MARIOH: Multiplicity-Aware Hypergraph Reconstruction

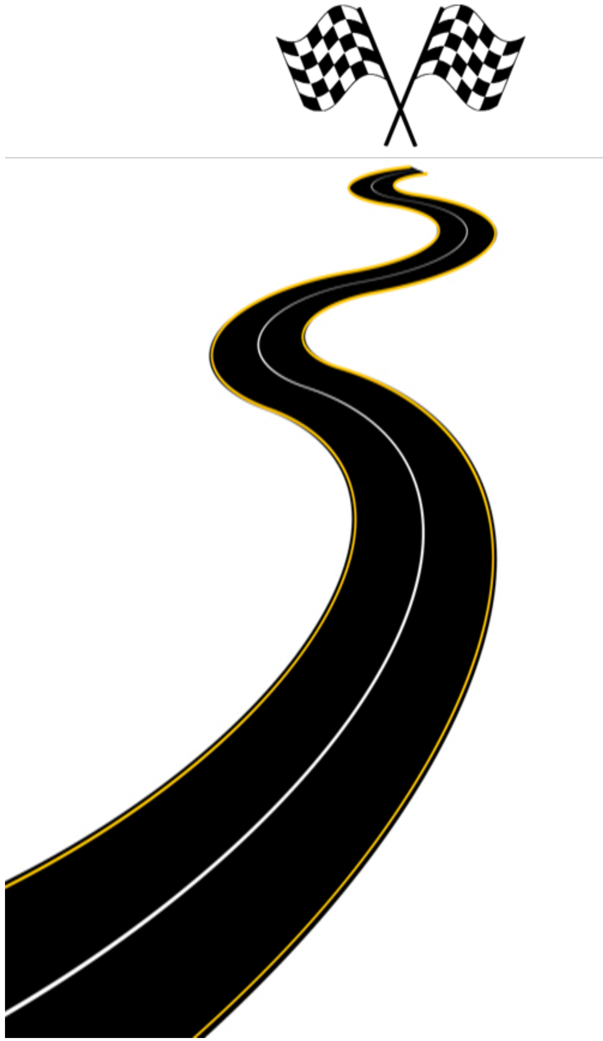


Kyuhan Lee[1,2] Geon Lee[1] Kijung Shin[1]

[1]Kim Jaechul Graduate School of AI, KAIST
[2]GraphAI

Road Map

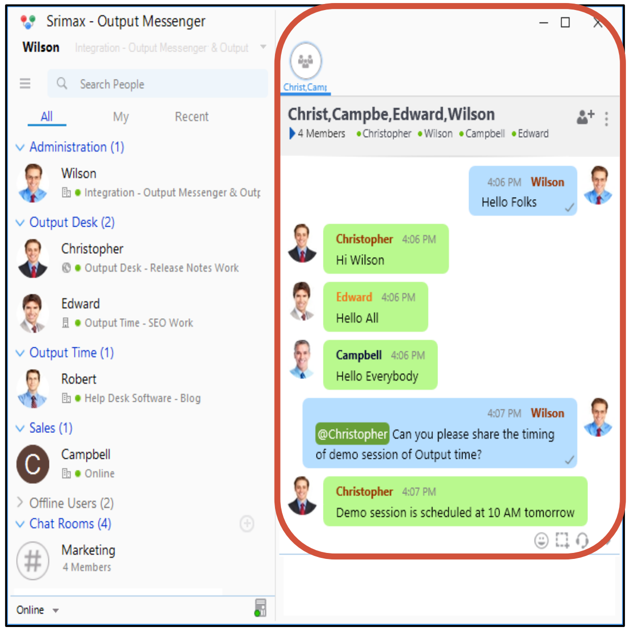
- Introduction <<
- Related Work
- Proposed Algorithm: MARIOH
- Experimental Results
- Conclusion

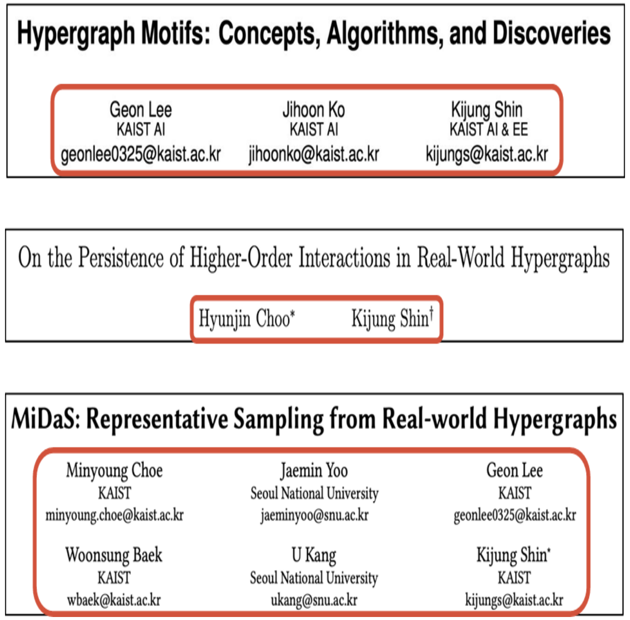


Motivation

Many relations in the real world involve more than two entities

- e.g., group chats, paper co-authorship, co-purchases





Hypergraph Motifs: Concepts, Algorithms, and Discoveries

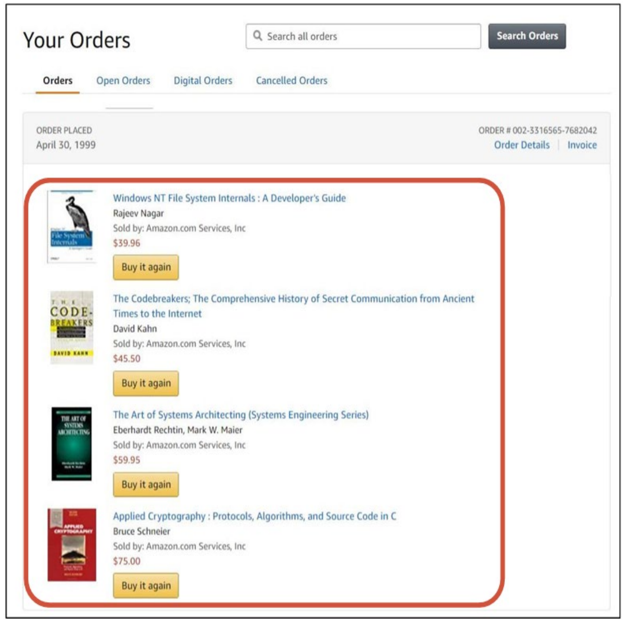
Geon Lee, KAIST AI, geonlee0325@kaist.ac.kr
Jihoon Ko, KAIST AI, jihoonko@kaist.ac.kr
Kijung Shin, KAIST AI & EE, kijungs@kaist.ac.kr

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



Motivation

These group relations are often recorded as pairwise links

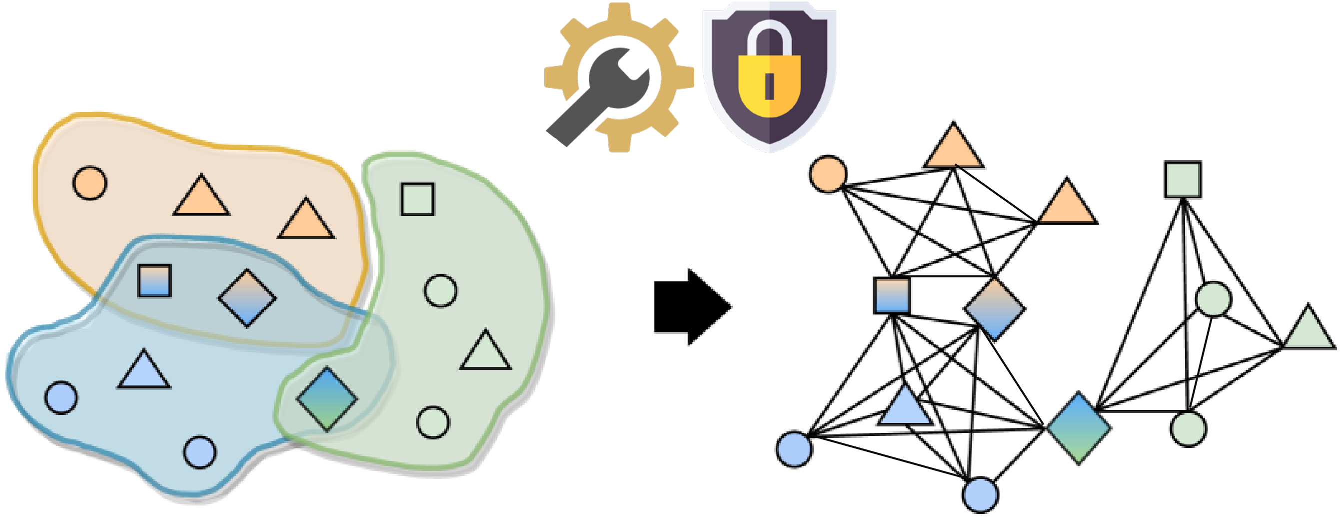
- Technical limitations: only pairwise interactions can be measured
- Privacy constraints: higher-order data is withheld by providers

They form cliques in the resulting pairwise graph

Biological Complexes

Social Networks

Brain Connectivity



Motivation

These group relations are often recorded as pairwise links

- Technical limitations: only pairwise interactions can be measured
- Privacy constraints: higher-order data is withheld by providers

They form cliques in the resulting pairwise graph

Biological Complexes

Social Networks

Brain Connectivity

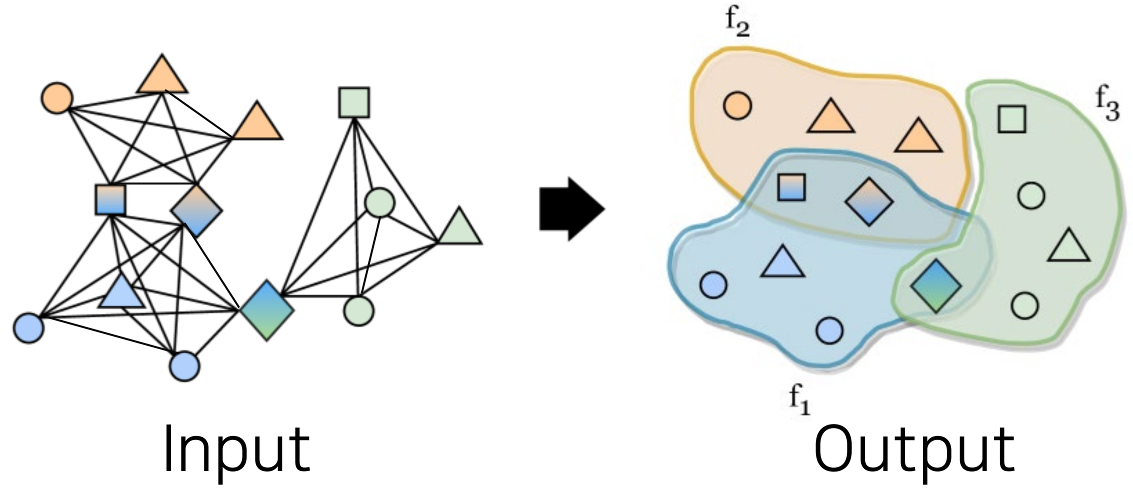
Goal: Reconstructing Higher-order Relations

To reconstruct higher-order (group) relations from graphs by identifying higher-order cliques

- Higher-order clique: A clique derived from a higher-order relation

Input & Output:

- Input: Graph G
- Output: Hypergraph H_G

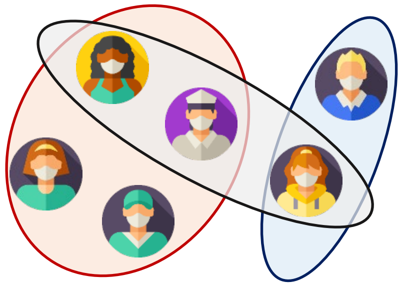


What is Hypergraph?

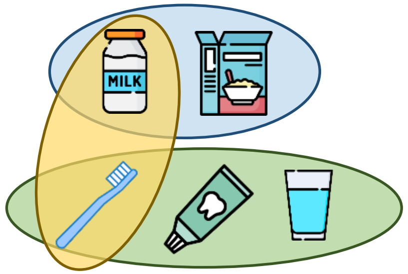
A generalization of a graph where an edge (hyperedge) connects more than two nodes

Hypergraph $H = (V, E)$

- V: set of vertices
- E: set of hyperedges (subsets of V)



Collaboration of Researchers



Co-purchases of Items

Advantages of Hypergraph Reconstruction

Intuitively interpretable

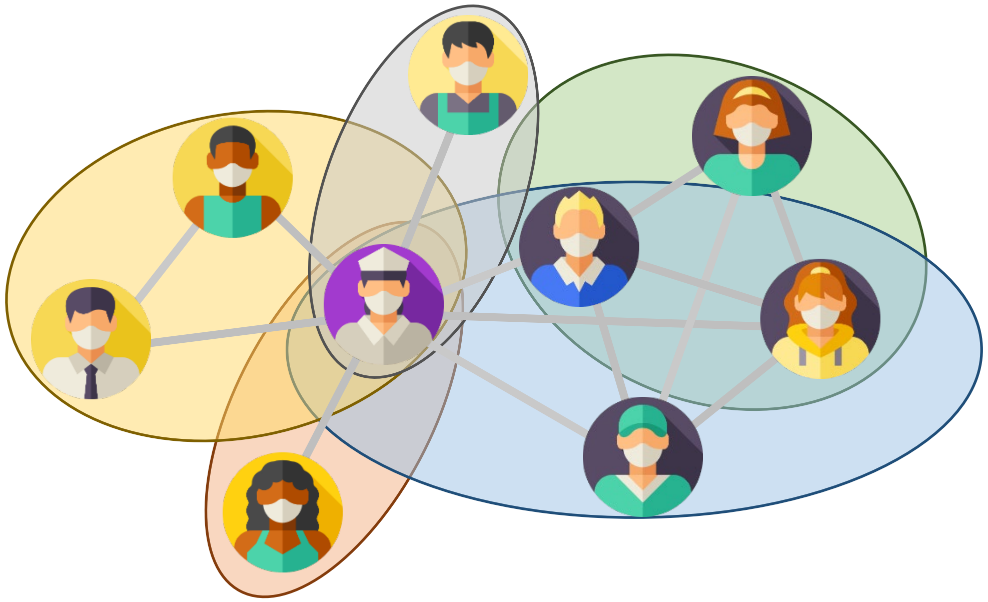
- Reconstructed output (i.e., hypergraph) is intuitively understandable
  - Each hyperedge corresponds to an actual higher-order relation
  - E.g., Each hyperedge in a co-author network corresponds to a paper (i.e., the collaboration of the co-authors of the paper)

Applicable

- Reconstructed output can be applied to hypergraph analysis tools
  - Hypergraph analysis tools often outperform graph counterparts [Tan et al., 2024, Aponte et al., 2022]

Our solution: Inferring Higher-Order Cliques

We infer higher-order cliques to identify higher-order relation from the cliques existing in the graph



From: Input graph G (with 14 edges)

To: Hypergraph H_G (with 14 nodes involved)

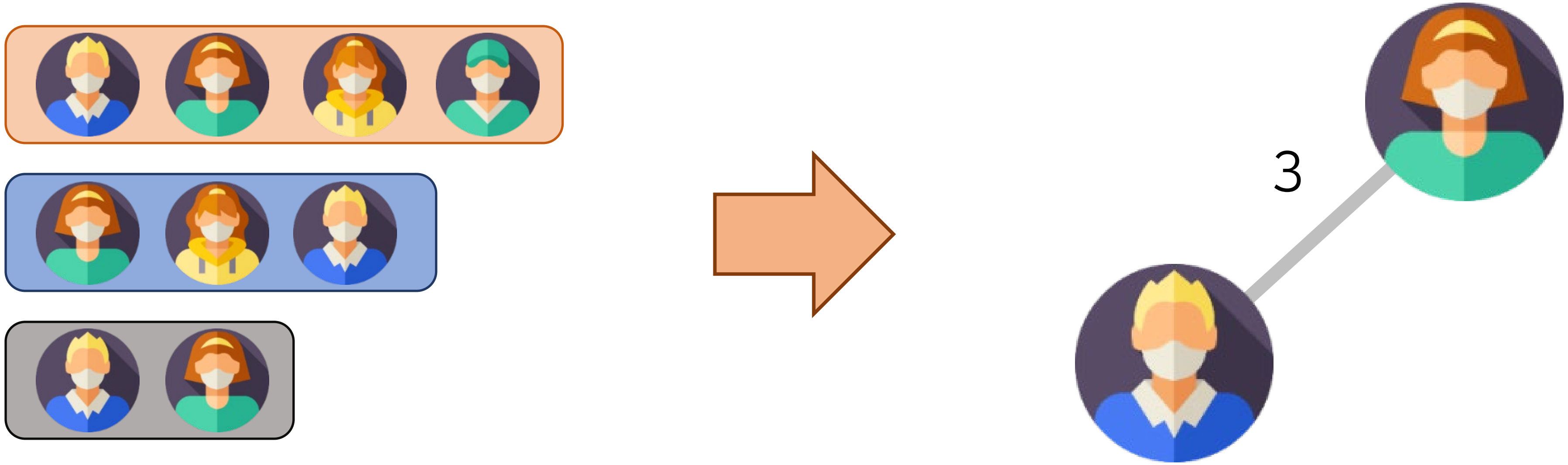
Challenges in Inferring Higher-Order Cliques

Inferring higher-order cliques also involves several challenges

- Large search space
  - Many cliques exist within the graph
- Ambiguity in hyperedge identification
  - Not all cliques are derived from higher-order group relations
  - Different higher-order group relations can result in same cliques

Edge Multiplicity in Higher-Order Relations

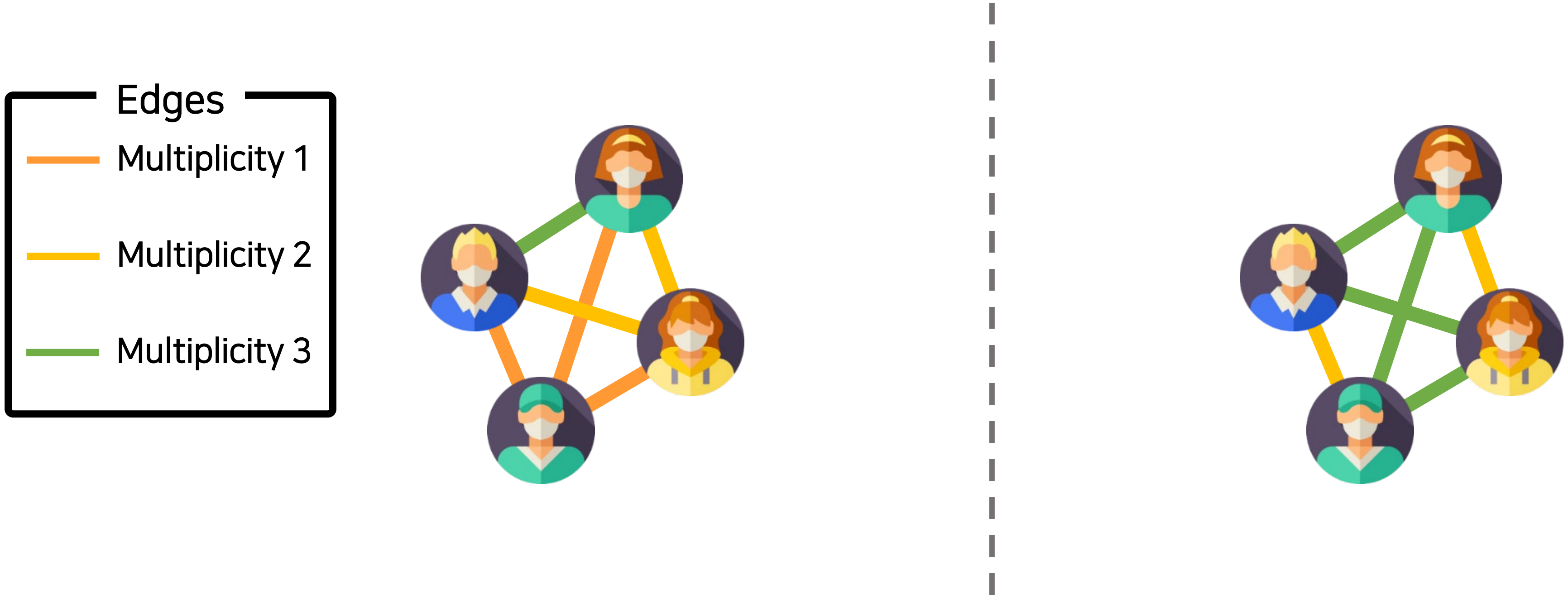
Captures higher-order relation frequency



Advantages of Leveraging Edge Multiplicity

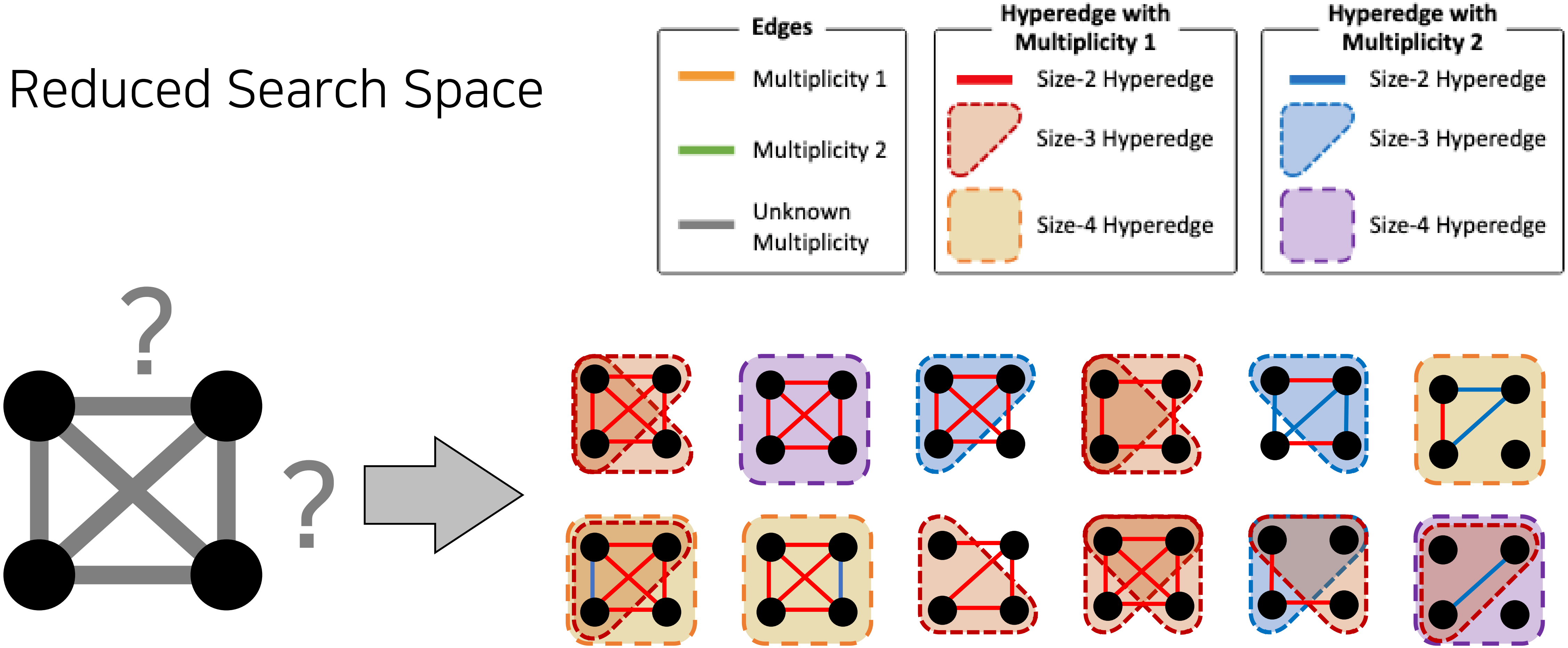
Clique feature enhancement

- Edge multiplicity within a clique provides distinguishing information, even for cliques of the same size



Advantages of Leveraging Edge Multiplicity

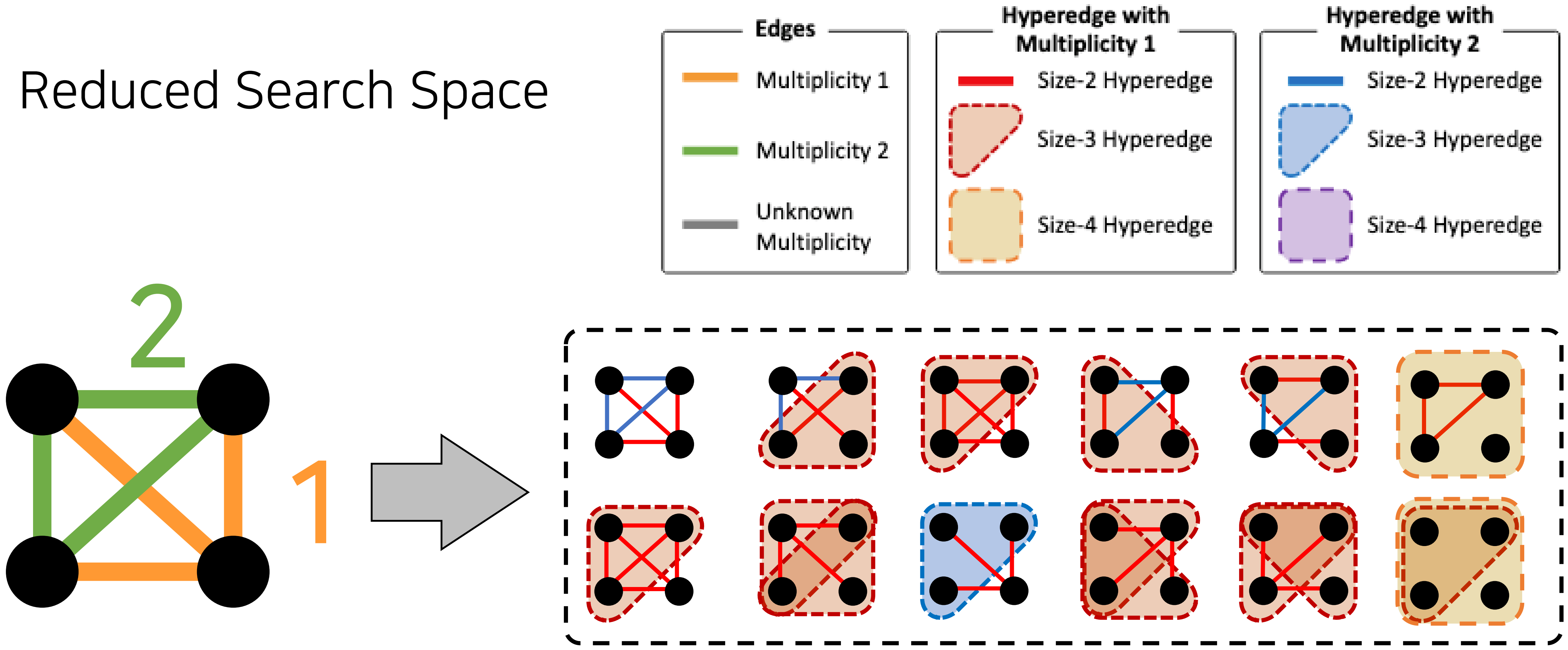
Reduced Search Space



+ Infinitely many cases

Advantages of Leveraging Edge Multiplicity

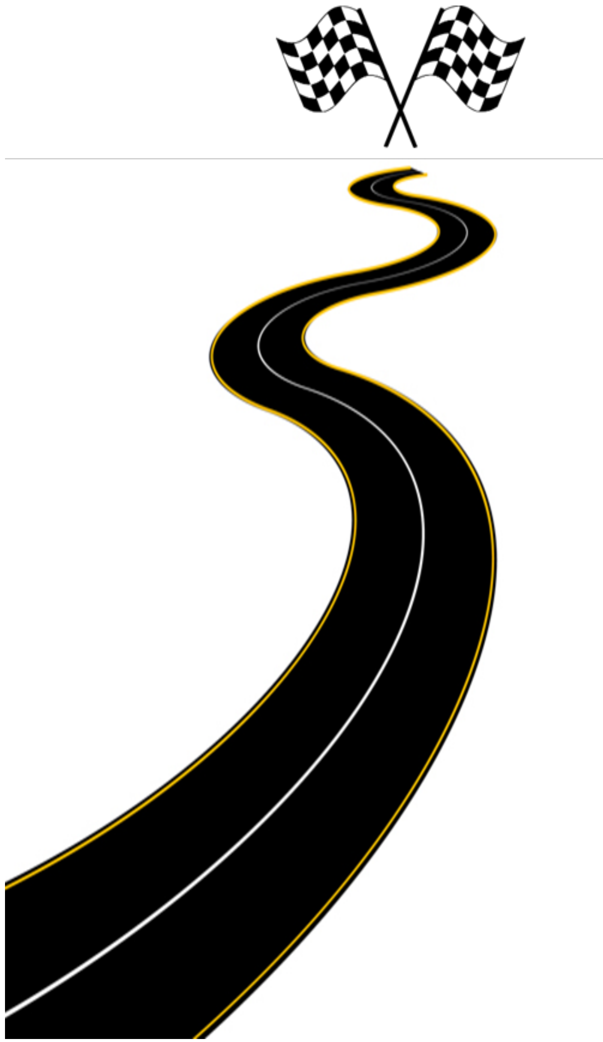
Reduced Search Space



Only 12 Potential Outputs

Road Map

- Introduction
- Related Work <<
- Proposed Algorithm: MARIOH
- Experimental Results
- Conclusion



Related Work: Multiplicity-Unaware Methods

Bayesian-MDL [Young et al., 2021]:

- Unsupervised approach
- Reconstructs hypergraphs by minimizing the description length of the observed graph

SHyRE [Wang and Kleinberg., 2024]:

- Supervised approach
- Samples cliques from the projected graph based on statistics
- Trains a classifier to predict whether the sampled cliques are true hyperedges

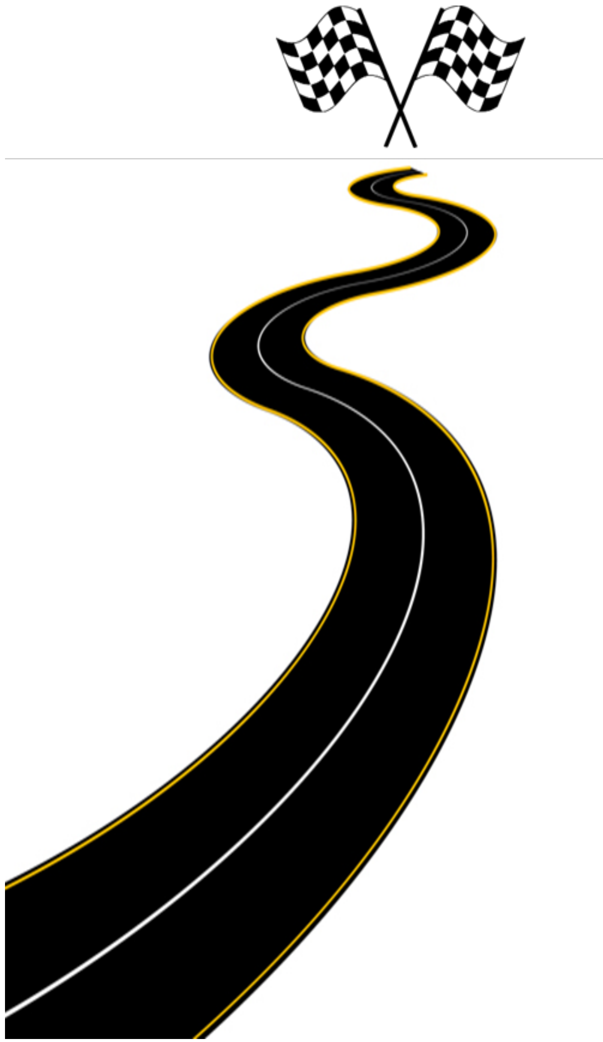
Related Work: Multiplicity-Aware Methods

SHyRE-Unsup [Wang and Kleinberg., 2024]:

- Unsupervised approach
- Iteratively selects and replaces maximal cliques as hyperedges, prioritizing cliques with higher structural importance
- (Limitation) High complexity due to replacing cliques one by one
- (Limitation) Ineffective utilization of edge multiplicity

Road Map

- Introduction
- Related Work
- Proposed Algorithm: MARIOH <<
- Experimental Results
- Conclusion



Proposed Method: MARIOH

MARIOH: Multiplicity-Aware Reconstruction of Hypergraphs

Main Ideas of MARIOH:

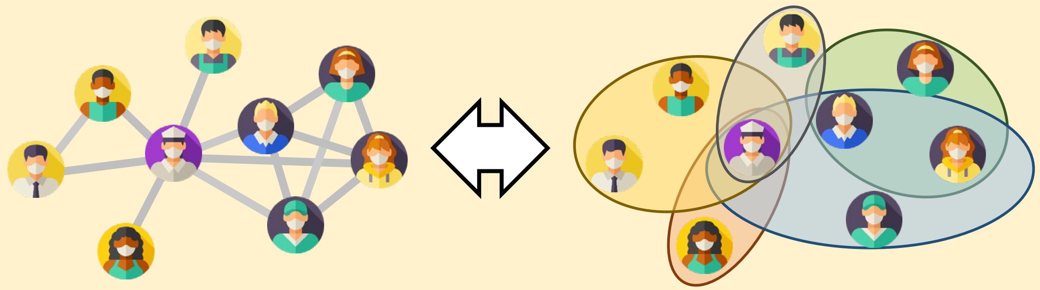
- **Multiplicity-Aware Features:** Leverages edge multiplicity to enhance reconstruction accuracy
- **Theoretically-Grounded Filtering:** Reduces the search space by identifying guaranteed size-2 hyperedges
- **Bidirectional Search:** Explores both highly promising cliques and overlooked sub-cliques to maximize recovery

Problem Formulation

Supervised Hypergraph Reconstruction

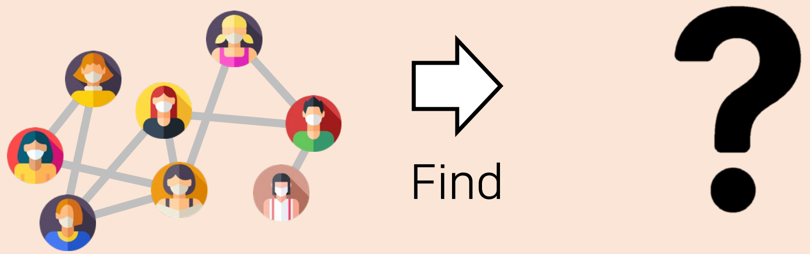
- **Given:**
  - Source Data: graph $\mathcal{G}^{(S)}$, hypergraph $\mathcal{H}^{(S)}$
  - Target Data: graph $\mathcal{G}^{(T)}$
- **To Reconstruct:** the hypergraph $\widehat{\mathcal{H}}^{(T)}$ for $\mathcal{G}^{(T)}$

Source: $\mathcal{G}^{(S)}, \mathcal{H}^{(S)}$



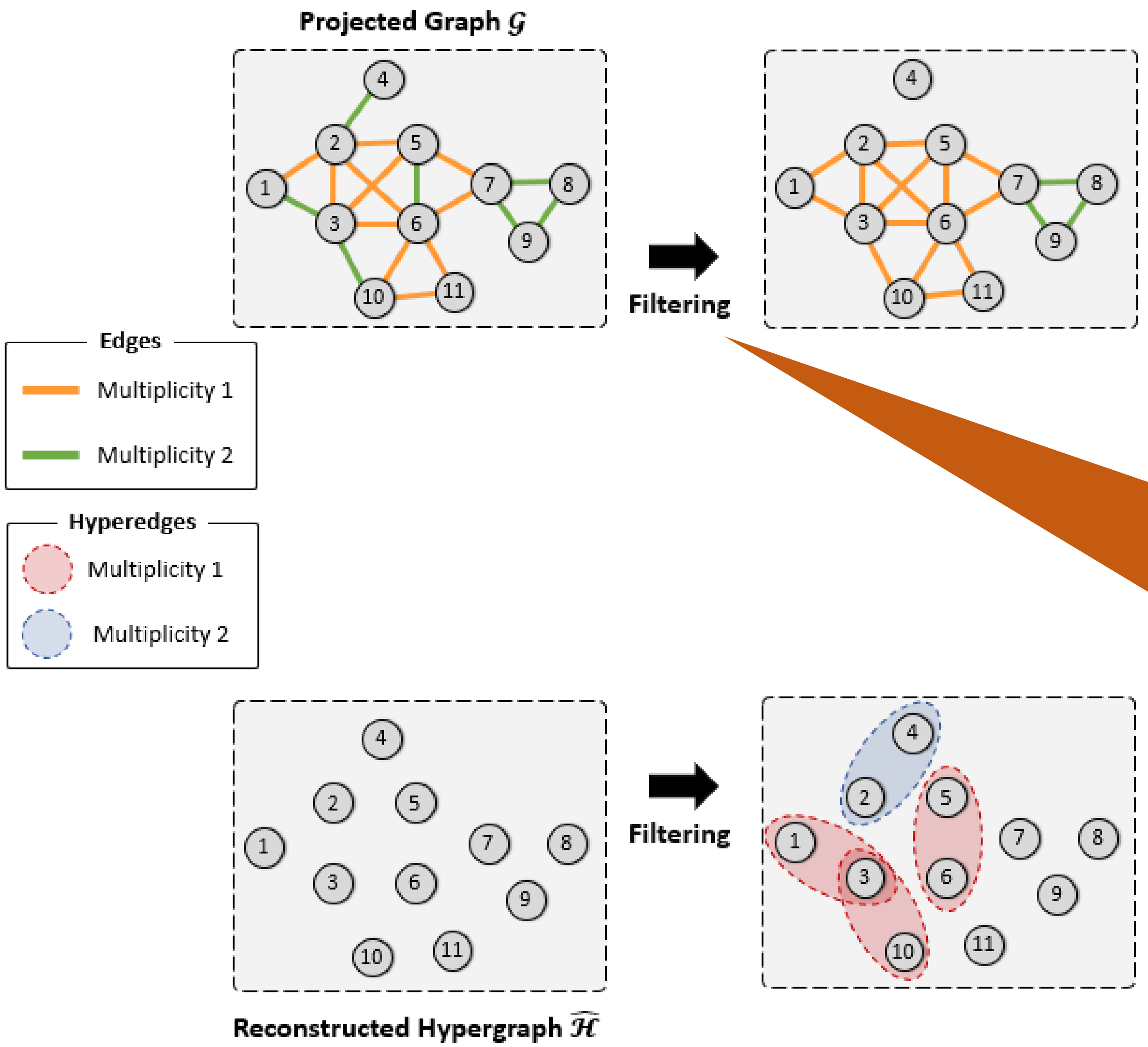
E.g.: Datasets from **Facebook Messenger**

Target: $\mathcal{G}^{(T)}$



E.g.: Datasets from **WhatsApp**

Example Procedure of MARIOH



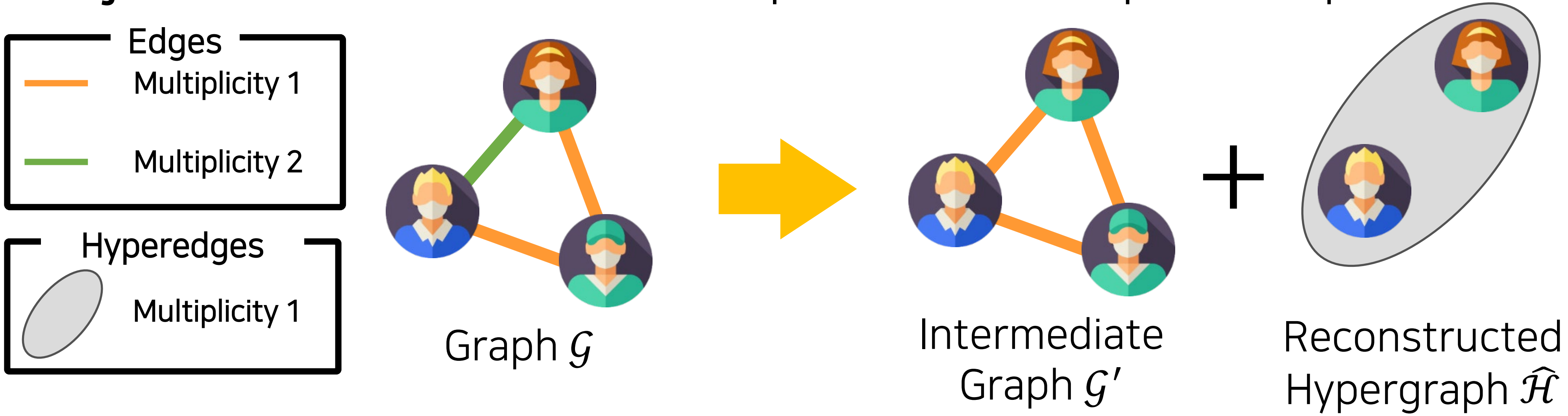
Objective: Identify edges in the projected graph ($\mathcal{G}$) that are theoretically guaranteed to correspond to size-2 hyperedges in the original hypergraph

Key Components of MARIOH

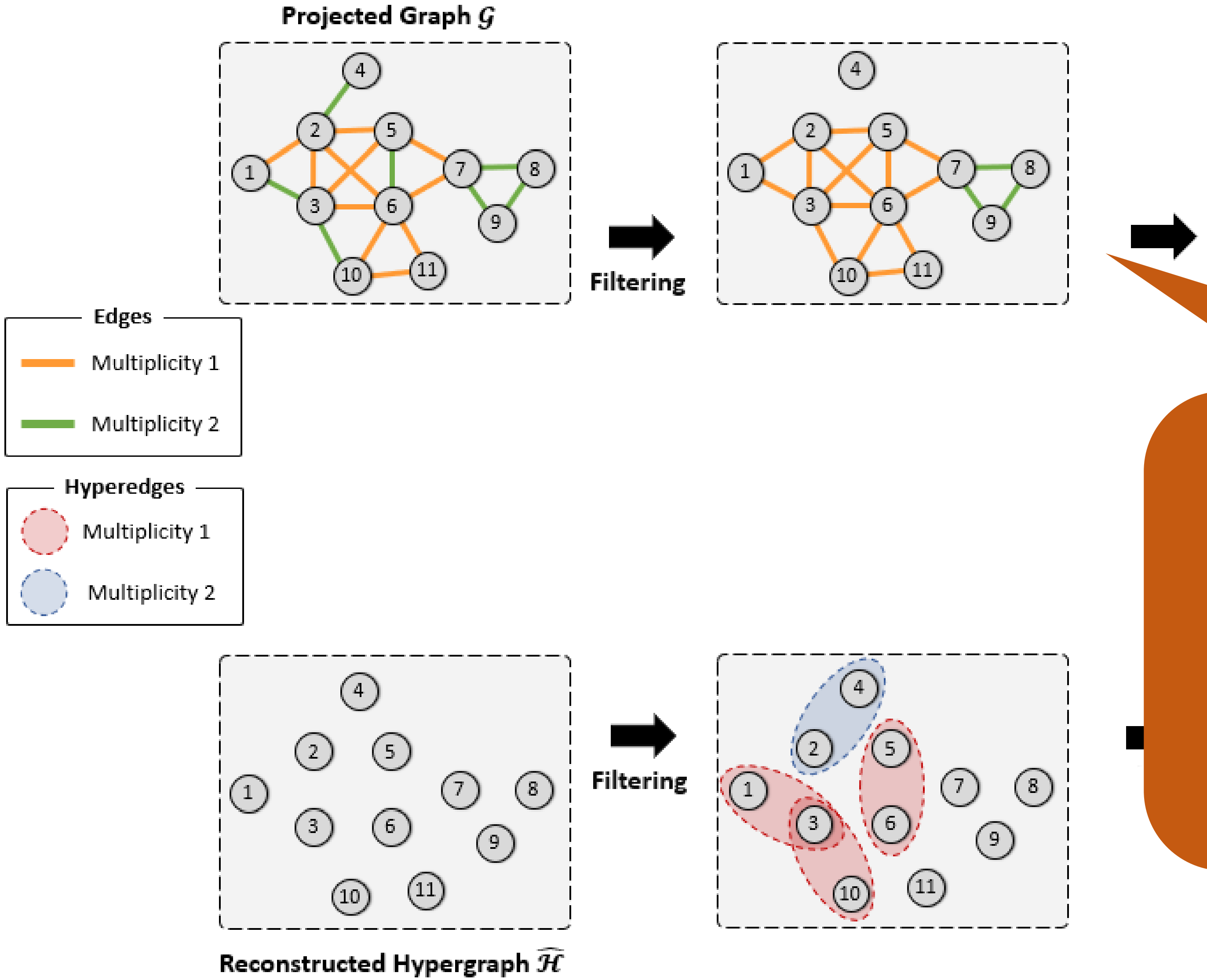
Theoretically-Guaranteed Filtering (Preprocessing Step)

- Identify theoretically-guaranteed size-2 hyperedges
- Remove them from the projected graph $\mathcal{G}$
- Add them to the reconstructed hypergraph ($\hat{\mathcal{H}}$)

- **Objective**: Reduces the search space for subsequent steps

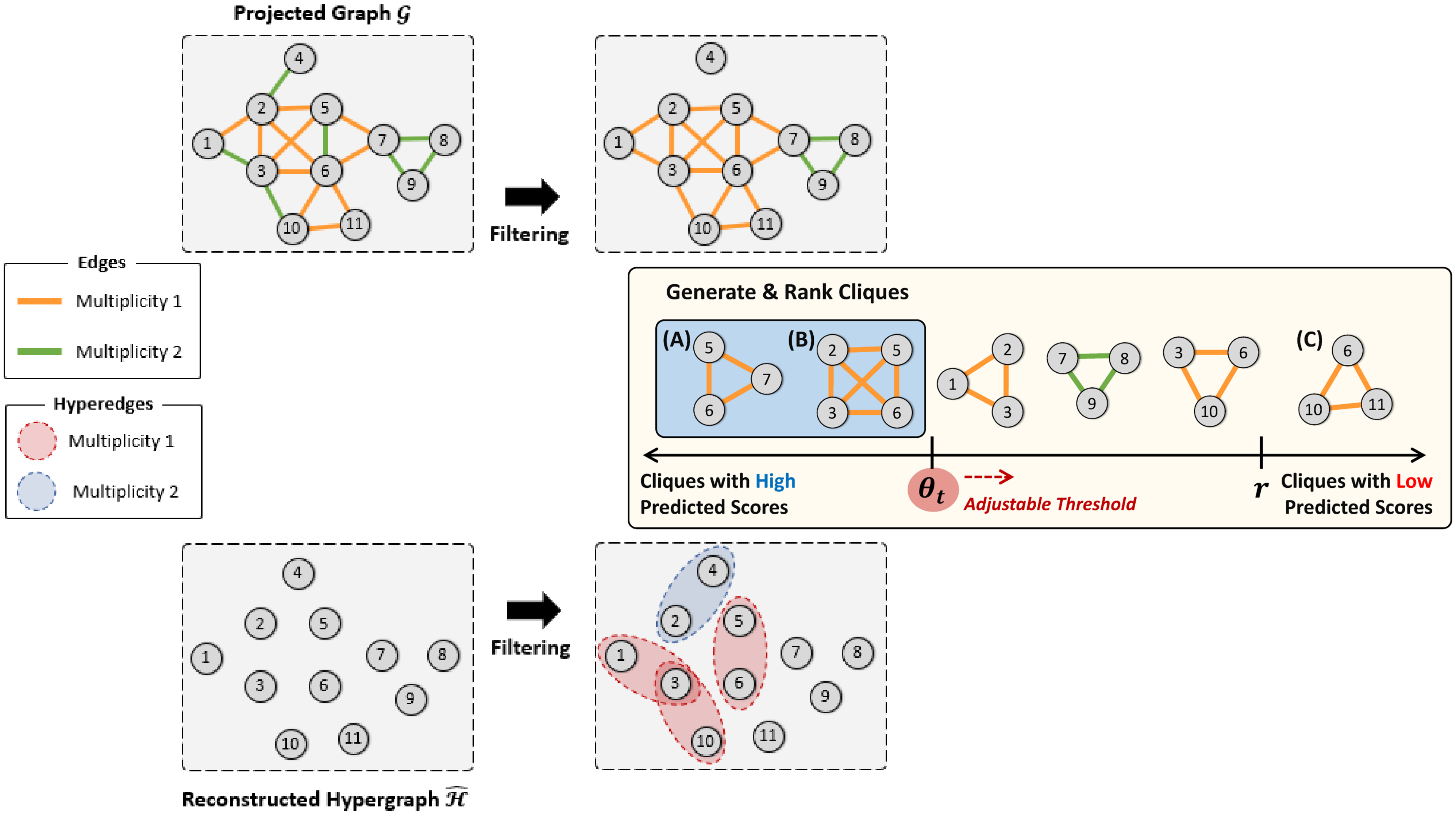


Example Procedure of MARIOH

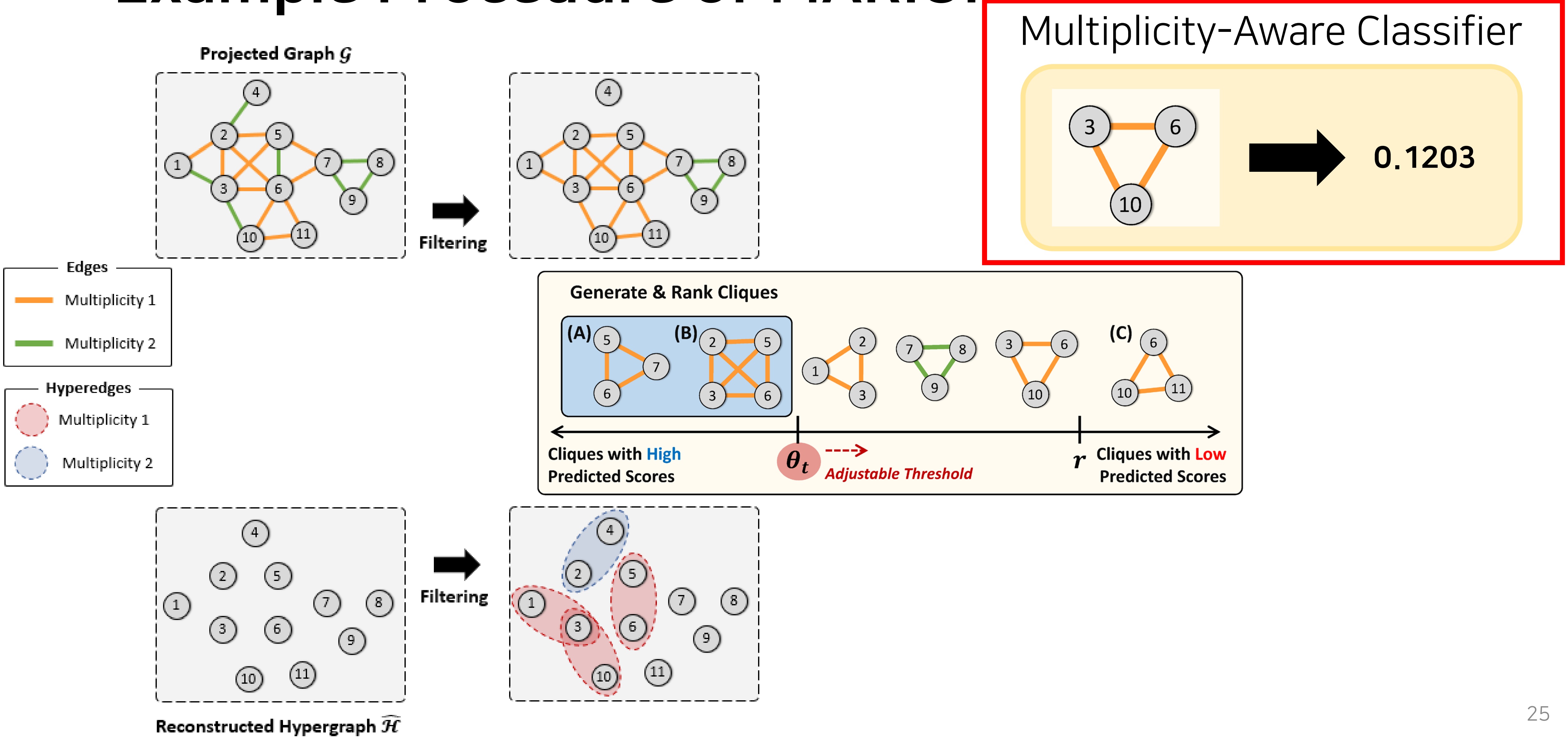


Objective: Identify all potential hyperedges in the intermediate graph ($\mathcal{G}'$) by evaluating maximal cliques and their substructures

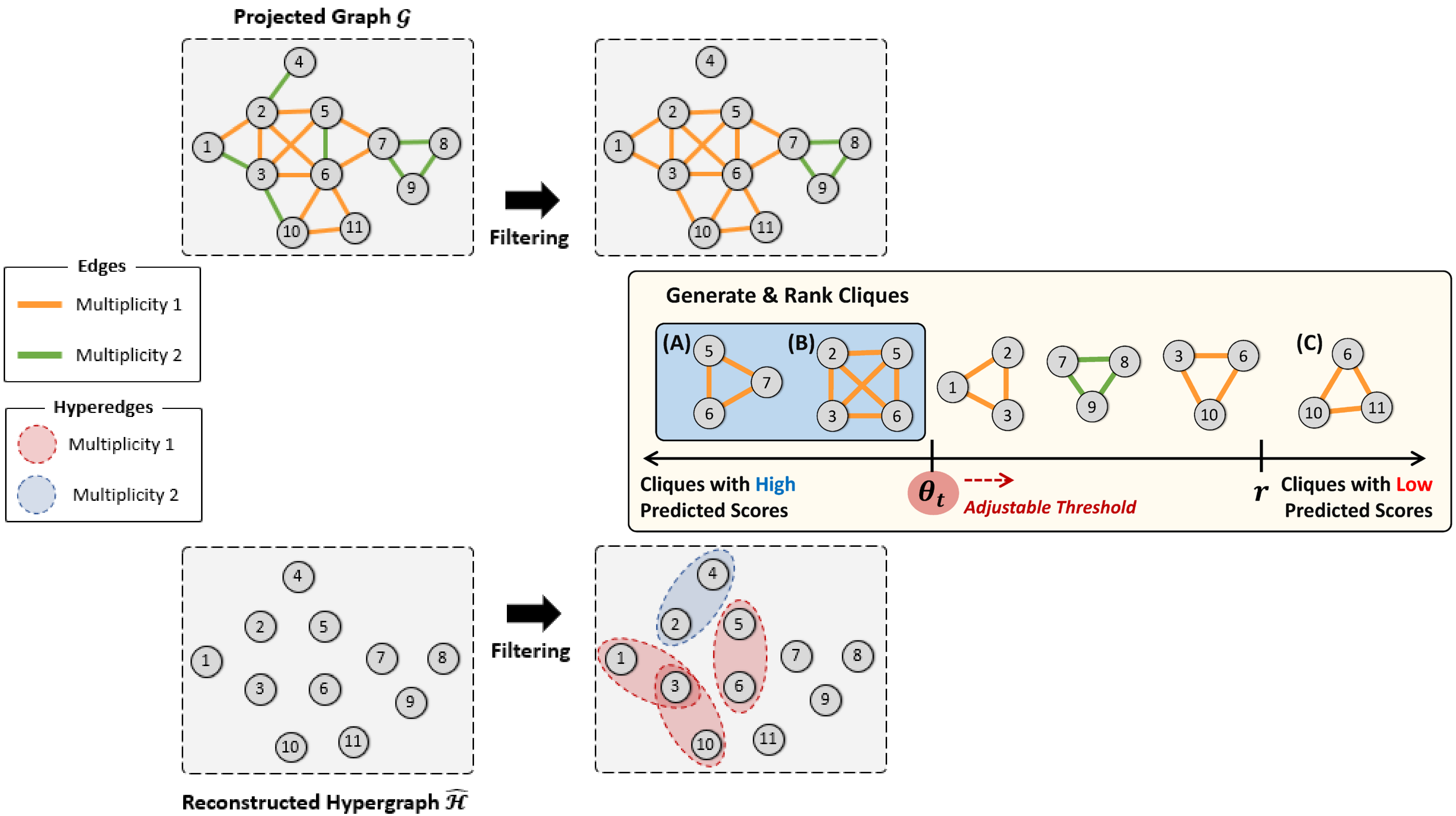
Example Procedure of MARIOH



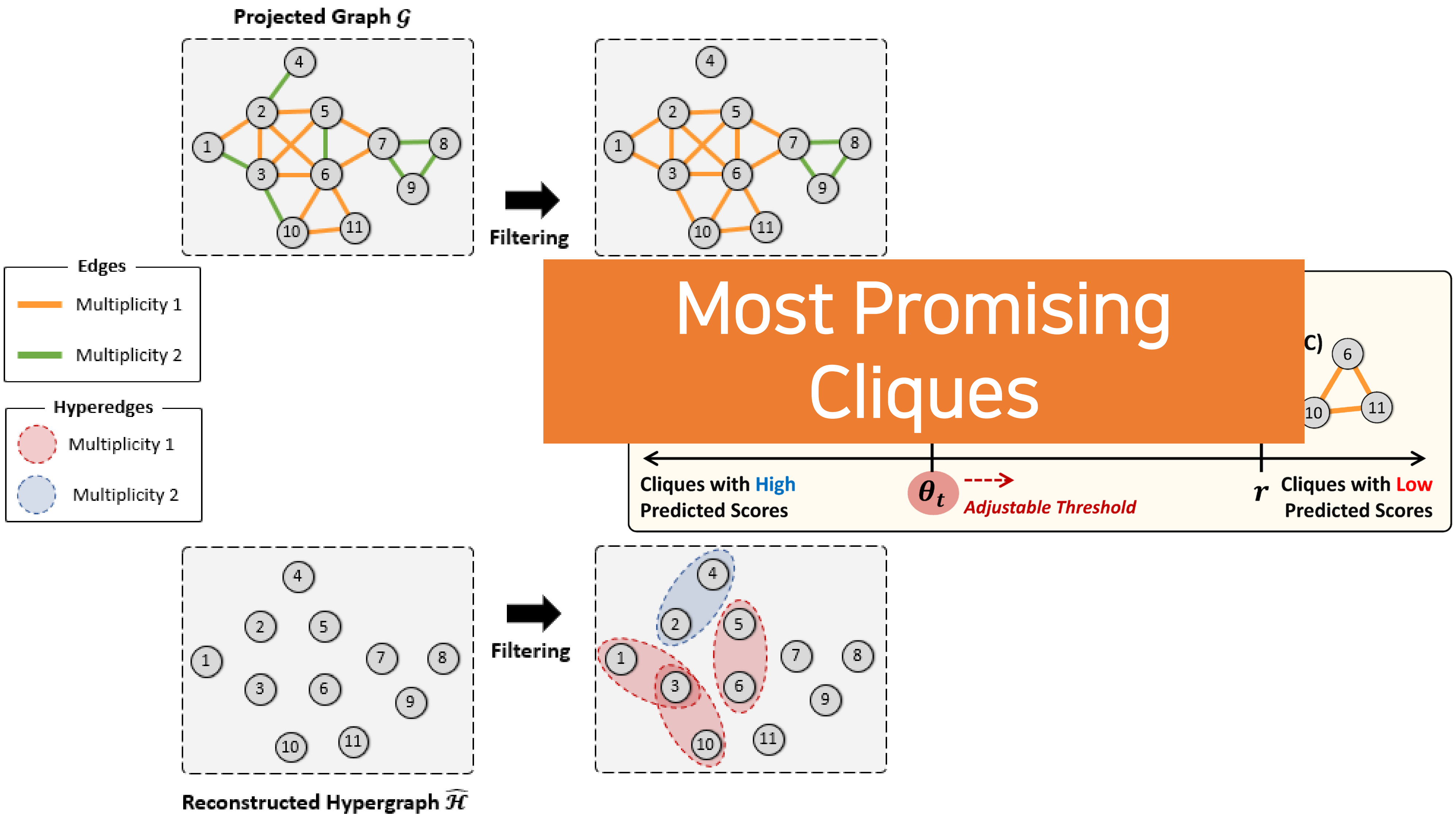
Example Procedure of MARIOH



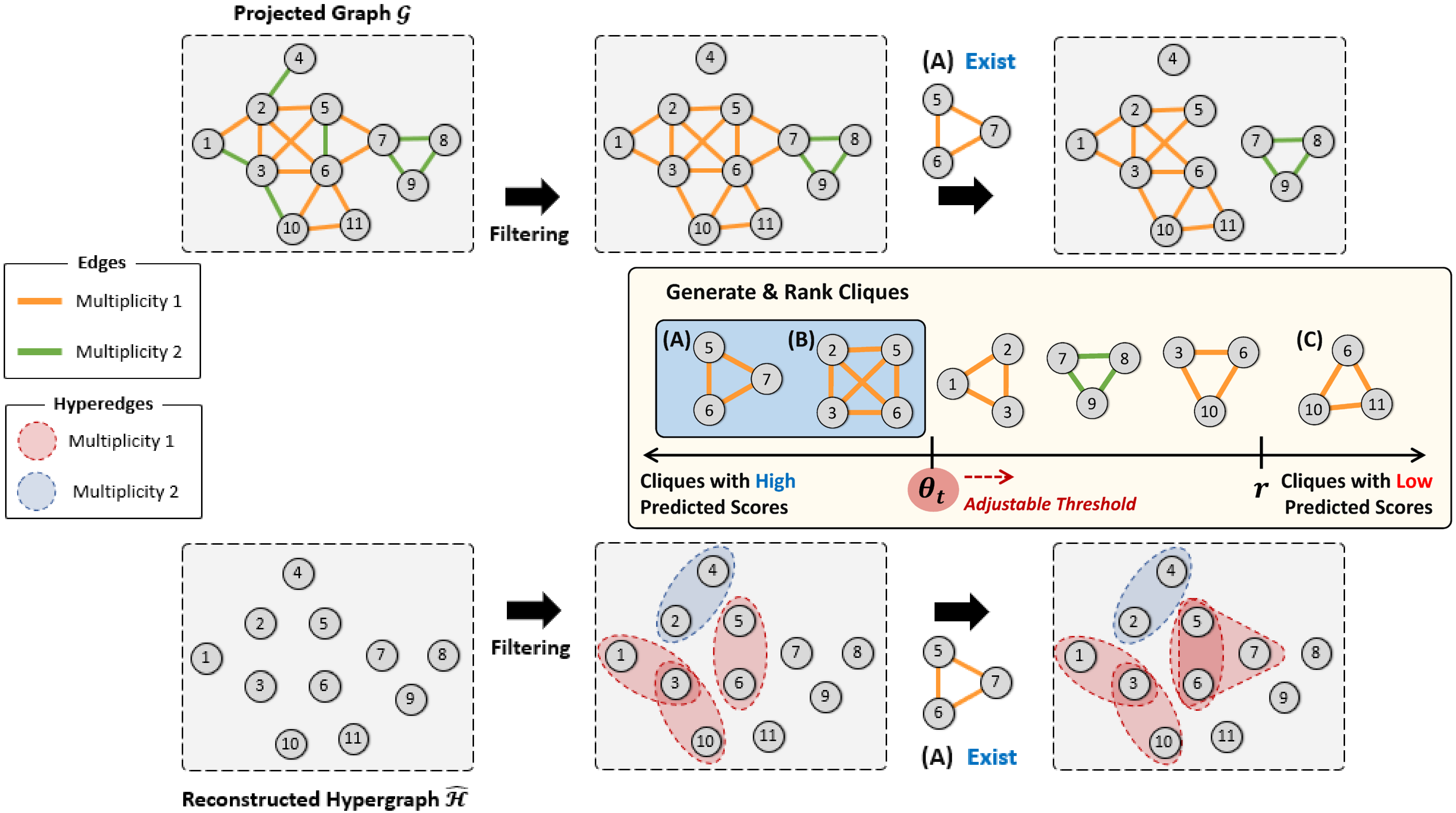
Example Procedure of MARIOH



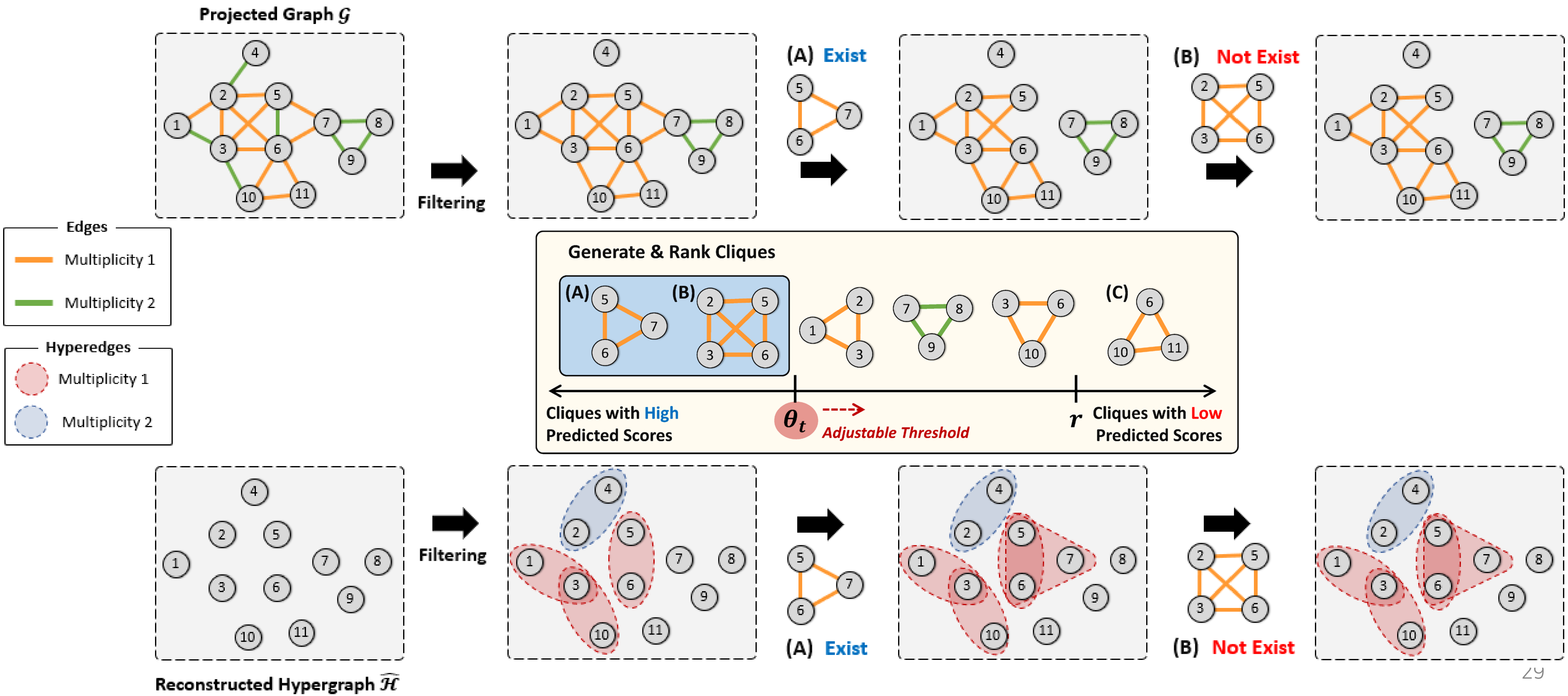
Example Procedure of MARIOH



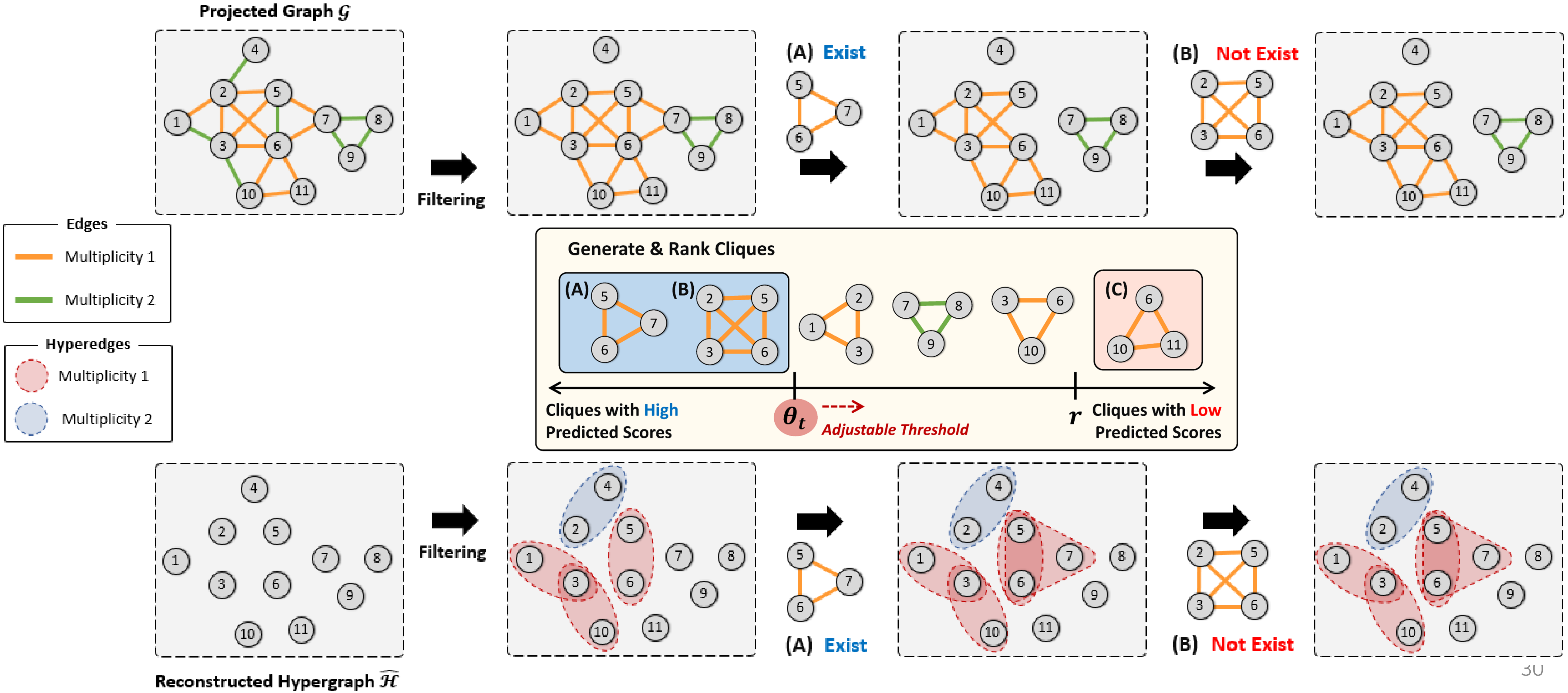
Example Procedure of MARIOH



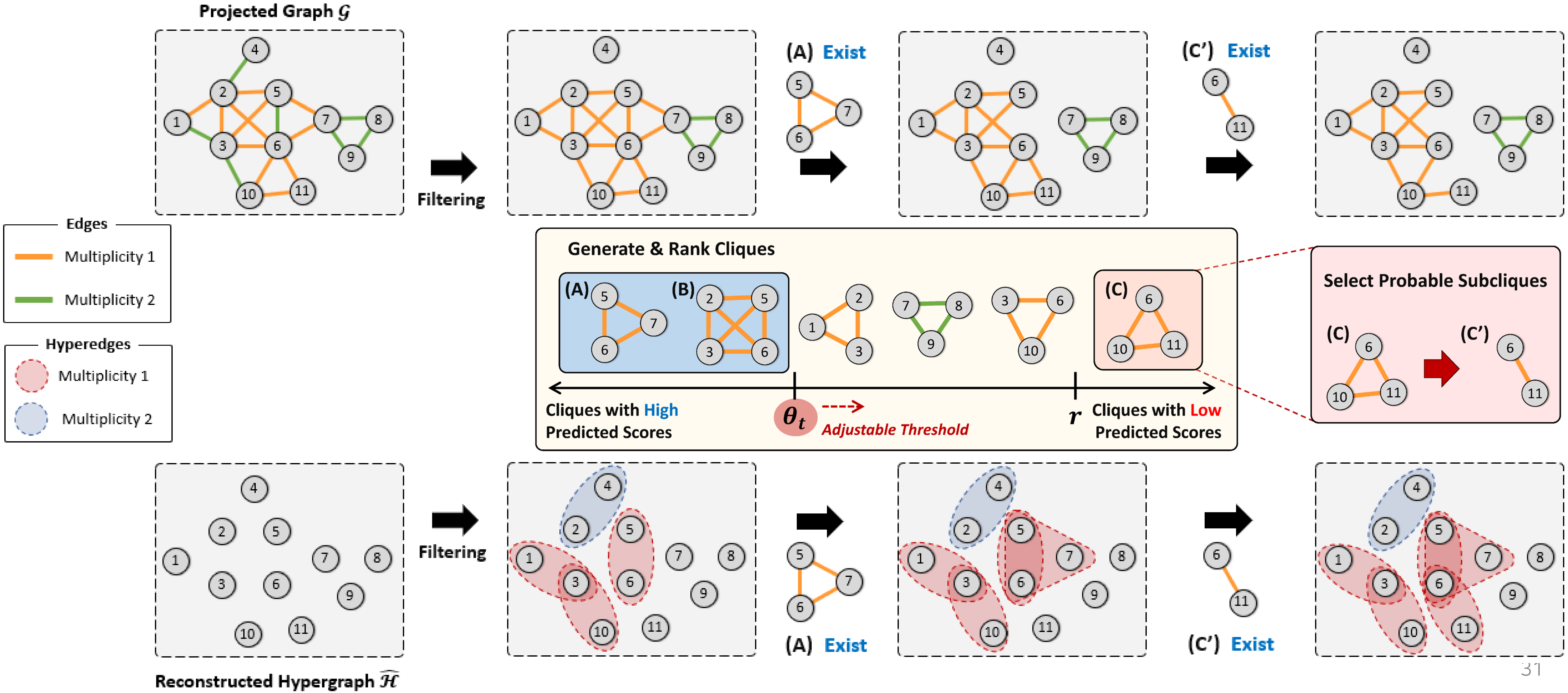
Example Procedure of MARIOH



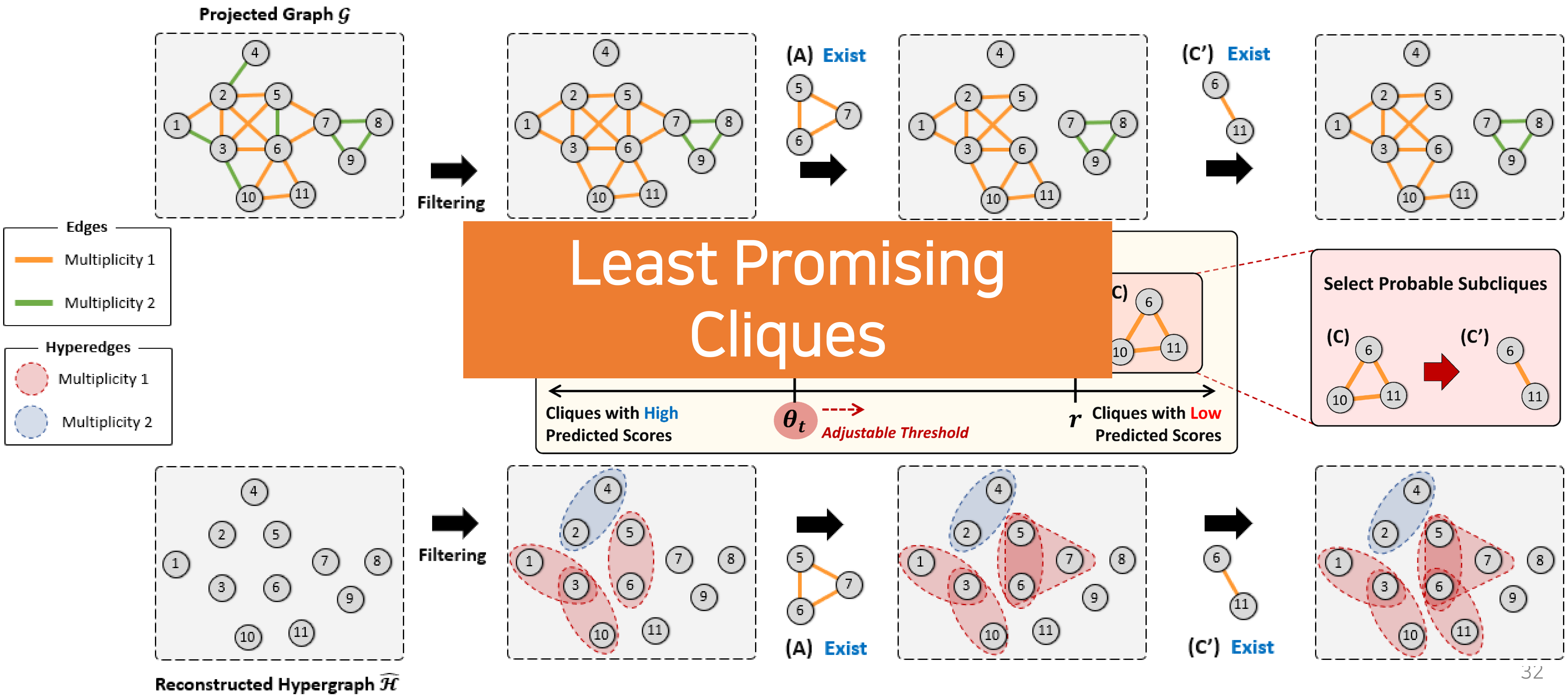
Example Procedure of MARIOH



Example Procedure of MARIOH



Example Procedure of MARIOH



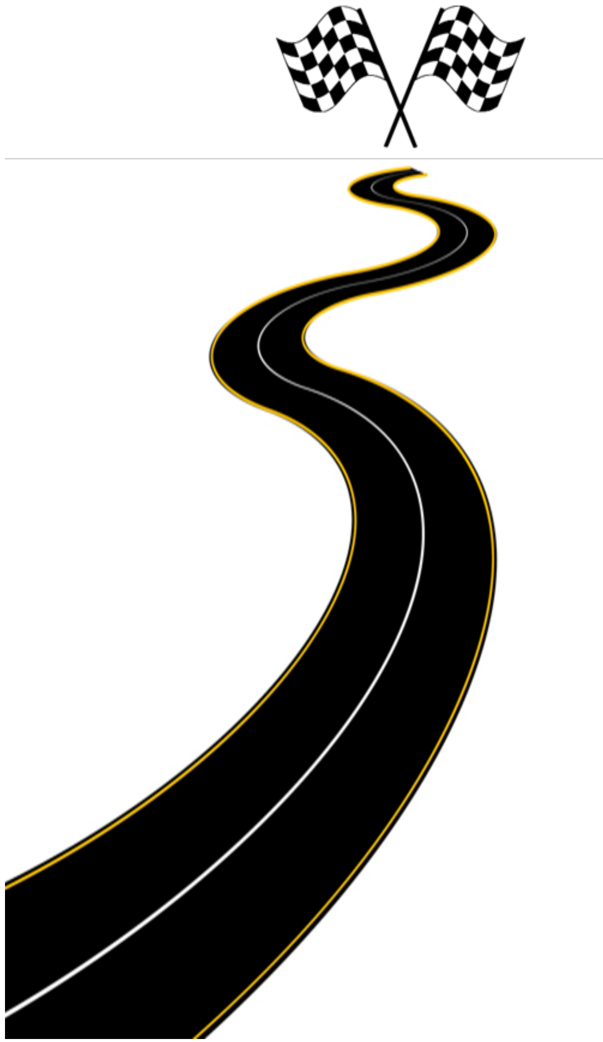
Key Components of MARIOH

Bidirectional Search: Iterate the following steps

- **Clique evaluation**
  - Assign a **score** to each maximal clique using multiplicity-aware classifier
- **Threshold-based selection**
  - Apply an **adjustable threshold** to select cliques with sufficiently high score
- **Sub-clique exploration**
  - Identify high-score **sub-cliques** hidden within low-score cliques
- **Threshold adjustment**
  - Gradually lower the threshold over iterations

Road Map

- Introduction
- Related Work
- Proposed Algorithm: MARIOH
- Experimental Results <<
- Conclusion

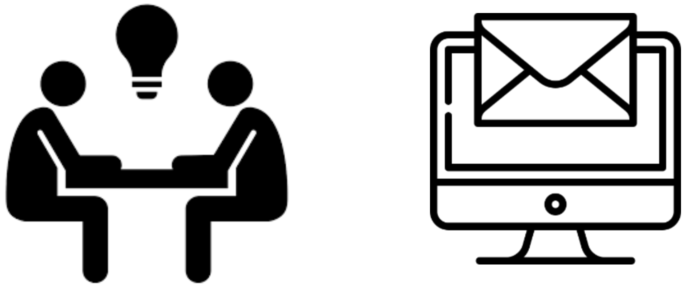


Experimental Settings

Datasets: 10 Real-world Graphs

Hyperedges are split into two halves:

- Timestamp-based split if timestamps are available
- Random split otherwise



Collaboration Email

Chat Restaurant And others!

Source and Target Hypergraphs:

- Source hypergraph is used to train MARIOH
- Target hypergraph is used for evaluation

Multiplicity Settings:

- Multiplicity-Reduced Hypergraphs: all hyperedges appear at most once
- Multiplicity-Preserved Hypergraphs: hyperedges may appear multiple times

Experimental Settings

Competitors:

Overlapping Community Detection-based Methods

- Demon [Coscia et al., 2012], Cfinder [Palla et al., 2005]

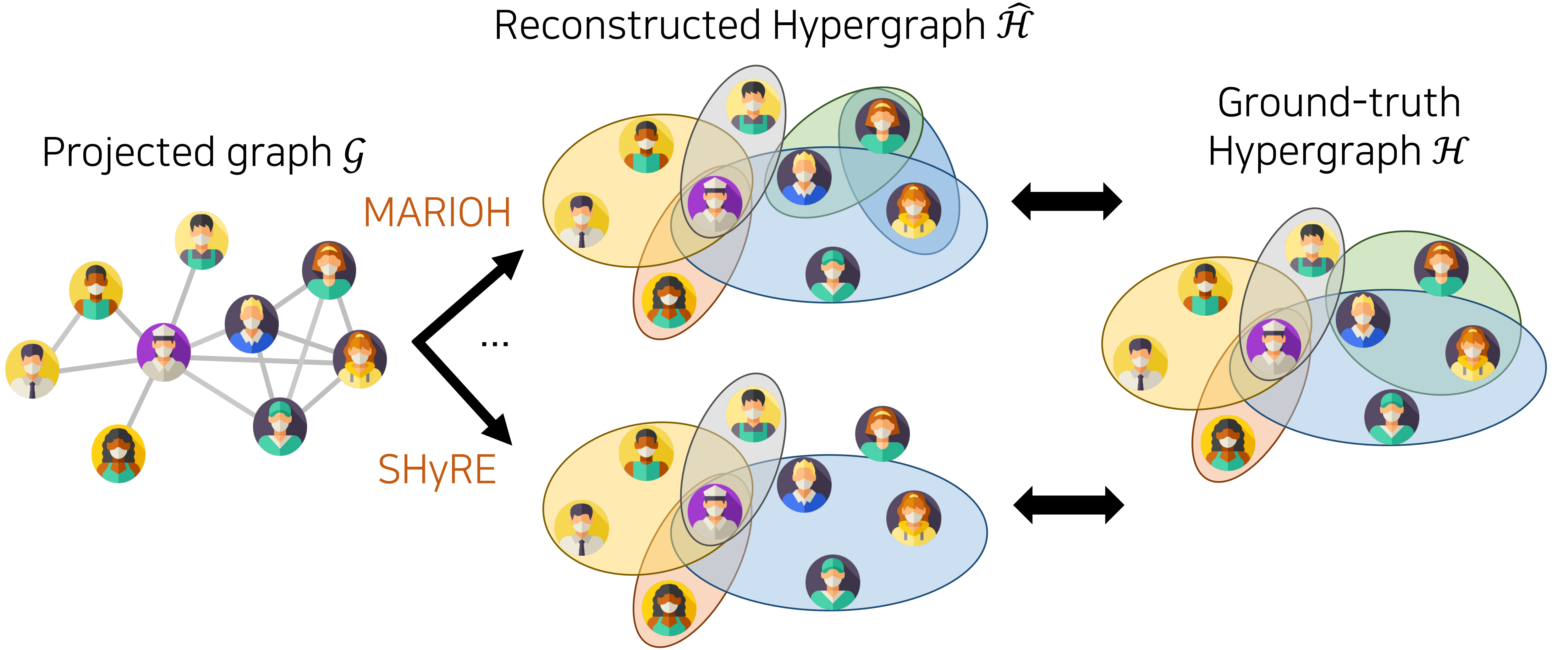
Clique Decomposition-based Methods

- Max Clique [Tomita et al., 2006]. Clique Covering [Conte et al. 2016]

Hypergraph Reconstruction Methods

- Bayesian-MDL [Young et al., 2021], SHyRE-Count, SHyRE-Motif, SHyRE-Unsup. [Wang and Kleinberg, 2024]

Comparison of Reconstruction Accuracy



MARIOH Excels in Reconstruction Accuracy

MARIOH consistently and significantly achieves higher reconstruction accuracy compared to its competitors across all datasets

Method	Enron	P.School	H.School	Crime	Hosts	Directors	Foursquare	DBLP	Eu	MAG-TopCS
CFinder	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	24.04 ± 0.00	2.50 ± 0.00	41.18 ± 0.00	7.81 ± 0.00	21.50 ± 0.00	0.01 ± 0.00	25.34 ± 0.00
Demon	2.43 ± 0.00	0.09 ± 0.00	2.97 ± 0.00	76.27 ± 0.48	9.21 ± 0.62	90.14 ± 1.39	17.01 ± 0.35	49.05 ± 0.02	0.01 ± 0.00	24.75 ± 0.05
Max Clique	4.31 ± 0.00	0.09 ± 0.00	2.38 ± 0.00	92.82 ± 0.00	23.19 ± 0.00	**100.00 ± 0.00**	9.55 ± 0.00	84.51 ± 0.00	0.98 ± 0.00	82.12 ± 0.00
Clique Covering	6.84 ± 0.00	1.95 ± 0.00	6.89 ± 0.00	93.24 ± 0.00	54.19 ± 0.00	**100.00 ± 0.00**	93.69 ± 0.00	83.87 ± 0.00	7.11 ± 0.00	84.27 ± 0.00
Bayesian-MDL	4.77 ± 0.26	0.18 ± 0.02	3.57 ± 0.05	93.15 ± 0.20	53.10 ± 0.32	**100.00 ± 0.00**	80.65 ± 1.06	86.06 ± 0.01	4.76 ± 0.03	87.26 ± 0.02
SHyRe-Unsup	13.74 ± 0.00	8.34 ± 0.00	17.00 ± 0.00	94.86 ± 0.00	50.38 ± 0.00	**100.00 ± 0.00**	92.27 ± 0.00	OOT	5.19 ± 0.00	94.31 ± 0.00
SHyRe-Motif	14.14 ± 3.27	OOT	54.21 ± 0.31	92.82 ± 0.00	49.69 ± 0.59	**100.00 ± 0.00**	70.92 ± 6.65	85.99 ± 0.04	OOM	86.04 ± 0.11
SHyRe-Count	14.36 ± 0.50	42.87 ± 1.66	54.19 ± 0.22	92.82 ± 0.00	56.64 ± 0.55	**100.00 ± 0.00**	85.96 ± 0.06	86.07 ± 0.00	11.14 ± 0.27	87.14 ± 0.09
MARIOH-M	19.48 ± 0.45	47.04 ± 0.56	55.69 ± 0.18	93.65 ± 0.30	56.17 ± 2.64	**100.00 ± 0.00**	94.03 ± 3.04	95.86 ± 0.44	11.85 ± 0.05	96.57 ± 9.24
MARIOH-F	22.95 ± 0.92	47.26 ± 0.64	57.66 ± 0.40	96.33 ± 7.35	57.47 ± 3.46	**100.00 ± 0.00**	97.27 ±1.80	97.82 ± 0.11	12.48 ± 0.41	97.01 ± 0.12
MARIOH-B	21.23 ± 1.79	9.59 ± 0.38	18.04 ± 0.32	**100.00 ± 0.00**	60.08 ± 5.34	**100.00 ± 0.00**	**100.00 ± 0.00**	**98.25 ± 0.02**	11.87 ± 0.13	**98.20 ± 0.09**
MARIOH	**25.06 ± 0.60**	**47.48 ± 0.65**	**57.75 ± 0.29**	**100.00 ± 0.00**	**61.52 ± 2.63**	100.00 ± 0.00	99.19 ± 0.34	98.13 ± 0.06	**13.98 ± 0.23**	98.04 ± 0.12

Multiplicity-Reduced Hypergraphs: All hyperedges appear at most once

MARIOH Excels in Reconstruction Accuracy

MARIOH achieves higher reconstruction accuracy than its variants:

- **MARIOH-M**: MARIOH without multiplicity-aware features
- **MARIOH-F**: MARIOH without filtering step
- **MARIOH-B**: MARIOH without considering least promising cliques

Method	Enron	P.School	H.School	Crime	Hosts	Directors	Foursquare	DBLP	Eu	MAG-TopCS
MARIOH-M	48.89 ± 0.40	51.49 ± 0.47	68.66 ± 0.21	91.84 ± 2.11	50.32 ± 12.86	**100.00 ± 0.00**	84.49 ± 1.74	92.84 ± 1.78	10.63 ± 0.75	95.73 ± 0.12
MARIOH-F	48.37 ± 0.72	53.26 ± 0.80	69.29 ± 0.25	96.33 ± 0.00	47.16 ± 16.38	93.91 ± 6.31	96.24 ± 4.51	97.08 ± 0.03	10.18 ± 0.54	96.18 ± 0.10
MARIOH-B	47.61 ± 0.19	27.81 ± 0.10	58.14 ± 0.05	**100.00 ± 0.00**	54.63 ± 19.68	**100.00 ± 0.00**	**99.01 ± 0.99**	**97.49 ± 0.07**	9.88 ± 0.41	**97.76 ± 0.06**
MARIOH	**52.26 ± 1.54**	**53.48 ± 1.42**	**69.85 ± 0.21**	**100.00 ± 0.00**	**58.41 ± 15.22**	**100.00 ± 0.00**	98.88 ± 0.46	97.48 ± 0.05	**11.55 ± 0.52**	97.62 ± 0.03

Multiplicity-Reduced Hypergraphs: All hyperedges appear at most once

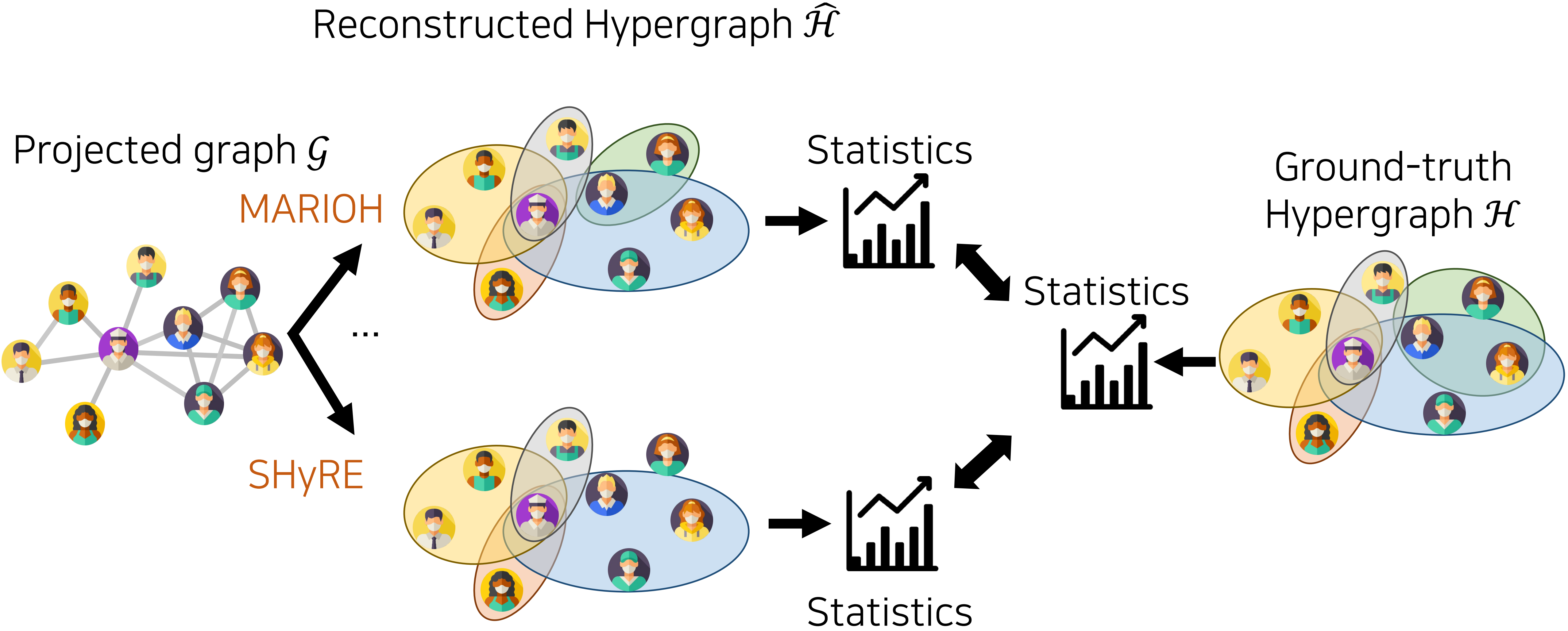
MARIOH Excels in Reconstruction Accuracy

MARIOH consistently and significantly achieves higher reconstruction accuracy compared to its competitors across all datasets

Method	Enron	P.School	H.School	Crime	Hosts	Directors	Foursquare	DBLP	Eu	MAG-TopCS
BAYESIAN-MDL	3.51 ± 0.16	0.11 ± 0.02	2.38 ± 0.11	95.06 ± 0.40	56.25 ± 1.29	**100.00 ± 0.00**	80.18 ± 0.76	85.10 ± 0.01	4.92 ± 0.05	87.36 ± 0.03
SHYRE-UNSUP	43.45 ± 0.00	25.30 ± 0.00	56.76 ± 0.00	**100.00 ± 0.00**	12.39 ± 0.00	**100.00 ± 0.00**	97.97 ± 0.00	OOT	5.62 ± 0.00	93.51 ± 0.00
MARIOH-M	48.89 ± 0.40	51.49 ± 0.47	68.66 ± 0.21	91.84 ± 2.11	50.32 ± 12.86	**100.00 ± 0.00**	84.49 ± 1.74	92.84 ± 1.78	10.63 ± 0.75	95.73 ± 0.12
MARIOH-F	48.37 ± 0.72	53.26 ± 0.80	69.29 ± 0.25	96.33 ± 0.00	47.16 ± 16.38	93.91 ± 6.31	96.24 ± 4.51	97.08 ± 0.03	10.18 ± 0.54	96.18 ± 0.10
MARIOH-B	47.61 ± 0.19	27.81 ± 0.10	58.14 ± 0.05	**100.00 ± 0.00**	54.63 ± 19.68	**100.00 ± 0.00**	**99.01 ± 0.99**	**97.49 ± 0.07**	9.88 ± 0.41	**97.76 ± 0.06**
MARIOH	**52.26 ± 1.54**	**53.48 ± 1.42**	**69.85 ± 0.21**	**100.00 ± 0.00**	**58.41 ± 15.22**	**100.00 ± 0.00**	98.88 ± 0.46	97.48 ± 0.05	**11.55 ± 0.52**	97.62 ± 0.03

Multiplicity-Preserved Hypergraphs: hyperedge may appear multiple times

Comparison of Structure Preservation

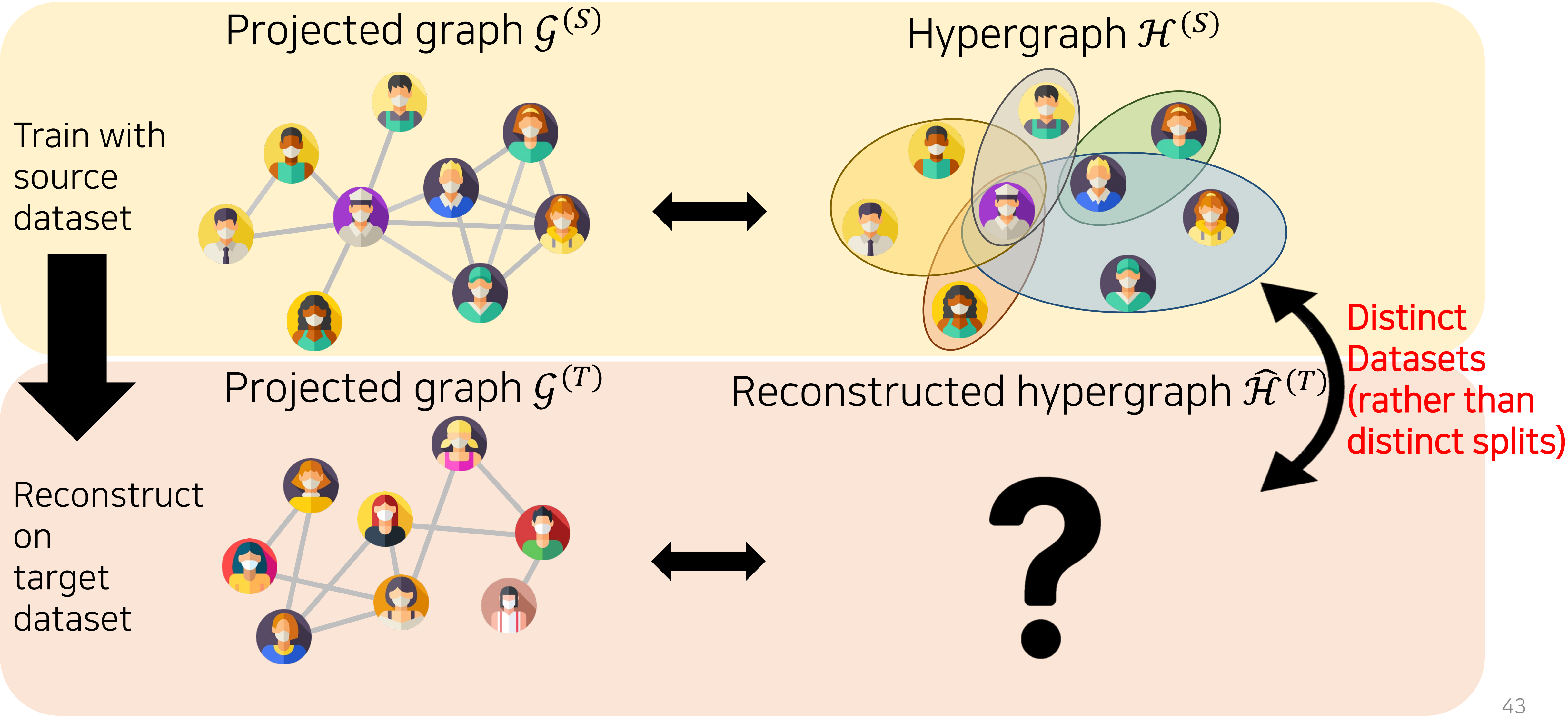


MARIOH Excels in Preserving Structures

MARIOH preserves macroscopic structural properties with the lowest error across 12 scalar and distributional metrics

Structural Properties		Bayesian-MDL	SHyRe-Count	SHyRe-Motif	SHyRe-Unsup	MARIOH
Scalar Properties	Number of Nodes	**0.000 ± 0.000**	0.013 ± 0.033	0.007 ± 0.017	**0.000 ± 0.000**	**0.000 ± 0.000**
	Number of Hyperedges	0.279 ± 0.339	0.180 ± 0.235	<u>0.166 ± 0.211</u>	0.167 ± 0.198	**0.087 ± 0.134**
	Average Node Degree	0.178 ± 0.247	0.151 ± 0.236	0.161 ± 0.227	<u>0.095 ± 0.149</u>	**0.044 ± 0.070**
	Average Hyperedge Size	0.187 ± 0.223	<u>0.085 ± 0.086</u>	0.087 ± 0.091	0.089 ± 0.093	**0.060 ± 0.079**
	Simplicial Closure Ratio	0.151 ± 0.314	0.178 ± 0.372	0.179 ± 0.368	<u>0.139 ± 0.236</u>	**0.107 ± 0.186**
	Hypergraph Density	0.279 ± 0.339	0.175 ± 0.224	<u>0.163 ± 0.205</u>	0.167 ± 0.198	**0.087 ± 0.134**
	Hypergraph Overlapness	0.178 ± 0.247	0.151 ± 0.236	0.161 ± 0.227	<u>0.095 ± 0.149</u>	**0.044 ± 0.070**
Distributional Properties	Node Degree	0.154 ± 0.235	0.118 ± 0.196	0.128 ± 0.192	<u>0.081 ± 0.155</u>	**0.033 ± 0.055**
	Node-Pair Degree	0.059 ± 0.073	0.087 ± 0.133	0.101 ± 0.128	**0.011 ± 0.020**	<u>0.025 ± 0.048</u>
	Node-Triple Degree	0.052 ± 0.071	0.155 ± 0.373	0.041 ± 0.049	<u>0.016 ± 0.027</u>	**0.015 ± 0.030**
	Hyperedge Homogeneity	0.151 ± 0.118	0.213 ± 0.229	0.241 ± 0.230	**0.055 ± 0.068**	<u>0.073 ± 0.104</u>
	Singular Values	0.035 ± 0.036	0.098 ± 0.124	0.102 ± 0.118	**0.034 ± 0.038**	**0.034 ± 0.038**
	Average (Overall)	0.142 ± 0.090	0.134 ± 0.055	0.128 ± 0.064	<u>0.079 ± 0.058</u>	**0.051 ± 0.032**

Comparison of Transferability



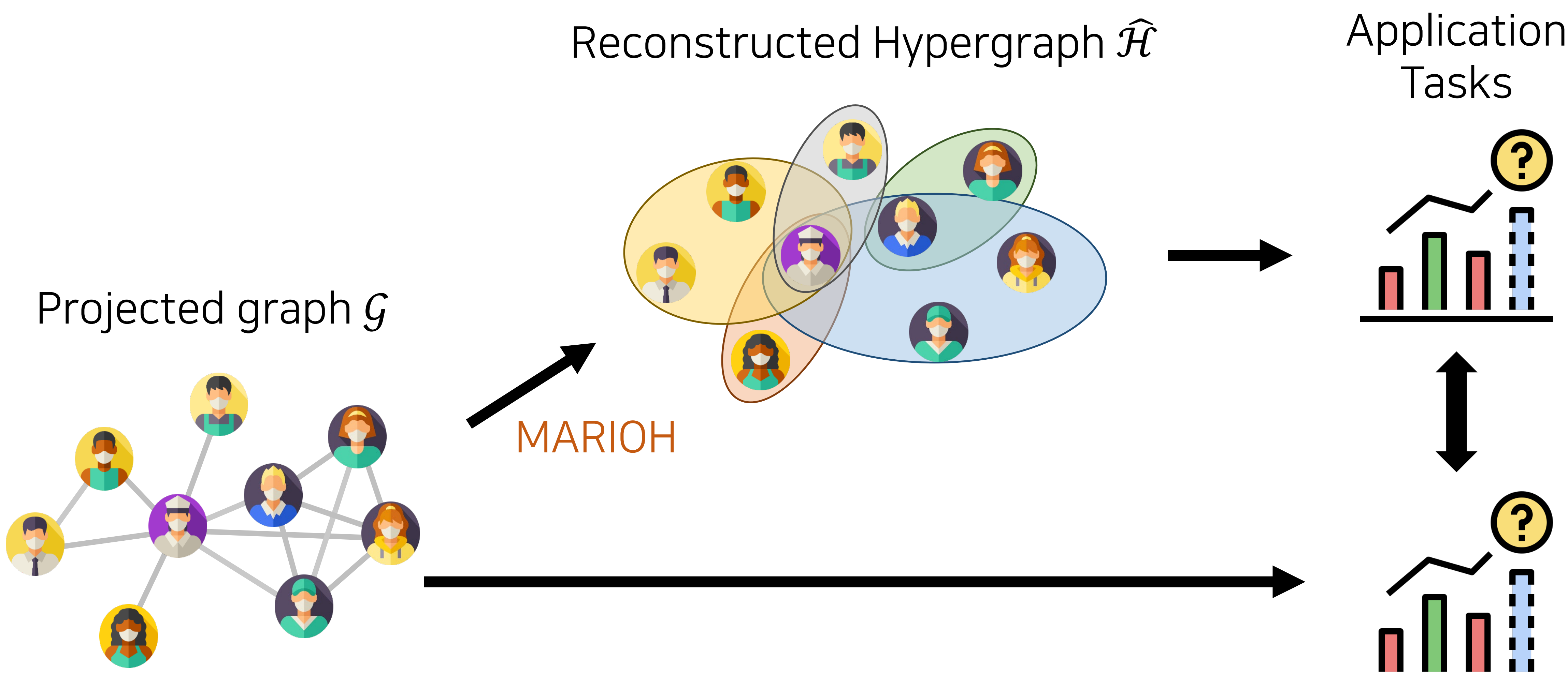
MARIOH is Transferable

MARIOH generalizes effectively across datasets, even with limited training data

Source Dataset	DBLP				Eu		P.School	
Target Dataset	**DBLP**	**MAG-History**	**MAG-TopCS**	**MAG-Geology**	**Eu**	**Enron**	**P.School**	**H.School**
SHyRE-Unsup	OOT	OOT	84.50 ± 0.00	OOT	5.19 ± 0.00	13.74 ± 0.00	8.34 ± 0.00	17.00 ± 0.00
SHyRE-Motif	85.99 ± 0.04	76.73 ± 0.52	OOT	OOT	OOM	OOM	OOT	OOT
SHyRE-Count	86.07 ± 0.00	93.86 ± 0.34	78.08 ± 0.36	48.64 ± 0.31	11.14 ± 0.27	23.08 ± 0.08	42.87 ± 1.66	43.83 ± 0.28
MARIOH	**98.13 ± 0.06**	**97.80 ± 0.57**	**92.92 ± 0.22**	**81.66 ± 0.73**	**13.98 ± 0.23**	**27.59 ± 2.50**	**47.48 ± 0.65**	**49.52 ± 2.09**

Method	DBLP	Hosts	Enron
Bayesian-MDL	86.06 ± 0.01	53.10 ± 0.32	4.77 ± 0.26
SHyRE-Motif	85.99 ± 0.04	49.69 ± 0.59	14.14 ± 3.27
SHyRE-Count	86.07 ± 0.00	56.64 ± 0.55	14.36 ± 0.50
MARIOH (10%)	96.52 ± 0.22	55.42 ± 2.17	22.29 ± 1.78
MARIOH (20%)	97.11 ± 0.15	56.24 ± 1.46	22.45 ± 1.14
MARIOH (50%)	97.88 ± 0.08	58.17 ± 2.74	24.10 ± 1.32
MARIOH (100%)	**98.13 ± 0.06**	**61.52 ± 2.63**	**25.06 ± 0.60**

MARIOH is Applicable



MARIOH is Applicable: Link Prediction

Using hypergraphs reconstructed by MARIOH achieves the highest AUC in link prediction

Input	Enron	P.School	H.School	Crime	Hosts	Directors	Foursquare	DBLP	Eu	MAG-TopCS	Avg. Rank
Projected graph $\mathcal{G}$	83.94 ± 2.99	90.76 ± 0.37	**92.51 ± 1.37**	78.32 ± 1.11	98.61 ± 0.06	81.90 ± 3.10	99.44 ± 0.02	**96.54 ± 0.09**	**97.46 ± 0.06**	94.86 ± 0.22	3.90 ± 2.33
$\widehat{\mathcal{H}}$ by SHyRe-Unsup	80.07 ± 4.26	91.21 ± 0.31	90.08 ± 1.40	81.38 ± 0.80	98.75 ± 0.05	82.59 ± 3.80	**99.60 ± 0.02**	OOT	97.18 ± 0.10	95.05 ± 0.22	3.50 ± 1.51
$\widehat{\mathcal{H}}$ by SHyRe-Motif	78.65 ± 3.73	OOT	89.92 ± 1.64	81.54 ± 0.94	98.76 ± 0.03	82.59 ± 3.80	99.57 ± 0.02	96.40 ± 0.11	OOT	94.94 ± 0.14	4.40 ± 1.58
$\widehat{\mathcal{H}}$ by SHyRe-Count	77.76 ± 3.09	91.04 ± 0.77	90.16 ± 1.57	81.54 ± 0.94	**98.77 ± 0.11**	82.58 ± 3.80	99.57 ± 0.01	96.40 ± 0.04	95.82 ± 0.61	94.95 ± 0.19	3.80 ± 1.32
$\widehat{\mathcal{H}}$ by MARIOH	79.75 ± 4.74	91.72 ± 0.22	90.21 ± 1.51	**81.60 ± 1.00**	98.68 ± 0.06	**82.75 ± 3.71**	**99.60 ± 0.02**	96.46 ± 0.06	97.19 ± 0.09	**95.10 ± 0.17**	**2.30 ± 1.42**
Original Hypergraph $\mathcal{H}$	**84.55 ± 3.20**	**92.16 ± 0.19**	90.35 ± 1.18	81.59 ± 0.82	98.70 ± 0.06	82.58 ± 3.80	**99.60 ± 0.02**	96.39 ± 0.06	97.26 ± 0.13	95.06 ± 0.14	2.40 ± 1.43

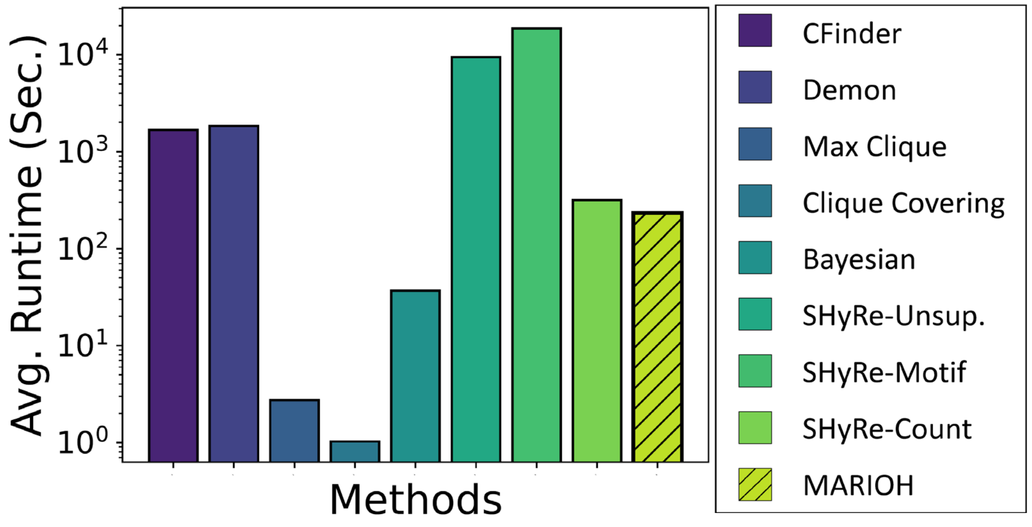
MARIOH is Applicable: Clustering

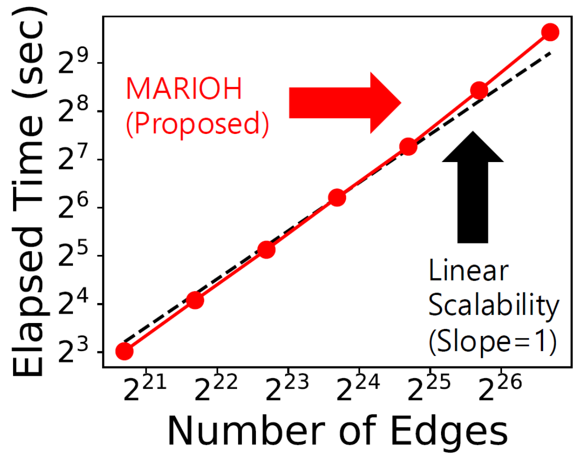
Spectral Clustering on hypergraphs reconstructed by MARIOH achieves higher NMI than those reconstructed by the competitors

Input of Spectral Clustering	**P.School**	**H.School**
Projected graph $\mathcal{G}$	0.8488	0.9392
$\widehat{\mathcal{H}}$ by SHyRe-Unsup	0.8982	0.9635
$\widehat{\mathcal{H}}$ by SHyRe-Motif	OOT	0.9830
$\widehat{\mathcal{H}}$ by SHyRe-Count	0.9095	0.9874
$\widehat{\mathcal{H}}$ by MARIOH	<u>0.9234</u>	**0.9936**
Original Hypergraph $\mathcal{H}$	**0.9255**	**0.9936**

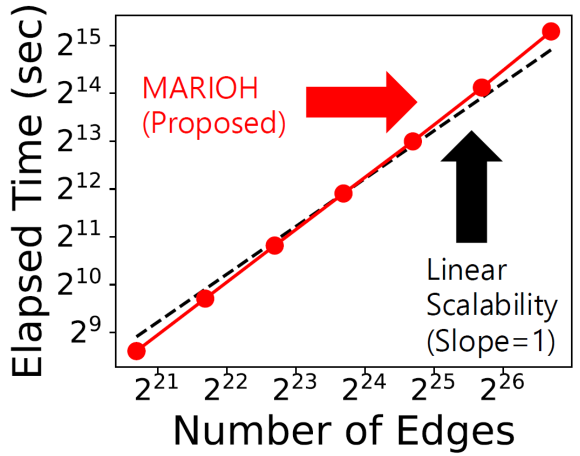
MARIOH is Fast and Scalable

MARIOH executes faster than recent advanced methods (e.g., SHyRE-Unsup., SHyRE-Count, and SHyRE-Motif) and scales near-linearly with graph size





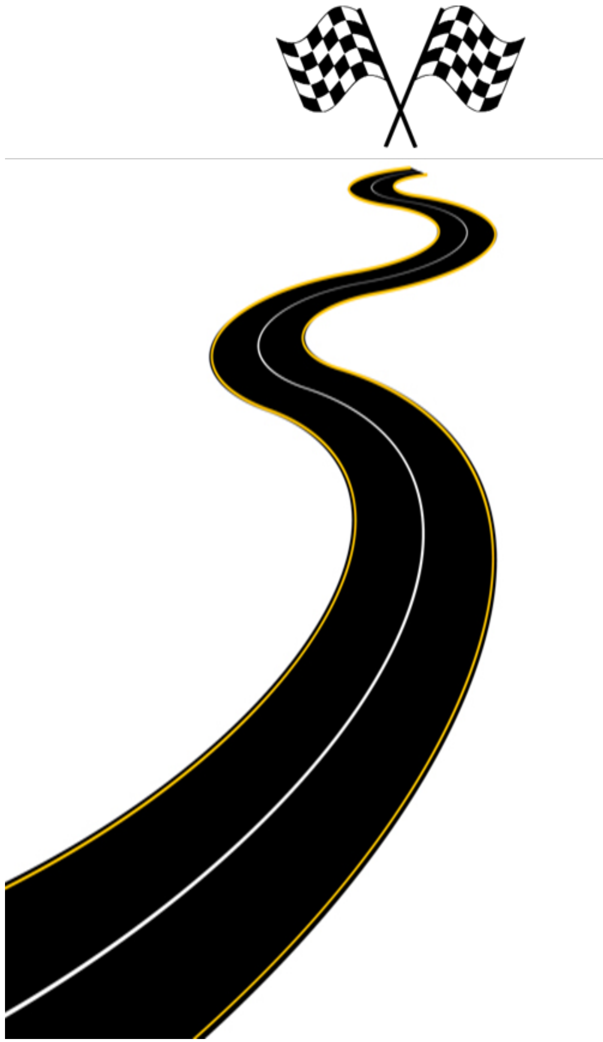
(a) Filtering Step



(b) Bidirectional Search Step

Road Map

- Introduction
- Related Work
- Proposed Algorithm: MARIOH
- Experimental Results
- Conclusion <<



Key Takeaways of MARIOH

🎯 Accurate Recovery

- Achieves up to 74.51% higher accuracy compared to existing methods

⇆ Transferability

- Generalizes effectively across datasets within the same domain

🚀 Applications

- Improves downstream tasks such as link prediction and node clustering

Code available at https://github.com/KyuhanLee/MARIOH

MARIOH: Multiplicity-Aware Hypergraph Reconstruction



Kyuhan Lee[1,2] Geon Lee[1] Kijung Shin[1]

[1]Kim Jaechul Graduate School of AI, KAIST
[2]GraphAI

Reconstruction Accuracy

Metric: Jaccard Similarity

- Measures the similarity between the original hyperedges ($\mathcal{E}_{\mathcal{H}}^{(T)}$) and the reconstructed hyperedges ($\mathcal{E}_{\widehat{\mathcal{H}}}^{(T)}$)
- Even if a single node differs, two hyperedges are considered entirely distinct making this metric strict and challenging

$$\mathcal{J}\left(\mathcal{H}^{(T)}, \widehat{\mathcal{H}}^{(T)}\right) = \frac{\left|\mathcal{E}_{\mathcal{H}}^{(T)} \cap \mathcal{E}_{\widehat{\mathcal{H}}}^{(T)}\right|}{\left|\mathcal{E}_{\mathcal{H}}^{(T)} \cup \mathcal{E}_{\widehat{\mathcal{H}}}^{(T)}\right|}$$

Reconstruction Accuracy

Metric: Multi-Jaccard Similarity

- Extends Jaccard Similarity by incorporating the multiplicity of hyperedges
- Measures the similarity between the original hyperedges ($\mathcal{E}_{\mathcal{H}}^{(T)}$) and the reconstructed hyperedges ($\mathcal{E}_{\widehat{\mathcal{H}}}^{(T)}$)

$$\mathcal{J}_{multi}\left(\mathcal{H}^{(T)}, \widehat{\mathcal{H}}^{(T)}\right) = \frac{\sum_{e \in \mathcal{E}_{union}} \min(M_{\mathcal{H}}(e), M_{\widehat{\mathcal{H}}}(e))}{\sum_{e \in \mathcal{E}_{union}} \max(M_{\mathcal{H}}(e), M_{\widehat{\mathcal{H}}}(e))}$$

where $\mathcal{E}_{union} = \mathcal{E}_{\mathcal{H}}^{(T)} \cup \mathcal{E}_{\widehat{\mathcal{H}}}^{(T)}$

- Multiplicity $M_{\mathcal{H}}(e)$:
  - Refers to the number of times a hyperedge e appears in the hypergraph.

Theoretically-Guaranteed Filtering

Why are these edges theoretically guaranteed?

MHH (Maximum Higher-Order Hyperedges)

- For each edge (u, v), the MHH is the maximum possible number of higher-order hyperedges (size-3 or larger) involving both u and v.

$$\mathrm{MHH}(u, v) = \sum_{z \in N(u) \cap N(v)} \min(\omega_{u,z}, \omega_{v,z})$$

- Residual edge multiplicity $r_{u,v}$ is defined as:

$$\mathrm{r}(u, v) = \omega_{u,v} - \mathrm{MHH}(u, v)$$

Multiplicity-Aware Classifier

Predicts the score of cliques being hyperedges

Aggregates three levels of features to represent a clique C:

- Node-level: Weighted degree of each node
- Edge-level: Edge multiplicity ($\omega_{u,v}$), $\mathrm{MHH}(u, v)$, and $\mathrm{MHH}(u, v)/\omega_{u,v}$
- Clique-level: Clique size, Clique cut ratio, and maximal clique indicator

Classifier: 2-layered MLP

- Trained on structural and multiplicity-aware features of $\mathcal{G}^{(S)}$
- Negative samples obtained from $\mathcal{H}^{(S)}$