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. Materials and details of this tutorial can also be found at **https://sites.google.com/view/hypergraph-tutorial**.

CCS CONCEPTS

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

KEYWORDS

Hypergraph, Higher-order Interaction, Pattern, Generator

1 MOTIVATION

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. 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 [34], natural language processing [11], social network analysis [32], and circuit designs [16].

Motivated by the successful investigation of structural patterns in real-world pairwise graphs (e.g., power-law degree distribution [1, 14] and network motifs [25, 26]) 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.

2 OUTLINE

We provide a comprehensive overview of structural patterns discovered in real-world hypergraphs, advanced data mining tools, and hypergraph generative models based on the patterns.

- **Part I: Introduction**
  - Group interactions in the real-world
  - Power of hypergraph modeling [30, 33]
  - Data repositories and open-source software
- **Part II: Static Structural Patterns**
  - Basic patterns
    - Node-level properties [5, 12, 19, 20]
    - Hyperedge-level properties [19, 20, 27]
    - Hypergraph-level properties [3, 12, 19, 29]
  - Advanced patterns (sub-hypergraph-level properties) [3, 15, 17, 20, 21, 24]
- **Part III: Dynamic Structural Patterns in Hypergraphs**
  - Basic patterns
    - Node-level properties [4, 9]
    - Hyperedge-level properties [4, 6, 22]
    - Hypergraph-level properties [19]
  - Advanced patterns (sub-hypergraph-level properties) [3, 10, 22, 23]
- **Part IV: Generative Models of Hypergraphs**
  - Full-hypergraph generation
    - Static hypergraphs [7, 20, 28, 29]
    - Dynamic hypergraphs [2, 12, 13, 17–19, 31]
  - Sub-hypergraph generation
    - Static sub-hypergraphs [8]
    - Dynamic sub-hypergraphs [4, 10]

Note that this tutorial is an extension of the tutorials given by the same presenters at ACM CIKM 2022, IEEE ICDM 2022, and WebConf 2023. This updated version of the tutorial provides more comprehensive coverage, incorporating recent advances in the topic.

3 POTENTIAL IMPACTS

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

4 PRESENTERS

Geon Lee 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 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 is an Associate 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 70 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.

ACKNOWLEDGMENTS

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. 2019-0-00075, Artificial Intelligence Graduate School Program (KAIST)).

REFERENCES

[1] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.

[2] Natalie C Behague, Anthony Bonato, Melissa A Huggan, Rehan Malik, and Trent G Marbach. 2021. The iterated local transitivity model for hypergraphs. *arXiv preprint arXiv:2101.12560* (2021).

[3] 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.

[4] Austin R Benson, Ravi Kumar, and Andrew Tomkins. 2018. Sequences of sets. In *KDD*.

[5] Fanchen Bu, Geon Lee, and Kijung Shin. 2023. Hypercore Decomposition for Non-Fragile Hyperedges: Concepts, Algorithms, Observations, and Applications. *arXiv preprint arXiv:2301.08440* (2023).

[6] 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.

[7] Philip S Chodrow. 2020. Configuration models of random hypergraphs. *Journal of Complex Networks* 8, 3 (2020), cnaa018.

[8] Minyoung Choe, Jaemin Yoo, Geon Lee, Woonsung Baek, U Kang, and Kijung Shin. 2022. Midas: Representative sampling from real-world hypergraphs. In *WWW*.

[9] Hyunjin Choo and Kijung Shin. 2022. On the Persistence of Higher-Order Interactions in Real-World Hypergraphs. *SDM* (2022).

[10] Cazamere Comrie and Jon Kleinberg. 2021. Hypergraph Ego-networks and Their Temporal Evolution. In *ICDM*.

[11] 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*.

[12] Manh Tuan Do, Se-eun Yoon, Bryan Hooi, and Kijung Shin. 2020. Structural patterns and generative models of real-world hypergraphs. In *KDD*.

[13] Nicole Eikmeier, Arjun S Ramani, and David Gleich. 2018. The hyperkron graph model for higher-order features. In *ICDM*.

[14] 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.

[15] Jonas L Juul, Austin R Benson, and Jon Kleinberg. 2022. Hypergraph patterns and collaboration structure. *arXiv preprint arXiv:2210.02163* (2022).

[16] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. 1999. Multilevel hypergraph partitioning: Applications in VLSI domain. *VLSI* 7, 1 (1999), 69–79.

[17] Sunwoo Kim, Fanchen Bu, Minyoung Choe, Jaemin Yoo, and Kijung Shin. 2023. How Transitive Are Real-World Group Interactions? - Measurement and Reproduction. In *KDD*.

[18] Jihoon Ko, Yunbum Kook, and Kijung Shin. 2022. Growth patterns and models of real-world hypergraphs. *Knowledge and Information Systems* 64, 11 (2022), 2883–2920.

[19] Yunbum Kook, Jihoon Ko, and Kijung Shin. 2020. Evolution of real-world hypergraphs: Patterns and models without oracles. In *ICDM*.

[20] Geon Lee, Minyoung Choe, and Kijung Shin. 2021. How Do Hyperedges Overlap in Real-World Hypergraphs?-Patterns, Measures, and Generators. In *WWW*.

[21] Geon Lee, Jihoon Ko, and Kijung Shin. 2020. Hypergraph motifs: concepts, algorithms, and discoveries. *PVLDB* 13, 12 (2020), 2256–2269.

[22] Geon Lee and Kijung Shin. 2021. THyMe+: Temporal Hypergraph Motifs and Fast Algorithms for Exact Counting. In *ICDM*.

[23] Geon Lee and Kijung Shin. 2023. Temporal hypergraph motifs. *KAIS* (2023), 1–38.

[24] Quintino Francesco Lotito, Federico Musciotto, Alberto Montresor, and Federico Battiston. 2022. Higher-order motif analysis in hypergraphs. *Communications Physics* 5, 1 (2022), 1–8.

[25] 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.

[26] 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.

[27] Federico Musciotto, Federico Battiston, and Rosario N Mantegna. 2021. Detecting informative higher-order interactions in statistically validated hypergraphs. *Communications Physics* 4, 1 (2021), 218.

[28] Kazuki Nakajima, Kazuyuki Shudo, and Naoki Masuda. 2021. Randomizing hypergraphs preserving degree correlation and local clustering. *TNSE* 9, 3 (2021), 1139–1153.

[29] Marios Papachristou and Jon Kleinberg. 2022. Core-periphery Models for Hypergraphs. In *KDD*.

[30] 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.

[31] Alexei Vazquez. 2022. Growth principles of natural hypergraphs. *arXiv preprint arXiv:2208.03103* (2022).

[32] 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*.

[33] 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*.

[34] 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*.