KAIST AI
Kim Jaechul Graduate School

Carnegie Mellon University

THE WEB CONFERENCE
ACM

Kronecker Generative Models for Power-Law Patterns in Real-World Hypergraphs


Minyoung Choe


Jihoon Ko

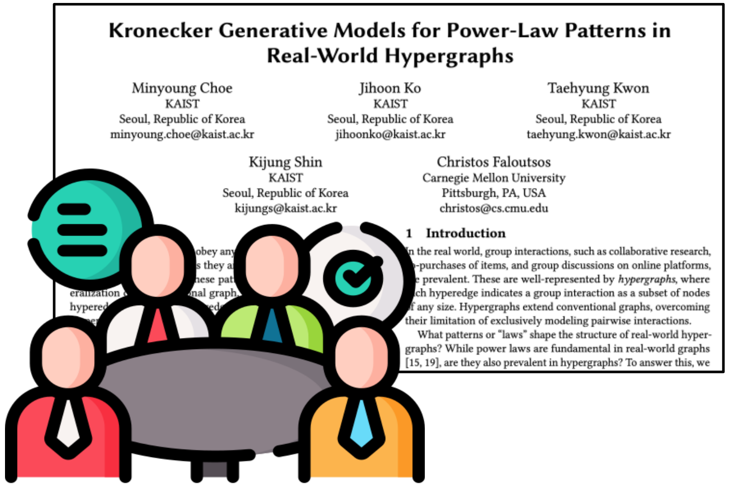

Taehyung Kwon


Kijung Shin

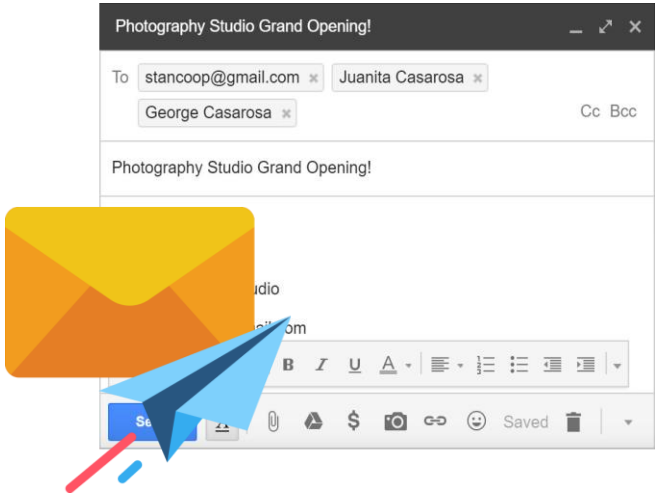

Christos Faloutsos

Group Interactions are EVERYWHERE

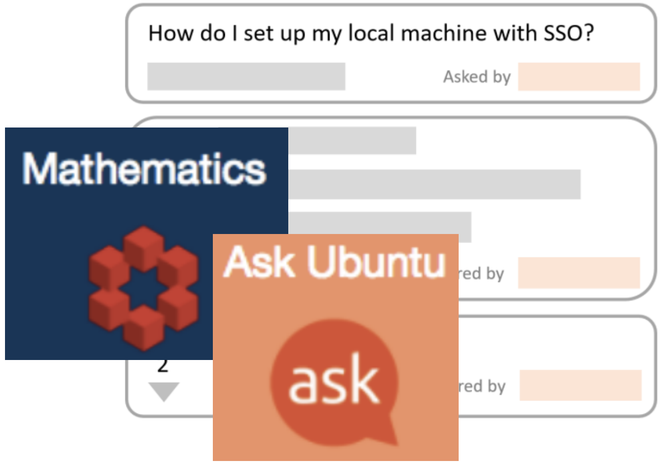
- Group interactions exist in many complex systems



Co-Authorship



Email



Q&A Platform

Hypergraphs represent group interactions

Authors (Nodes)

J. D. Watson	(W)	R. A. Laskowski	(L)
F. H. C. Crick	(C)	J. M. Thornton	(T)

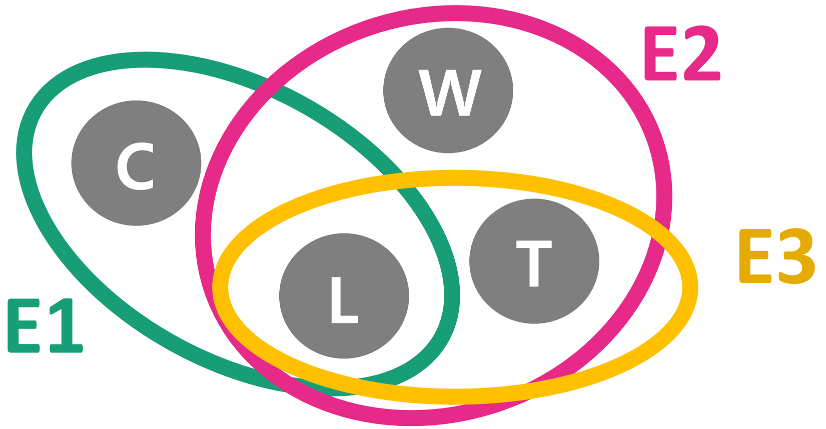
Publications (Hyperedges)

E1: THE STRUCTURE OF DNA
J. D. Watson, F. H. C. Crick – CSH' 1953

E2: Predicting protein function from sequence ...
J.D. Watson, R.A. Laskowski, J. M. Thornton – COSB' 2005

E3: Understanding the molecular machinery ...
R. A. Laskowski, J. M. Thornton – NRG' 2008

Hypergraph

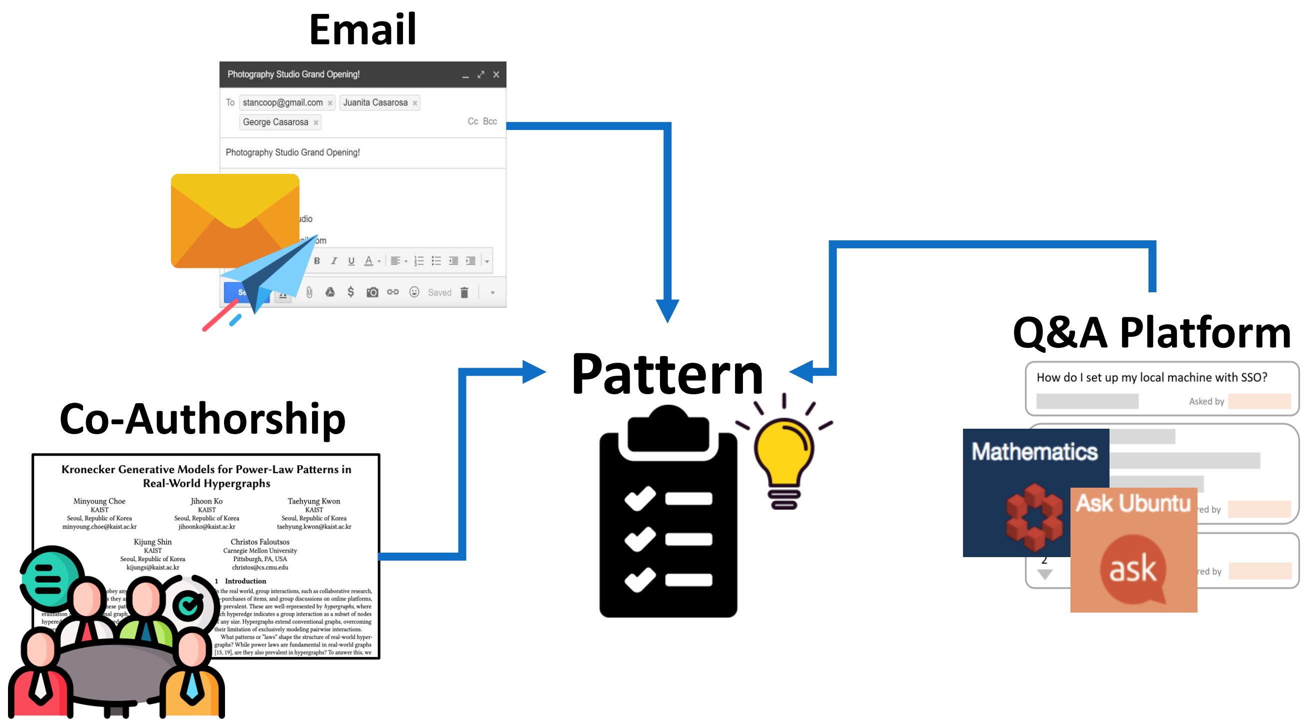


Incidence Matrix

	E1	E2	E3
W		1	
C	1		
L	1	1	1
T		1	1

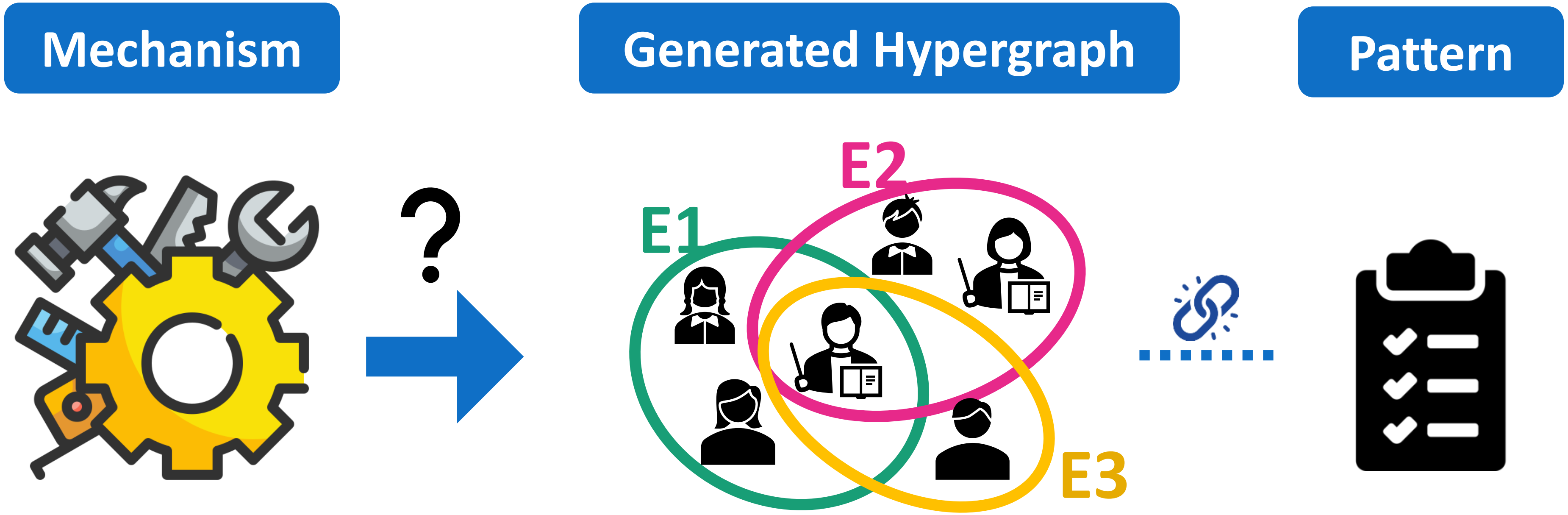
RQ 1. What Patterns Exist in Real-world Hypergraphs?

- Patterns refer to **frequently recurring** structural properties in real-world hypergraphs
- They capture **how group interactions typically form and evolve**
- Understanding these patterns helps us define what makes a hypergraph **realistic**



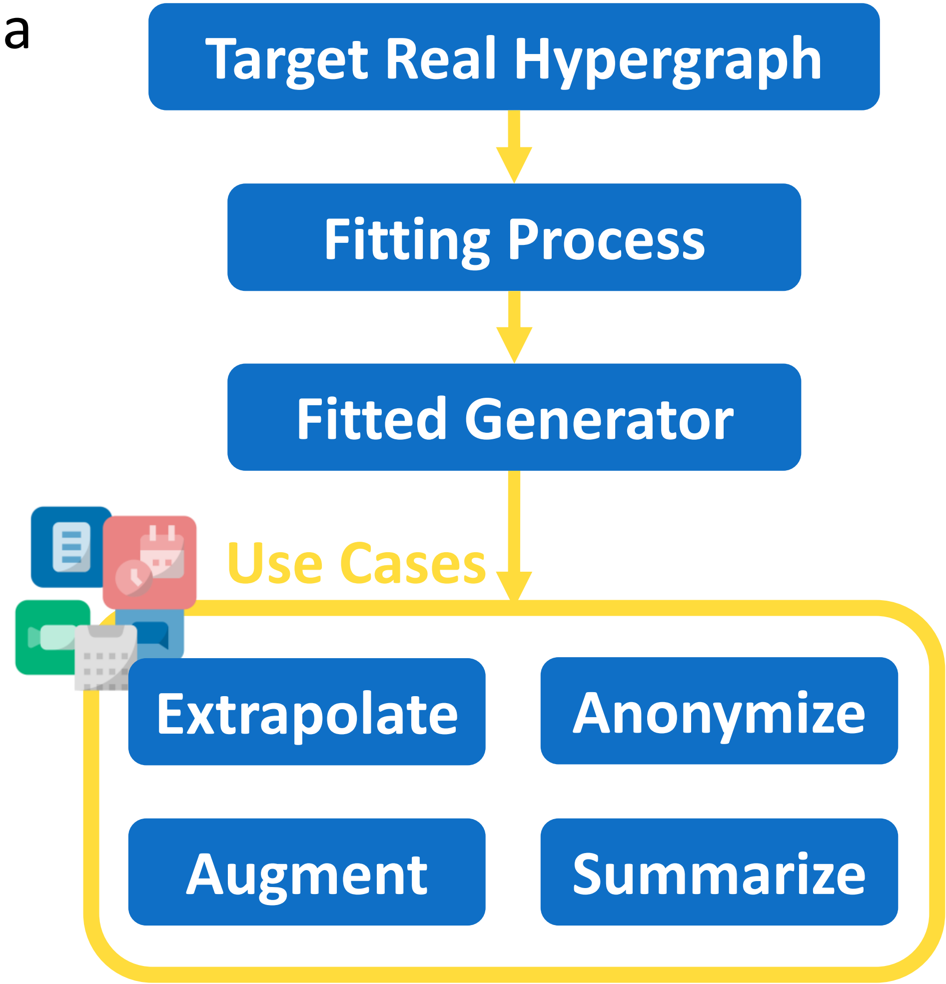
RQ 2. What Mechanisms Underlie These Patterns?

- Structural patterns in hypergraphs emerge from **underlying mechanisms**
- Identifying these mechanisms is key to designing **realistic generators**



RQ 3. How Can We Fit Generators to Real Hypergraphs?

- **Fitting** means tuning a generator to reproduce a target real hypergraph
- Once fitted, a generator can:
  - **Extrapolate:** predict future growth
  - **Anonymize:** protect privacy
  - **Augment:** create synthetic training data
  - **Summarize:** extract key patterns



Our Answers to the Research Questions

- **RQ 1. What Patterns Exist in Real-world Hypergraphs?**
  - → We discover **eight structural patterns**, including power-law and log-logistic distributions
- **RQ 2. What Mechanisms Underlie These Patterns?**
  - → We propose **HyRec**, a hypergraph generator based on **self-similar** (i.e., fractal) structure
- **RQ 3. How Can We Fit HyRec to Real Data?**
  - → We develop **SingFit**, an efficient algorithm that scales to **large real-world hypergraphs**

Discovering Eight Patterns in Real-World Hypergraphs

- **Studied 11 real-world hypergraphs from six domains**
  - Emails, Contacts, Drugs (NDC), Tags, Threads, and Co-authorship
- Identified **three key discoveries (D1–D3):**

Discovery (D1): Node pair degrees, intersection sizes, and singular values exhibit power-law distributions

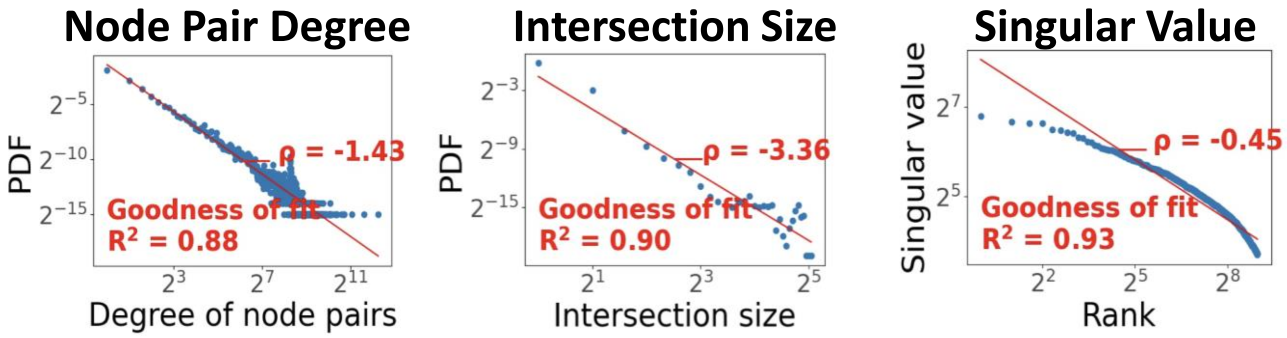
Discovery (D2): Node degrees and hyperedge sizes follow log-logistic distributions which are closely related to power laws

Discovery (D3): Clustering coefficients, density, and overlapness in egonets follow power-law patterns

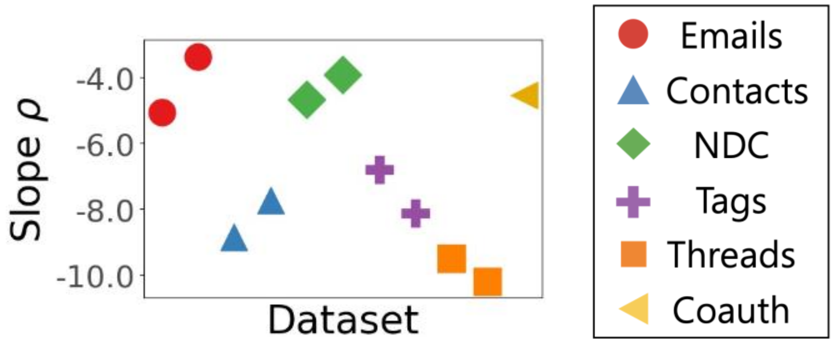
Discovery (D1) - Power-Law Distributions in Hypergraphs

Discovery (D1): Node pair degrees, intersection sizes, and singular values exhibit power-law distributions

- **Analysis:**
  - **High log-likelihood ratios (LRs) & strong linear regression fits on a log-log scale** confirm power-law behavior
  - **Similar slopes** within the same domain suggest domain-based similarities



Strong linear regression fitting (R^2 scores near 1)



Similar slopes within the same domains

Discovery (D2) - Log-Logistic Distributions in Hypergraphs

Discovery (D2): Node degrees and hyperedge sizes follow log-logistic distributions which are closely related to power laws

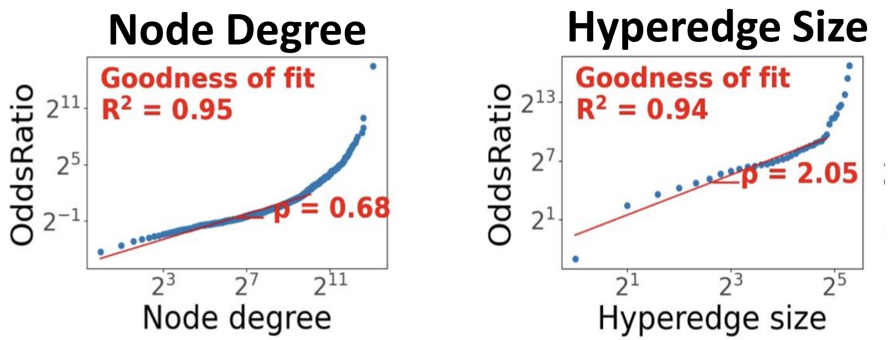
- **Analysis:**
  - **Power-law-like odds ratios** suggest a closer **fit to log-logistic distributions**
  - Linear regression on a log-log scale of odds ratios yields high R^2 scores, confirming strong log-logistic characteristics

x follows log-logistic

⬍

$$\text{OddsRatio}(x) := \frac{CDF(x)}{1-CDF(x)}$$

follows power-law

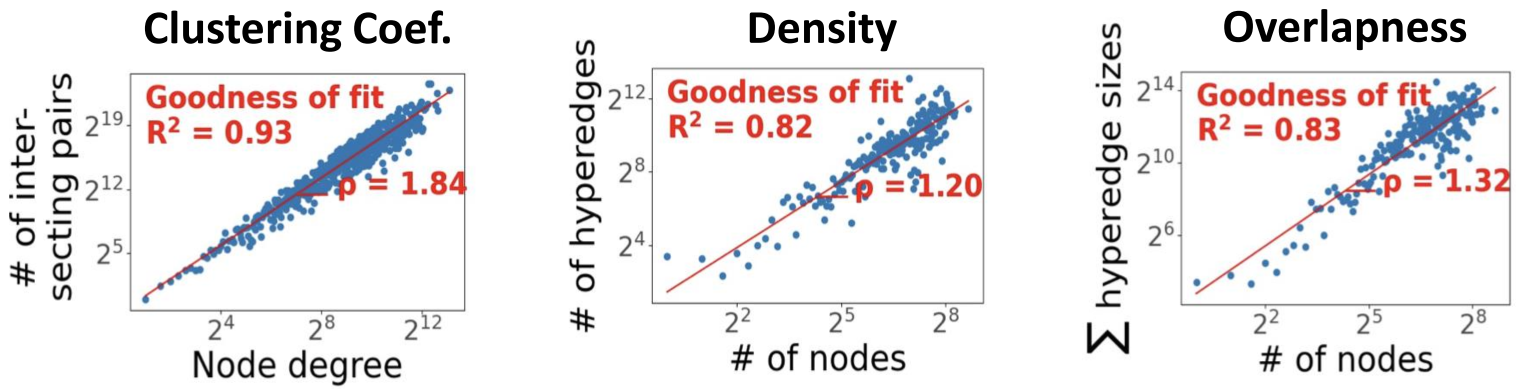


High R^2 scores validate log-logistic distribution

Discovery (D3) - Additional Power-Law Patterns

Discovery (D3): Clustering coefficients, density, and overlapness in egonets follow power-law patterns

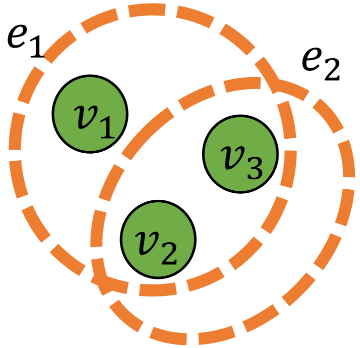
- **Analysis:**
  - **Clustering Coefficients:** Ratio of intersecting hyperedge pairs to node degree
  - **Density:** Ratio of hyperedge count to node count
  - **Overlapness:** Ratio of sum of hyperedge sizes to node count
  - **High R^2 values in log-log regression confirm power-law behavior**



HyRec: A Generative Model for Real-World Hypergraphs

- **HyRec** generates hypergraphs using the Kronecker product of incidence matrices
- *Given an **initiator hypergraph G** with incidence matrix I(G) and **order K**,*
- ***HyRec(G, K)** generates a hypergraph with an incidence matrix $\boldsymbol{I}(\boldsymbol{G})^{[\boldsymbol{K}]}$,*

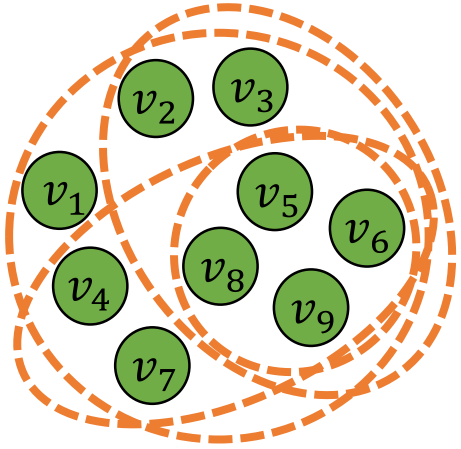
  representing the K-th Kronecker power of $\boldsymbol{I}(\boldsymbol{G})$



	e_1	e_2
v_1	1	
v_2	1	1
v_3	1	1

Initiator Hypergraph (G) and Corresponding Incidence Matrix ($\boldsymbol{I}(\boldsymbol{G})$)

Kronecker Power of Order 2



	e_1	e_2	e_3	e_4
v_1	1			
v_2	1	1		
v_3	1	1		
v_4	1		1	
v_5	1	1	1	1
v_6	1	1	1	1
v_7	1		1	
v_8	1	1	1	1
v_9	1	1	1	1

Theoretical Properties of HyRec

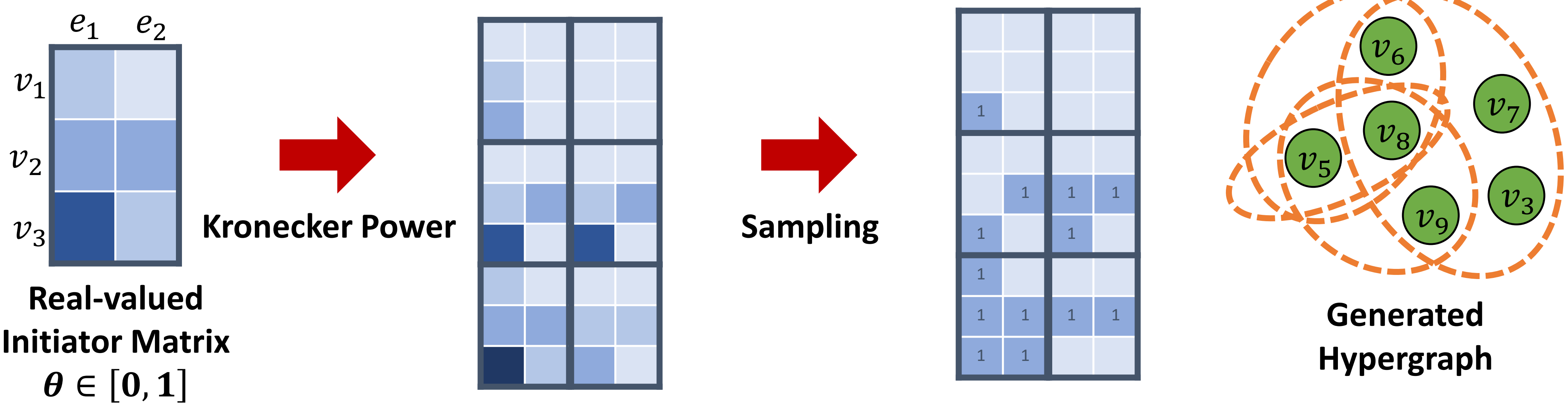
[Theorem: Structural Patterns] HyRec generates hypergraphs where key statistics follow **multinomial distributions***: (1) degrees, (2) hyperedge sizes, (3) pair degrees, (4) intersection sizes, and (5) singular values

** **Multinomial distributions** can resemble power-law and log-logistic distributions found in real-world hypergraphs*

[Theorem: Evolutionary Patterns] HyRec models hypergraph growth by simulating changes in (1) density and (2) (effective) diameter as the Kronecker power exponent K increases, reflecting the evolution of hypergraphs over time.

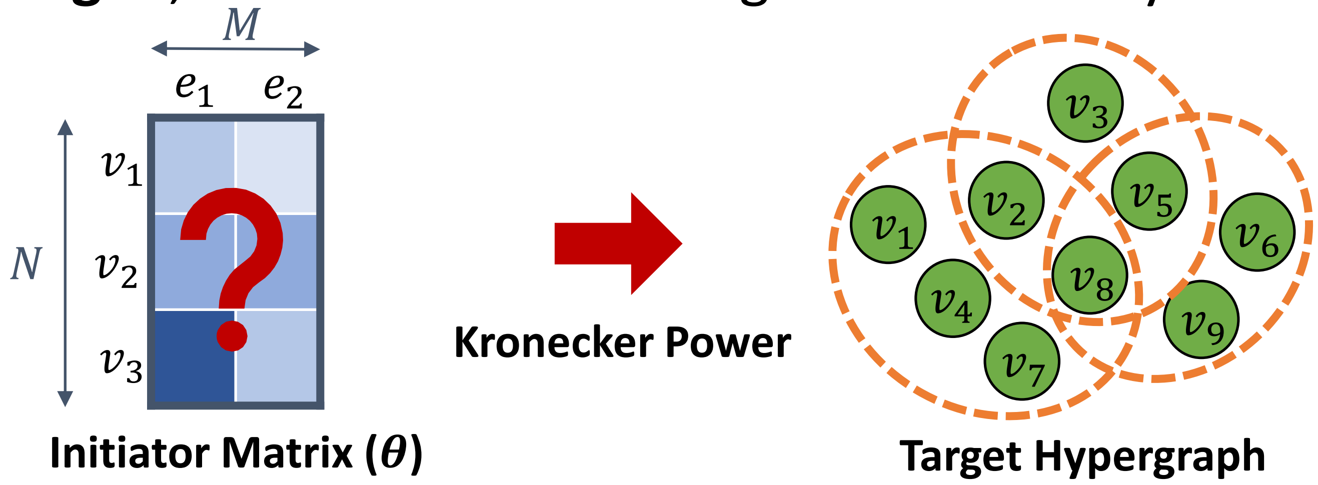
Stochastic Version of HyRec

- The Kronecker power of **binary** initiators always produces the same structure
- We adopt a **stochastic** version with **real-valued** initiators and Bernoulli sampling
- This introduces **variability**
- The following slides will focus on the **stochastic version of HyRec**



SingFit: Fitting HyRec to Real-World Hypergraphs

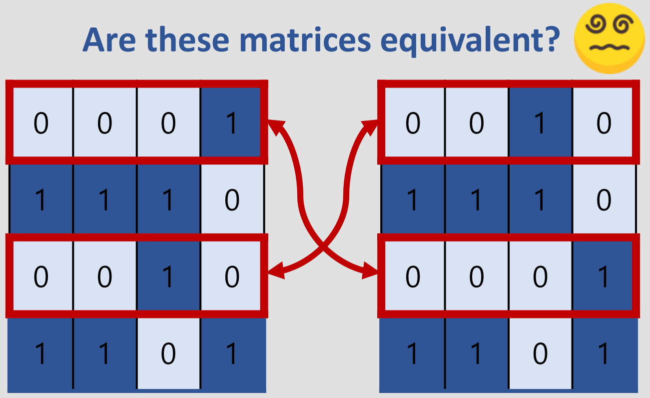
- **Problem**: How can we generate a Kronecker hypergraph that closely matches a given real-world hypergraph?
- **Goal**: **Find the initiator matrix ($\boldsymbol{\theta} \in [\mathbf{0}, \mathbf{1}]^{\boldsymbol{N \times M}}$)** that best captures **key properties** of the target hypergraph
- We develop **SingFit**, a fast and efficient algorithm to fit HyRec to real-world data



SingFit - Key Challenges in Fitting HyRec

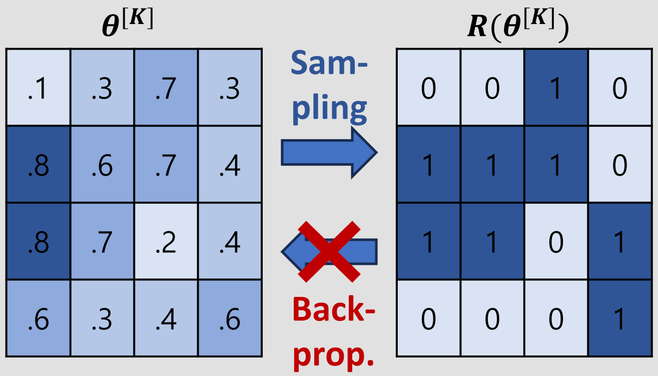
- **Fitting an initiator matrix (θ) requires overcoming three key challenges:**

C1. Computational Cost of Alignment



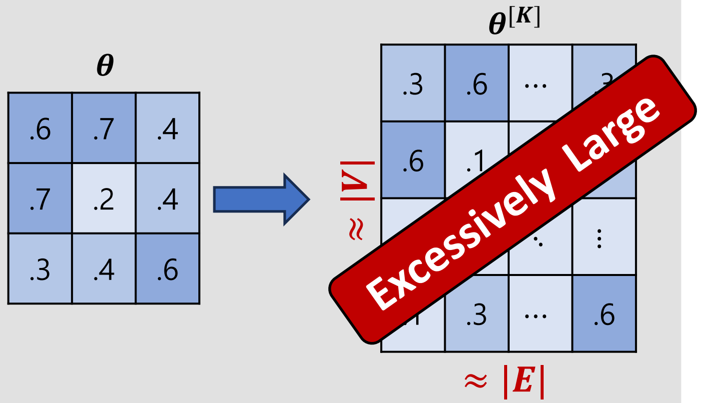
Incidence matrices remain **equivalent under row/column permutations, making alignment** with the target matrix require **an exhaustive search.**

C2. Non-differentiability of Generation



The **sampling process is non-differentiable**, preventing gradient-based optimization for parameter fitting

C3. Density of Probability Matrix

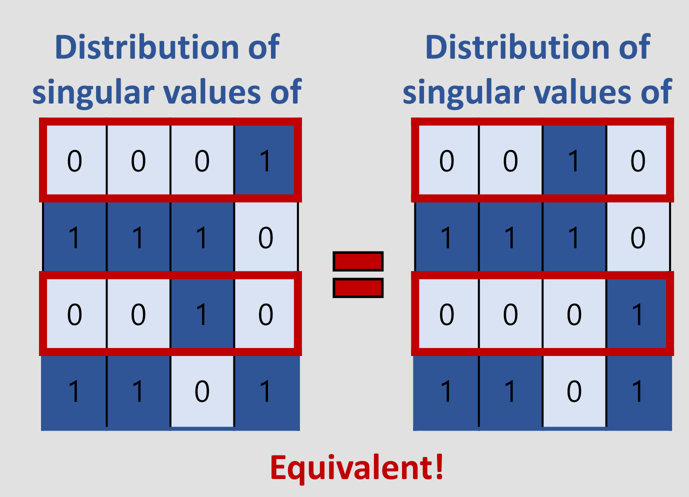


Generating the hypergraph involves **handling all node-hyperedge connections**, leading to **high computational and memory overhead**

SingFit - Overcoming Fitting Challenges

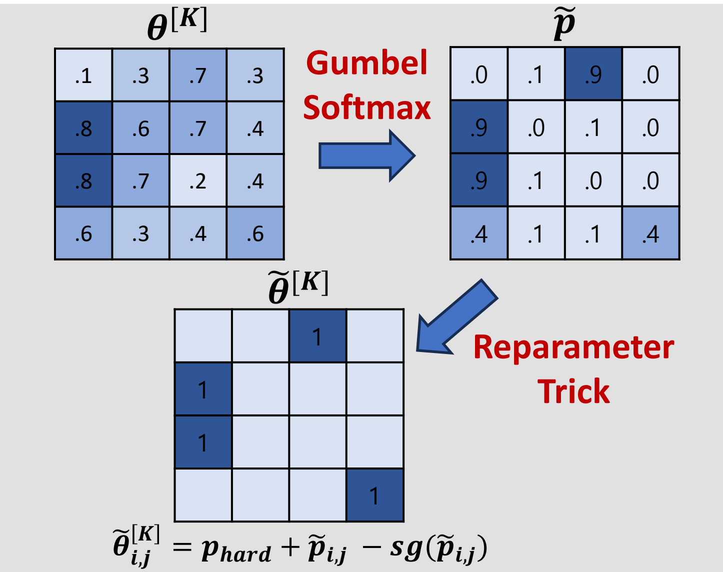
- **SingFit tackles key challenges with three strategies:**

S1. Singular Value Matching



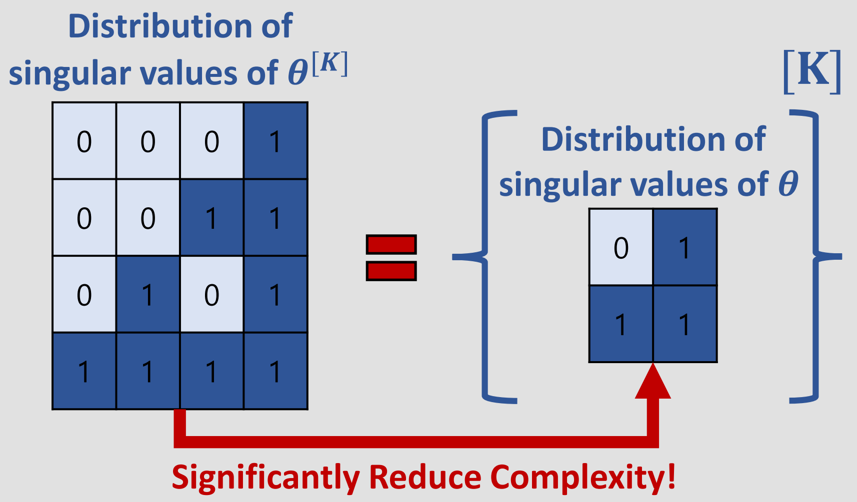
Aligns singular values, which remain invariant under row/column permutations, avoiding high alignment costs (C1).

S2. Differentiable Sampling



Uses **Gumbel Softmax Trick** to bridge the gap between the probability matrix and binary incidence matrix (C2).

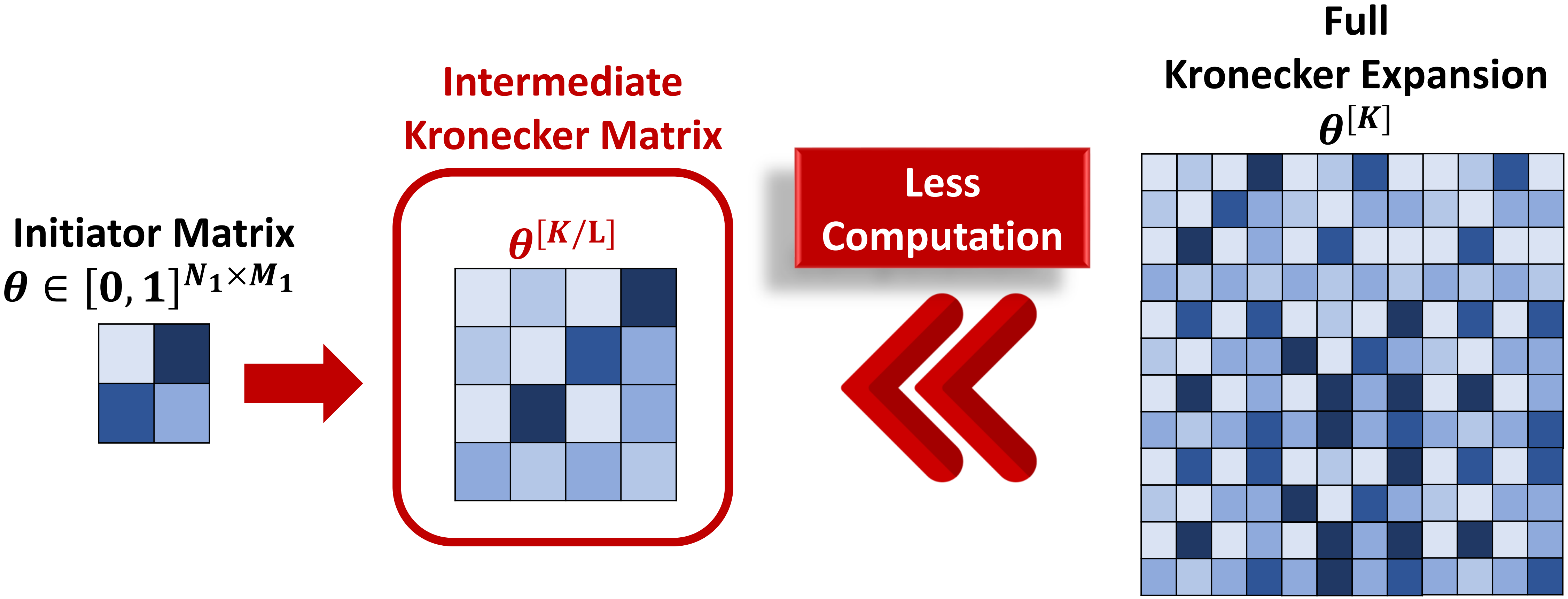
S3. Acceleration with Kronecker Product Properties



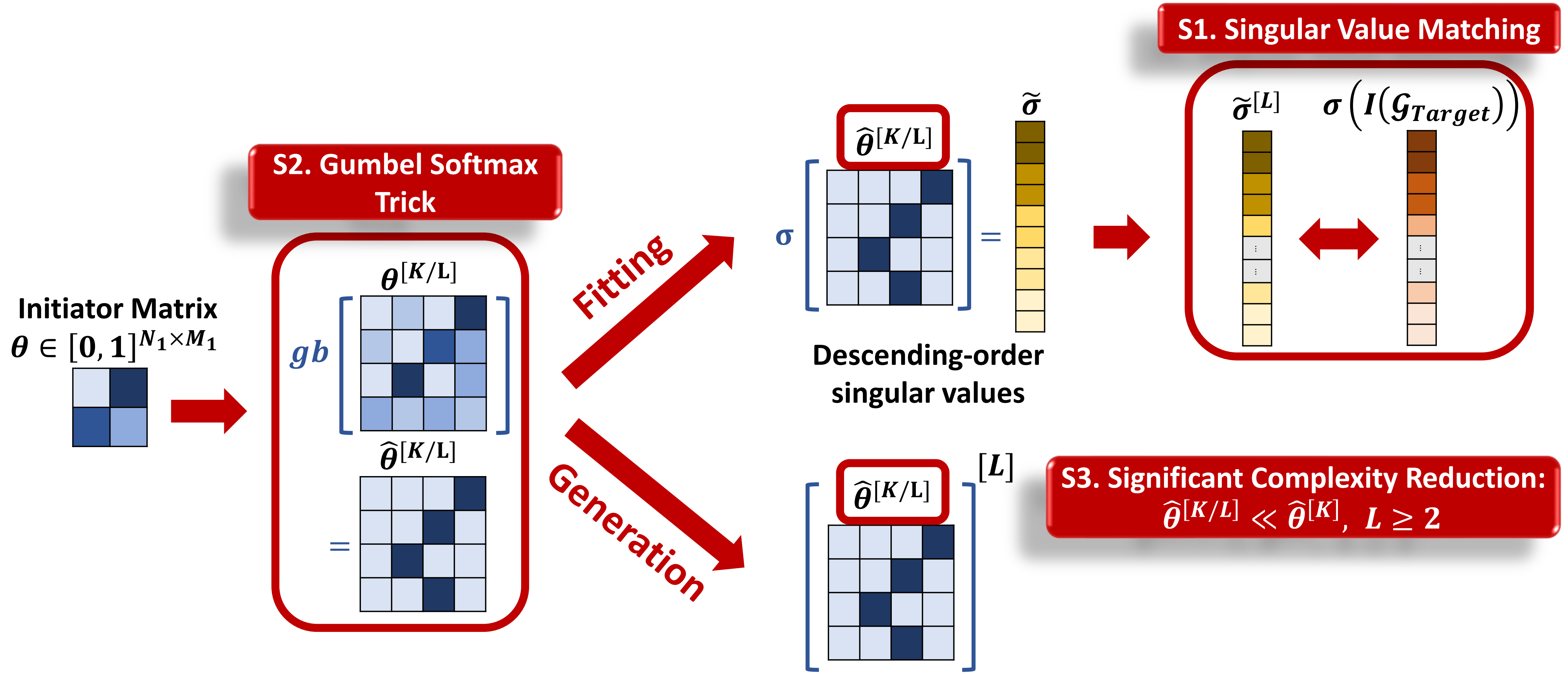
Kronecker product properties enable computing large-matrix properties from smaller matrices, significantly reducing computational overhead (C3).

SingFit: Description of Fitting and Generation

- **Avoids full Kronecker expansion and uses a smaller intermediate matrix**

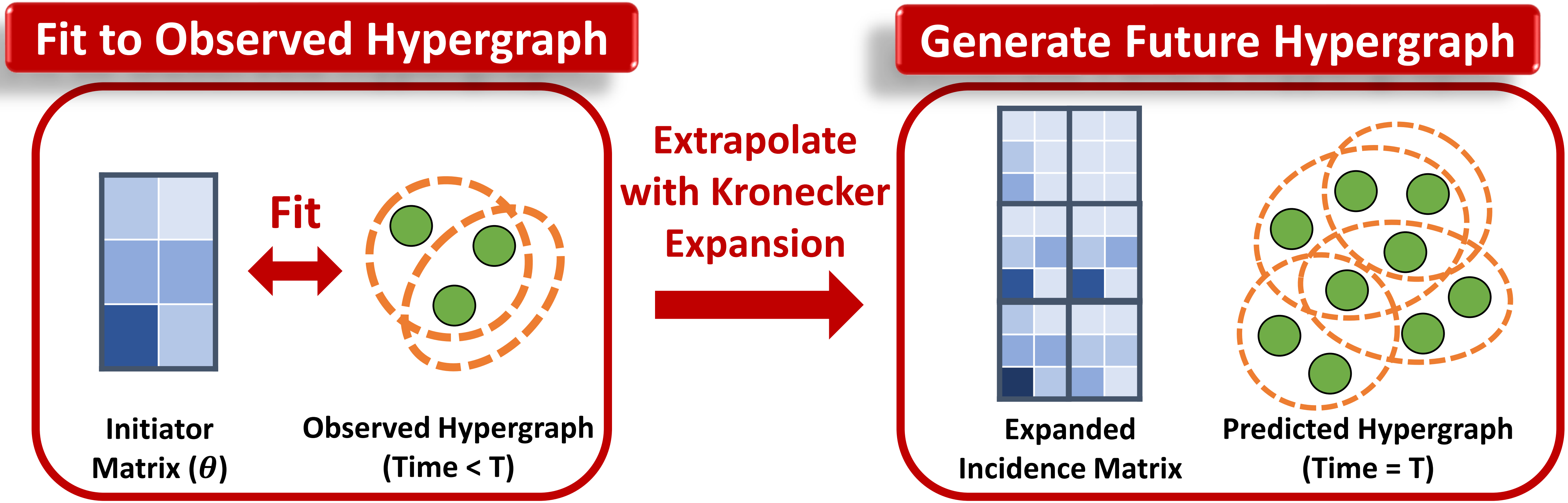


SingFit: Description of Fitting and Generation



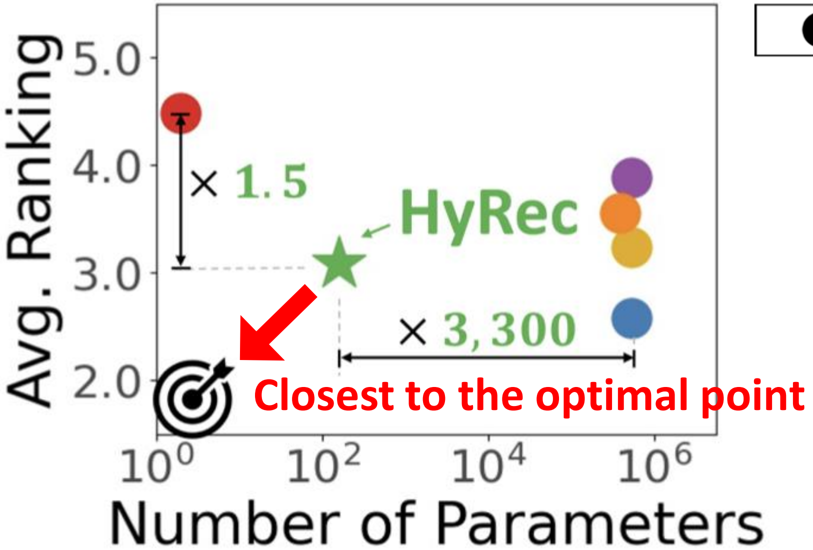
Application: Hypergraph Extrapolation using HyRec

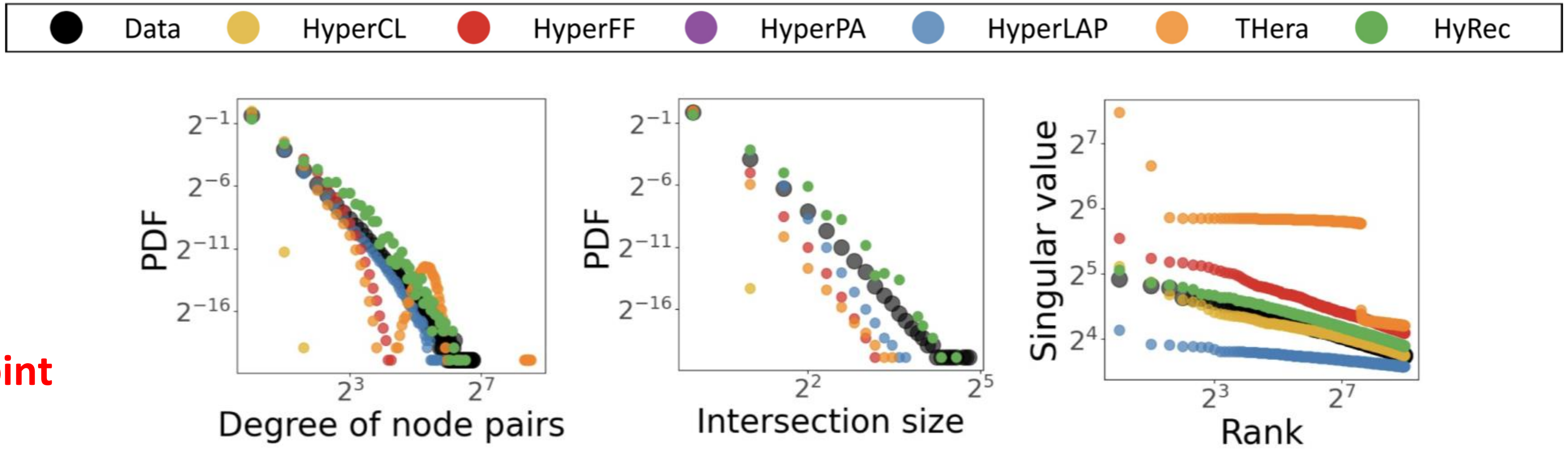
- **Extrapolation** is achieved by expanding the fitted HyRec model with Kronecker power



Experimental Results: Fitting to Real-World Hypergraphs

- **HyRec models real-world hypergraphs accurately and efficiently**
- It reproduces 9 hypergraph properties with a small number of parameters

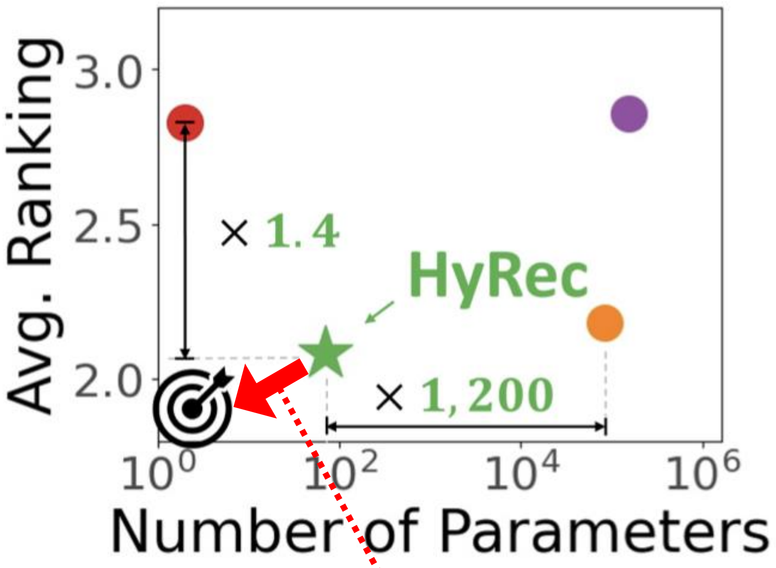




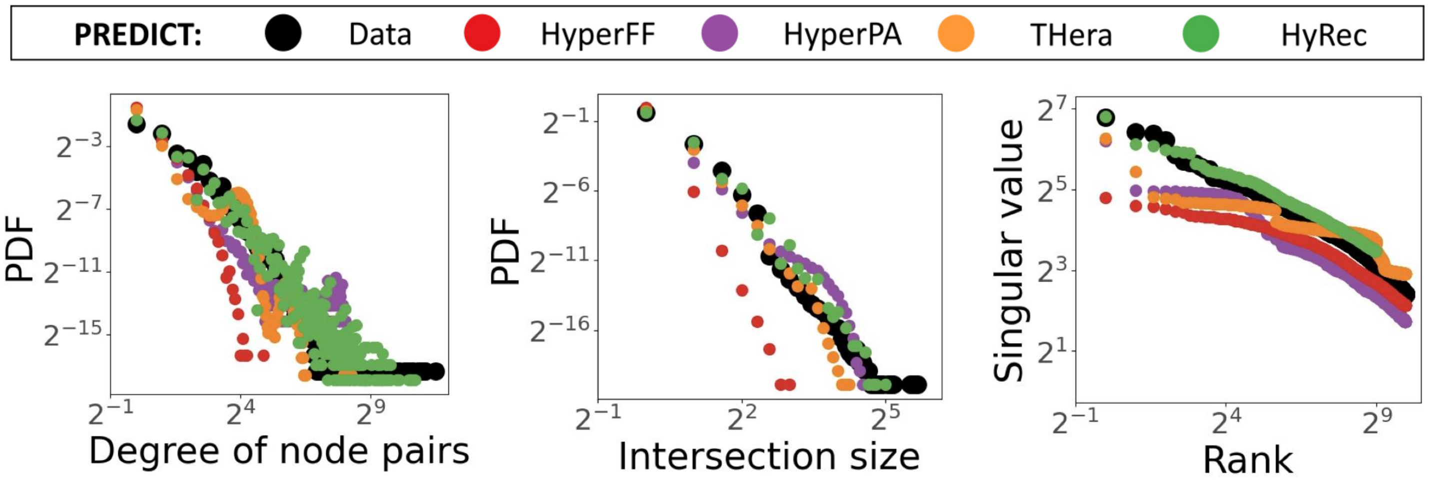
HyRec achieves the strong alignment with real-world data

Experimental Results: Strong Extrapolation

- **HyRec accurately extrapolates the future structure of real hypergraphs**
- It predicts 9 hypergraph properties, outperforming all baselines with minimal input



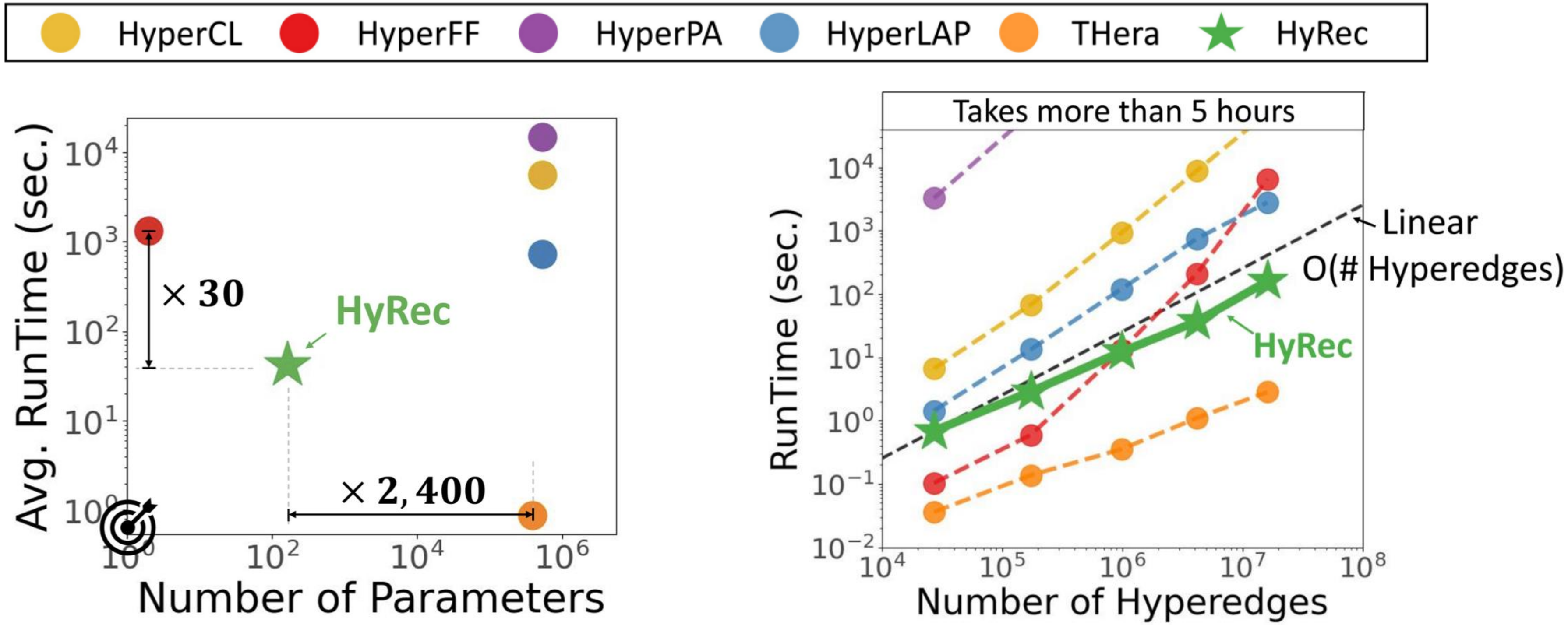
Closest to the optimal point



HyRec provides the closest forecast to real-world hypergraph data

Experimental Results: Efficient Fitting and Generation

- **SingFit scales to large hypergraphs with near-linear runtime**
- It enables efficient fitting and generation through Kronecker expansion



Conclusions

- Our contributions are summarized as follows:

✓ **Discoveries:** Identification of eight power-law-related patterns in eleven real-world hypergraphs

✓ **Model:** Design of HyRec, a tractable and realistic generative model of hypergraphs, supported by SingFit

✓ **Proofs:** Mathematical validation that HyRec adheres to these identified patterns

Code & Dataset: https://github.com/young917/HyRec

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

Code & Dataset: https://github.com/young917/HyRec

Minyoung Choe

Jihoon Ko

Taehyung Kwon

Kijung Shin

Christos Faloutsos