural language processing [9, 12], social network analysis [28], bioinformatics [14], and circuit designs [16].

Motivated by the successful investigation of structural patterns in real-world pairwise graphs (e.g., power-law degree distribution [1, 11], six degrees of separation [13, 15], and network motifs [24, 25]) and their wide range of applications, such patterns in real-world hypergraphs have been extensively studied recently. The flexibility in the size of each hyperedge, which provides the expressiveness of hypergraphs, brings about unique structural properties that have not been considered in pairwise graphs, and specialized tools have been developed to analyze their structural patterns. Moreover, several efforts have been made to reproduce and thus explain the patterns through intuitive hypergraph generative models.

In this half-day tutorial, we provide a comprehensive overview of structural patterns discovered in real-world hypergraphs, advanced data mining tools for hypergraphs, and hypergraph generative models based on the patterns.

While this topic *hypergraph mining* is in its infant stage, we believe it will be of interest of a much larger group of researchers,

- **Part I: Introduction**
  - Group interactions in the real-world
  - Power of hypergraph modeling [26, 29]
  - Data repositories and open-source software for hypergraph mining
- **Part II: Static Structural Patterns in Hypergraphs and Data Mining Tools for Their Discovery**
  - Basic patterns
    - Node-level properties [10, 18, 19]
    - Hyperedge-level properties [18, 19]
    - Hypergraph-level properties [2, 10, 18]
  - Advanced patterns (sub-hypergraph-level properties) [2, 19, 20, 23]
- **Part III: Dynamic Structural Patterns in Hypergraphs and Data Mining Tools for Their Discovery**
  - Basic patterns
    - Node-level properties [3, 7]
    - Hyperedge-level properties [3, 4, 21]
    - Hypergraph-level properties [18]
  - Advanced patterns (sub-hypergraph-level properties) [2, 8, 21]
- **Part IV: Generative Models of Hypergraphs**
  - Full-hypergraph generation
    - Static hypergraphs [5, 19]
    - Dynamic hypergraphs [10, 18]
  - Sub-hypergraph generation
    - Static sub-hypergraphs [6]
    - Dynamic sub-hypergraphs [3, 8]

Figure 2: The brief outline of the proposed tutorial.

especially those interested in graphs, when considering the representational power, usability, and omnipresence of hypergraphs. Moreover, patterns and generative models of hypergraph data will have a huge impact on our understanding of complex systems and also on various applications, including algorithm design, simulation, and anomaly detection, as those of graph data do. This tutorial aims to provide a starting point for further studies on this topic.

Relevance to CIKM: This tutorial on novel and interdisciplinary directions covers various aspects of data mining, including findings, algorithms, and applications. It should be noticed that more than half of the studies covered in the tutorial appeared in conferences in data mining (specifically, ICDM, KDD, WWW, VLDB, and SDM).

3 TARGET AUDIENCE AND PREREQUISITES

This tutorial is targeted at anyone interested in graph mining, graph learning, social network analysis, or network science, from researchers to the practitioners from industry. It should be noticed that hypergraphs have been used for modeling data from a variety of domains, including recommender systems [22, 30], computer vision [17, 27], natural language processing [9, 12], social network analysis [28], and thus they are of interest to practitioners.

Basic knowledge of linear algebra and probability theory will be helpful. For the audience new to this field, we will cover all necessary preliminaries and provide an intuitive overview of recent studies on the topic. We will also offer in-depth descriptions of advanced techniques for the audience with more experience in this field. Specifically, the audience of this tutorial will be able to (1) understand the basic hypergraph-related concepts, (2) use the concepts to model and analyze group interactions in various real-world complex systems, and (3) understand structural design principles of real-world hypergraphs.

4 OUTLINE AND CONTENTS

We provide a brief outline of the tutorial in Figure 2.

In this tutorial, we focus on providing a comprehensive overview of structural patterns discovered in real-world hypergraphs, and advanced data mining tools for large-scale hypergraphs. As an introduction, we present how hypergraphs are used to model various types of data from different domains.

During the first half of this tutorial, we introduce structural patterns pervasive in real-world hypergraphs. Specifically, we cover (a) static structural patterns[1] [2, 10, 18–20, 23] and (b) dynamic structural patterns[2] [2, 3, 7, 8, 18, 21] of real-world hypergraphs where the static patterns are further divided into (a) node-level patterns, (b) hyperedge-level patterns, (c) sub-hypergraph-level patterns, and (d) hypergraph-level patterns, as summarized in Table 1. The presented patterns include macroscopic (i.e., global) [10, 18, 19] and microscopic (i.e., local) [2, 20, 21, 23] patterns, and they also include patterns regarding connectivity [10, 18, 23], overlap [2, 8, 19–21], and repetition [3, 4, 7] of hyperedges. Together with the patterns, we present advanced data mining tools (e.g., hypergraph motifs [20, 21], multi-level decomposition [10], and a principled measure of "overlapness" [19]) developed for their discovery.

During the second half, we present generative models of hypergraphs, which are based on the observations made in real-world hypergraphs. They aim to reproduce and thus explain the structural patterns through intuitive mechanisms on individual nodes or hyperedges. These models can also be used for creating large-scale benchmark datasets, for anonymizing hypergraphs with sensitive information, and for comparing hypergraphs of different sizes. As categorized in Table 2, we cover four models for generating entire

[1] Patterns from static hypergraphs or a few snapshots.
[2] Patterns related to the dynamics of evolving hypergraphs.

Table 1: Categorization of structural properties in real-world hypergraphs that are covered in this tutorial.

	Static patterns* (Part II)				Dynamic patterns** (Part III)			
	Nodes	Hyperedges	Sub-hypergraphs	Hypergraphs	Nodes	Hyperedges	Sub-hypergraphs	Hypergraphs
Benson et al. (PNAS'18) [2]			✓	✓			✓	
Benson et al. (KDD'18) [3]					✓	✓		
Cencetti et al. (SciRep'21) [4]						✓		
Choo and Shin (SDM'22) [7]					✓			
Comrie and Kleinberg (ICDM'21) [8]							✓	
Do et al. (KDD'20) [10]	✓			✓				
Kook et al. (ICDM'20) [18]	✓	✓		✓				✓
Lee et al. (WWW'21) [19]	✓	✓	✓					
Lee et al. (VLDB'20) [20]			✓					
Lee and Shin (ICDM'21) [21]						✓	✓	
Lotito et al (CommsPhys'22) [23]			✓					

*Patterns from static hypergraphs or a few snapshots. **Patterns related to the dynamics of evolving hypergraphs.

Table 2: Comparison of the hypergraph generative models covered in this tutorial.

	Full hypergraphs		Sub-hypergraphs	
	Static	Dynamic	Static	Dynamic
Benson et al. (KDD'18) [3]				✓
Comrie and Kleinberg (ICDM'21) [8]				✓
Chodrow (J. Complex Netw'20) [5]	✓			
Choe et al (WWW'22) [6]			✓	
Do et al. (KDD'20) [10]		✓		
Kook et al. (ICDM'20) [18]		✓		
Lee et al. (WWW'21) [19]	✓			

hypergraphs models [5, 10, 18, 19] and two models for generating sub-hypergraphs [3, 6, 8]. In addition to their technical details, we present how realistic they are in various aspects.

5 IMPORTANT REFERENCES

The important references covered in the tutorial are provided below.

- [PNAS'18] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. *Simplicial closure and higher-order link prediction.* PNAS, 115(48):E11221–E11230, 2018.
- [KDD'18] Austin R Benson, Ravi Kumar, and Andrew Tomkins. *Sequences of sets.* In KDD, 2018
- [SciRep'21] Giulia Cencetti, Federico Battiston, Bruno Lepri, and Márton Karsai. *Temporal properties of higher-order interactions in social networks.* Scientific reports, 11(1):1–10, 2021.
- [J. Complex Netw'20] Philip S Chodrow. *Configuration models of random hypergraphs.* J. Complex Networks, 8(3):cnaa018, 2020.
- [WWW'22] Minyoung Choe, Jaemin Yoo, Geon Lee, Woonsung Baek, U Kang, and Kijung Shin. *Midas: Representative sampling from real-world hypergraphs.* In WWW, 2022.
- [SDM'22] Hyunjin Choo and Kijung Shin. *On the persistence of higher-order interactions in real-world hypergraphs.* SDM, 2022.
- [ICDM'21a] Cazamere Comrie and Jon Kleinberg. *Hypergraph ego-networks and their temporal evolution.* In ICDM, 2021.
- [KDD'20] Manh Tuan Do, Se-eun Yoon, Bryan Hooi, and Kijung Shin. *Structural patterns and generative models of real-world hypergraphs.* In KDD, 2020.
- [ICDM'20] Yunbum Kook, Jihoon Ko, and Kijung Shin. *Evolution of real-world hypergraphs: Patterns and models without oracles.* In ICDM, 2020.
- [WWW'21] Geon Lee, Minyoung Choe, and Kijung Shin. *How do hyperedges overlap in real-world hypergraphs?-patterns, measures, and generators.* In WWW, 2021.
- [VLDB'20] Geon Lee, Jihoon Ko, and Kijung Shin. *Hypergraph motifs: concepts, algorithms, and discoveries.* PVLDB, 13(12):2256–2269, 2020.
- [ICDM'21b] Geon Lee and Kijung Shin. *THyMe+: Temporal hypergraph motifs and fast algorithms for exact counting.* In ICDM, 2021.
- [CommPhys'22] Quintino Francesco Lotito, Federico Musciotto, Alberto Montresor, and Federico Battiston. *Higher-order motif analysis in hypergraphs.* Comm. Phys, 5(1):1–8, 2022.
- [SIREV'21] Leo Torres, Ann S. Blevins, Danielle Bassett, and Tina Eliassi-Rad. *The why, how, and when of representations for complex systems.* SIAM Review 63(3):435-485, 2021

6 RELEVANT TUTORIALS

There have been tutorials on mining of graphs in general, including:

- *Graph Structures in Data Mining* in KDD 2004
  - http://www.cs.cmu.edu/~christos/TALKS/KDD04-tut/
  - This tutorial focuses on (1) topological properties of nodes and edges, (2) importance measures of nodes, and (3) similarity and influence between nodes in graphs.
- *Large Graph Mining: Patterns, Tools, and Case Studies* in CIKM 2008 & ICDE 2009
  - http://tonghanghang.org/pdfs/tut-icde09-part1_patterns.pdf
  - This tutorial focuses on (1) structural patterns, (2) matrix & tensor tools, (3) proximity measures between nodes, and (4) case studies of real-world graphs.
- *Mining Billion-Scale Graphs: Patterns and Algorithms* in SIGMOD 2012
  - https://www.cs.cmu.edu/~christos/TALKS/12-SIGMOD-tutorial/
  - This tutorial focuses on (1) patterns in real-world graphs, (2) tools for pattern mining in graphs, and (3) scalable algorithms for large-scale graphs.
- *Advanced Graph Mining for Community Evaluation in Social Networks and the Web* in WSDM 2013
  - http://www.lix.polytechnique.fr/~mvazirg/WSDM2013_tutorial
  - This tutorial focuses on detection and evaluation methods of communities in graphs.
- *Big Graph Mining: Algorithms, Anomaly Detection, and Applications* in ASONAM 2013

  - https://datalab.snu.ac.kr/~ukang/talks/13-ASONAM-tutorial/
  - This tutorial focuses on (1) scalable graph mining, (2) graph-based anomaly detection, and (3) applications.
- *Core Decomposition of Networks: Concepts, Algorithms and Applications* in ICDM 2016 & PKDD 2017
  - https://fragkiskos.me/projects/core_tutorial/
  - This tutorial focuses on (a) the concept and properties of core decomposition, (b) efficient computation, and (c) applications.
- *Roles in Networks - Foundations, Methods and Applications* in ICDM 2021
  - https://cswzhang.github.io/icdm-tutorial-2021/
  - This tutorial focuses on (a) the taxonomy of role analytic methods, (b) role-based embedding methods, (c) and applications.

To the best of our knowledge, however, no tutorial that focuses on **hypergraphs** has been offered in data-mining and related venues. The patterns, tools and models covered in this tutorial are clearly distinguished from those for ordinary graphs. We plan to deliver the same tutorial at DSAA 2022 and ICDM 2022.

7 SHORT BIO OF PRESENTERS

Geon Lee (https://geonlee0325.github.io) is a Ph.D. student at the Kim Jaechul Graduate School of AI at KAIST. He received his B.S. degree in Computer Science and Engineering from Sungkyunkwan University in 2019. His research interests include graph mining and its applications. Especially, his studies of hypergraphs have appeared in major data mining venues, including VLDB, WWW, and ICDM.

Jaemin Yoo (https://jaeminyoo.github.io) is a postdoctoral research fellow in the Heinz College of Information Systems and Public Policy at Carnegie Mellon University. He received his Ph.D. and B.S. in Computer Science and Engineering from Seoul National University. His research interests include probabilistic mining and machine learning on graphs. His work has been published in major venues including WWW, KDD, and NeurIPS. He is a recipient of the Google PhD Fellowship and the Qualcomm Innovation Fellowship.

Kijung Shin (https://kijungs.github.io/) is an Ewon Endowed Assistant Professor (jointly affiliated) in the Kim Jaechul Graduate School of AI and the School of Electrical Engineering at KAIST. He received his Ph.D. in Computer Science from Carnegie Mellon University in 2019. He has published more than 50 referred articles at major data mining venues, and he won the best research paper award at KDD 2016. His research interests span a wide range of topics on graph mining, with a focus on scalable algorithm design and empirical analysis of real-world hypergraphs.

ACKNOWLEDGEMENTS

This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1C1C1008296) and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00157, Robust, Fair, Extensible Data-Centric Continual Learning).

REFERENCES

[1] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.
[2] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. 2018. Simplicial closure and higher-order link prediction. *PNAS* 115, 48 (2018), E11221–E11230.
[3] Austin R Benson, Ravi Kumar, and Andrew Tomkins. 2018. Sequences of sets. In *KDD*.
[4] Giulia Cencetti, Federico Battiston, Bruno Lepri, and Márton Karsai. 2021. Temporal properties of higher-order interactions in social networks. *Scientific reports* 11, 1 (2021), 1–10.
[5] Philip S Chodrow. 2020. Configuration models of random hypergraphs. *Journal of Complex Networks* 8, 3 (2020), cnaa018.
[6] Minyoung Choe, Jaemin Yoo, Geon Lee, Woonsung Baek, U Kang, and Kijung Shin. 2022. Midas: Representative sampling from real-world hypergraphs. In *WWW*.
[7] Hyunjin Choo and Kijung Shin. 2022. On the Persistence of Higher-Order Interactions in Real-World Hypergraphs. *SDM* (2022).
[8] Cazamere Comrie and Jon Kleinberg. 2021. Hypergraph Ego-networks and Their Temporal Evolution. In *ICDM*.
[9] Kaize Ding, Jianling Wang, Jundong Li, Dingcheng Li, and Huan Liu. 2020. Be More with Less: Hypergraph Attention Networks for Inductive Text Classification. In *EMNLP*.
[10] Manh Tuan Do, Se-eun Yoon, Bryan Hooi, and Kijung Shin. 2020. Structural patterns and generative models of real-world hypergraphs. In *KDD*.
[11] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. 1999. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review* 29, 4 (1999), 251–262.
[12] Mikhail Galkin, Priyansh Trivedi, Gaurav Maheshwari, Ricardo Usbeck, and Jens Lehmann. 2020. Message Passing for Hyper-Relational Knowledge Graphs. In *EMNLP*.
[13] Carsten Grabow, Stefan Grosskinsky, Jürgen Kurths, and Marc Timme. 2015. Collective relaxation dynamics of small-world networks. *Physical Review E* 91, 5 (2015), 052815.
[14] TaeHyun Hwang, Ze Tian, Rui Kuangy, and Jean-Pierre Kocher. 2008. Learning on weighted hypergraphs to integrate protein interactions and gene expressions for cancer outcome prediction. In *ICDM*.
[15] U Kang, Charalampos E Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec. 2010. Radius plots for mining tera-byte scale graphs: Algorithms, patterns, and observations. In *SDM*.
[16] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. 1999. Multilevel hypergraph partitioning: Applications in VLSI domain. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 7, 1 (1999), 69–79.
[17] Eun-Sol Kim, Woo Young Kang, Kyoung-Woon On, Yu-Jung Heo, and Byoung-Tak Zhang. 2020. Hypergraph attention networks for multimodal learning. In *CVPR*.
[18] Yunbum Kook, Jihoon Ko, and Kijung Shin. 2020. Evolution of real-world hypergraphs: Patterns and models without oracles. In *ICDM*.
[19] Geon Lee, Minyoung Choe, and Kijung Shin. 2021. How Do Hyperedges Overlap in Real-World Hypergraphs?-Patterns, Measures, and Generators. In *WWW*.
[20] Geon Lee, Jihoon Ko, and Kijung Shin. 2020. Hypergraph motifs: concepts, algorithms, and discoveries. *PVLDB* 13, 12 (2020), 2256–2269.
[21] Geon Lee and Kijung Shin. 2021. THyMe+: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting. In *ICDM*.
[22] Yicong Li, Hongxu Chen, Xiangguo Sun, Zhenchao Sun, Lin Li, Lizhen Cui, Philip S Yu, and Guandong Xu. 2021. Hyperbolic hypergraphs for sequential recommendation. In *CIKM*.
[23] Quintino Francesco Lotito, Federico Musciotto, Alberto Montresor, and Federico Battiston. 2022. Higher-order motif analysis in hypergraphs. *Communications Physics* 5, 1 (2022), 1–8.
[24] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. 2004. Superfamilies of evolved and designed networks. *Science* 303, 5663 (2004), 1538–1542.
[25] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. 2002. Network motifs: simple building blocks of complex networks. *Science* 298, 5594 (2002), 824–827.
[26] Leo Torres, Ann S Blevins, Danielle Bassett, and Tina Eliassi-Rad. 2021. The why, how, and when of representations for complex systems. *SIAM Rev.* 63, 3 (2021), 435–485.
[27] Xiangping Wu, Qingcai Chen, Wei Li, Yulun Xiao, and Baotian Hu. 2020. AdaHGNN: Adaptive Hypergraph Neural Networks for Multi-Label Image Classification. In *MM*.
[28] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. 2019. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *WWW*.
[29] Se-eun Yoon, Hyungseok Song, Kijung Shin, and Yung Yi. 2020. How much and when do we need higher-order information in hypergraphs? a case study on hyperedge prediction. In *WWW*.
[30] Junwei Zhang, Min Gao, Junliang Yu, Lei Guo, Jundong Li, and Hongzhi Yin. 2021. Double-scale self-supervised hypergraph learning for group recommendation. In *CIKM*.