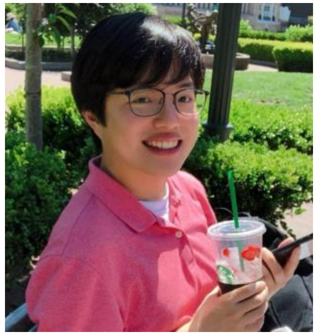
KAIST AI
Kim Jaechul Graduate School

ICDM 2022

Reciprocity in Directed Hypergraphs: Measures, Findings, and Generators



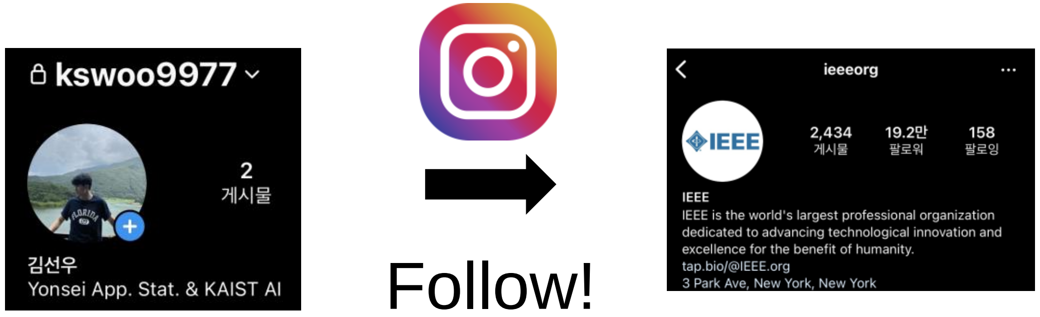
Sunwoo Kim | Minyoung Choe | Jaemin Yoo | Kijung Shin

Roadmap

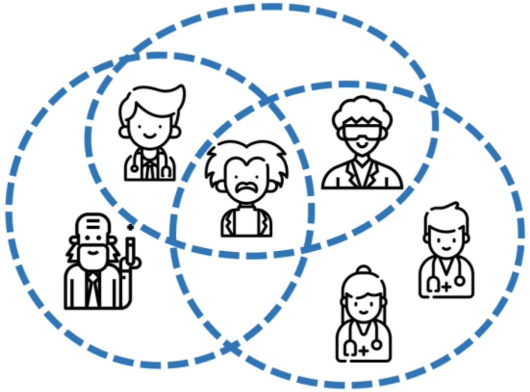
- Overview
- Reciprocity Measure
- Observations
- Generators
- Conclusions

What is Directed Hypergraph?

- In the real world, there are…
  - Directional relations: “Follow” in a social media
  - Group relations: Academic collaborations



Social Media

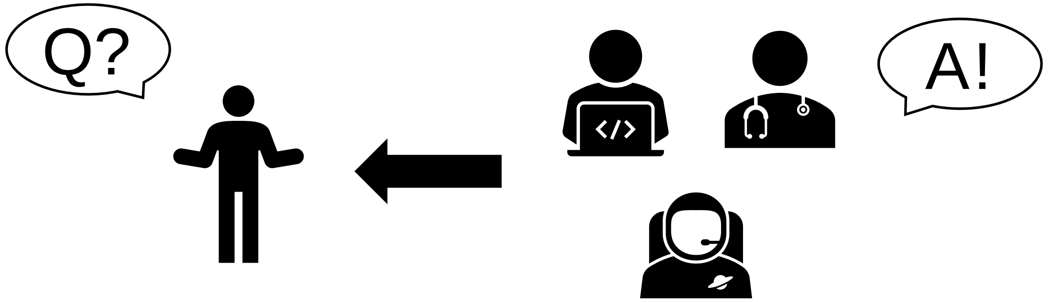


Collaborations of Researchers

What is Directed Hypergraph?

- Directional relations + Group relations = Directional group relations
  - Chemical reactions
  - Question & Answering

$$\begin{array}{ccc} CH_4 & & CO_2 \\ + & \longrightarrow & + \\ 2O_2 & & 2H_2O \end{array}$$



stackoverflow askubuntu

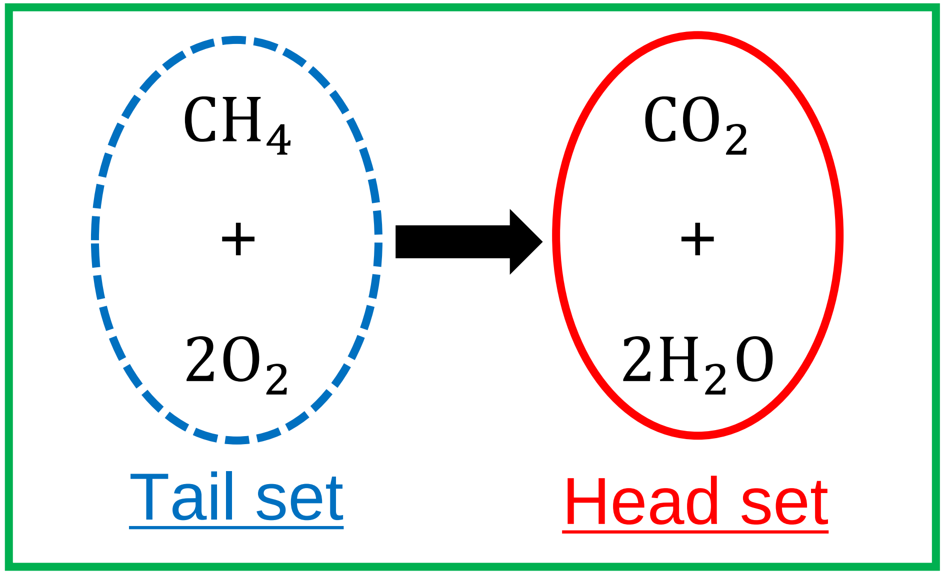
Chemical Reaction

Question & Answering

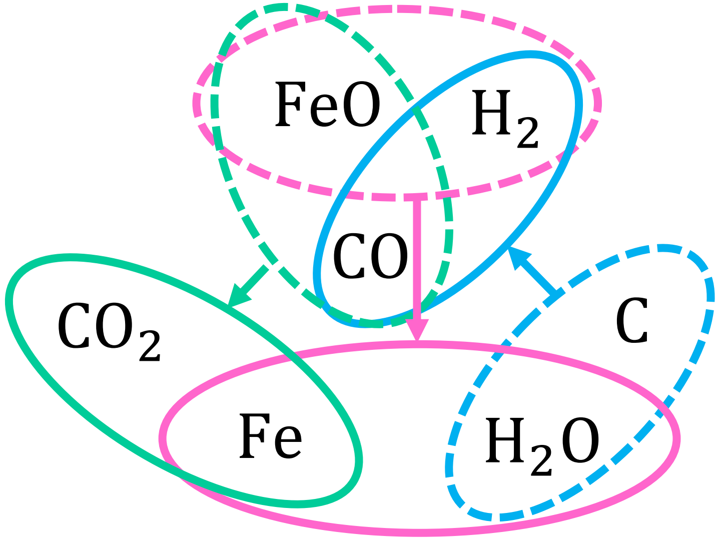
Modeling with Directed Hypergraphs!

What is Directed Hypergraph?

- Directed hypergraph consists of nodes and hyperarcs.
- Each hyperarc consists of two node sets: Head set and Tail set.
- Head set: Destination & Tail set: Source



Hyperarc

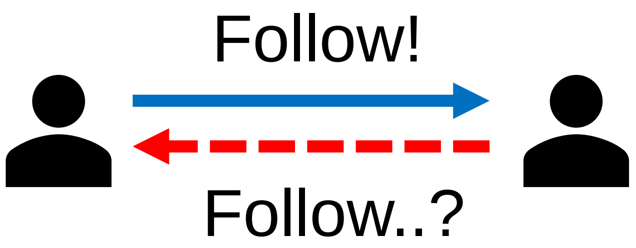


Directed Hypergraph

What is Reciprocity?

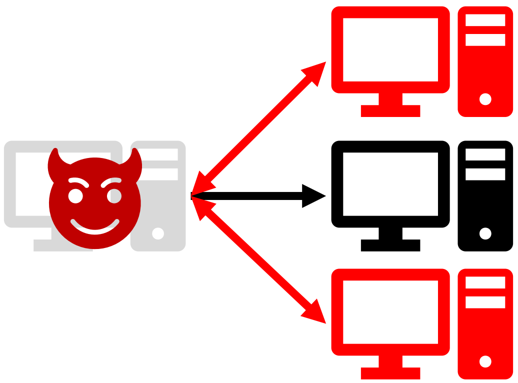
- How mutually are nodes linked?

In a graph: $\frac{\text{\# of Mutually Connected Edges}}{\text{\# of Edges}}$

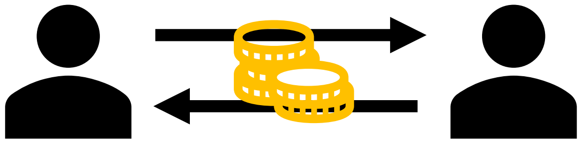


- Useful in understanding human interaction, computer virus spread model, and behavior (trust) prediction.

Social network



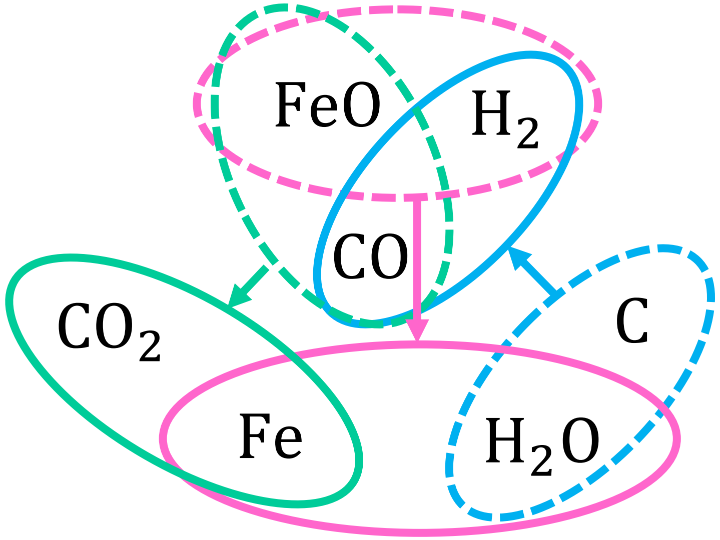
Virus spread

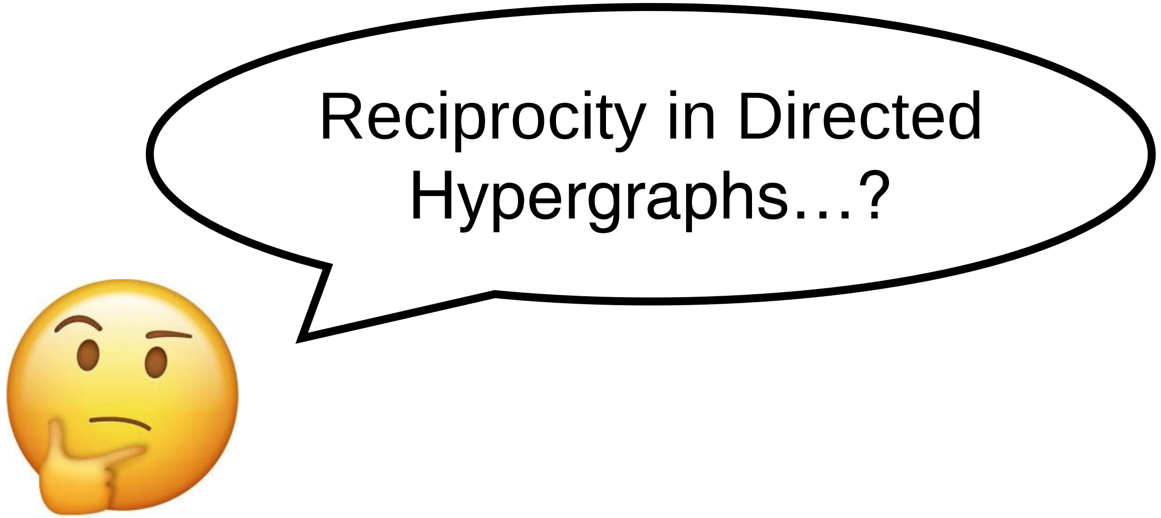


Behavior prediction

Our Research Question

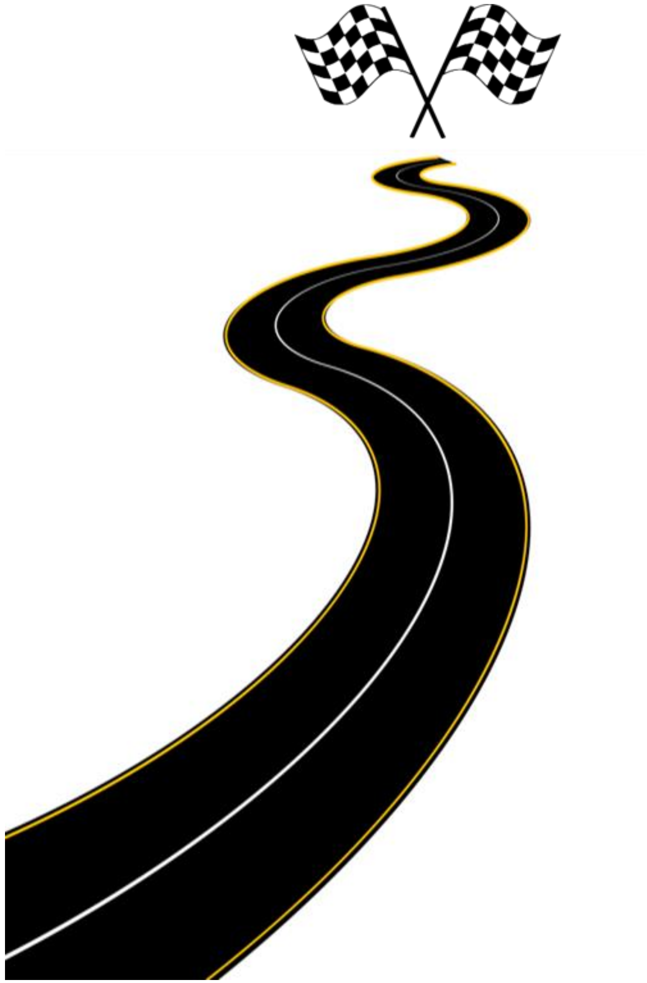
- RQ1) How can we measure reciprocity in a directed hypergraph?
- RQ2) What are the real-world hypergraphs' reciprocal patterns?
- RQ3) Can we generate real-world hypergraphs' reciprocal patterns?





Roadmap

- Overview
- Reciprocity Measure
- Observations
- Generators
- Conclusions

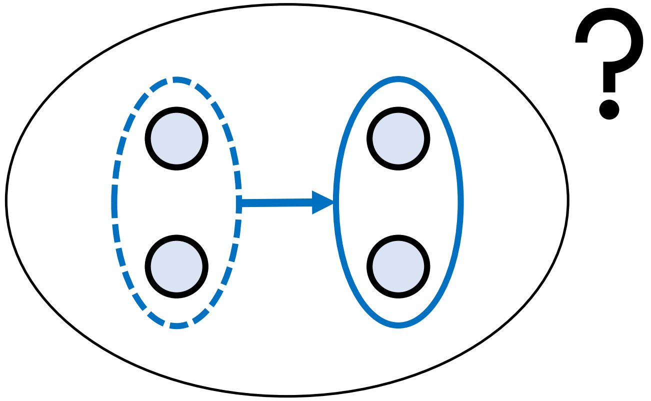


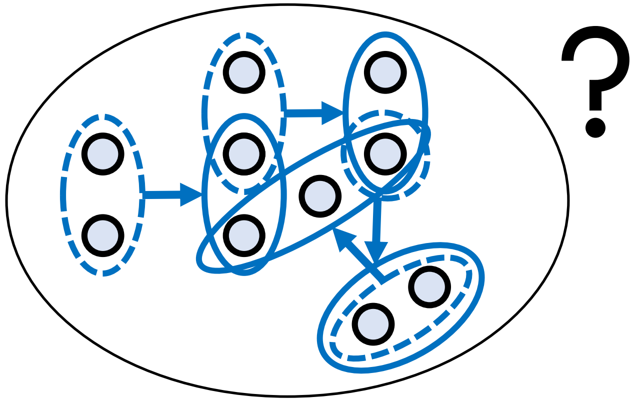
Overall Framework

AXIOM (8 axioms):
8 properties that a proper reciprocity measure should have.

HyperRec:
A hypergraph reciprocity measure that satisfies all the proposed axioms.

- We provide reciprocity measures for a hyperarc and a hypergraph.

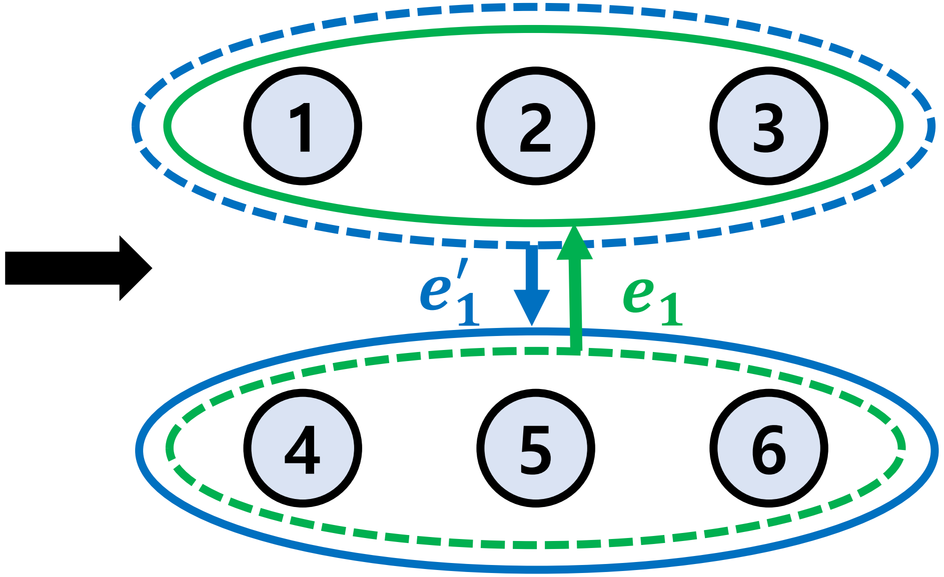


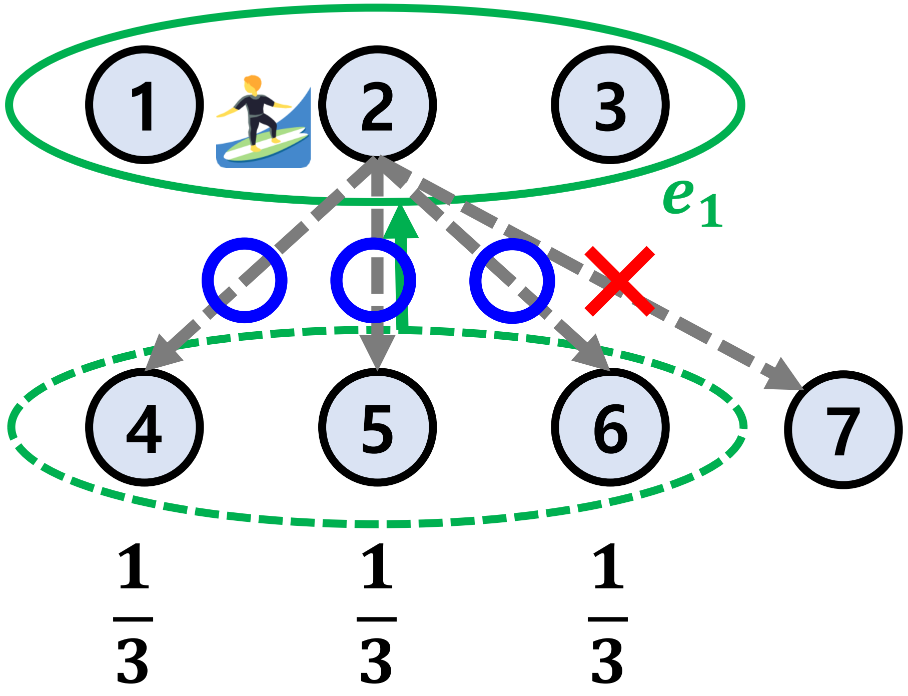


HyperRec: Proposed Reciprocity Measure

- HyperRec is based on the **random surfer's** return probability.
- Consider measuring the reciprocity of a specific hyperarc $\boldsymbol{e_1}$.
- Ideally, it should return to each of its sources with prob. 1/(tail set size).

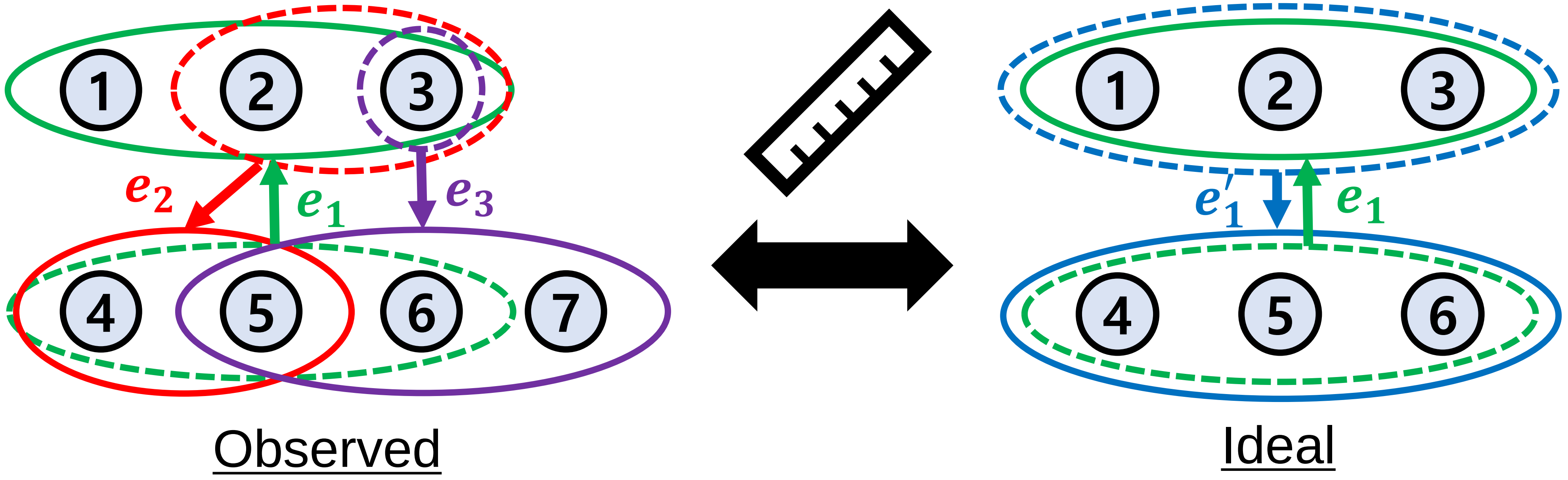
Optimally reciprocal hyperarc $\boldsymbol{e_1'}$





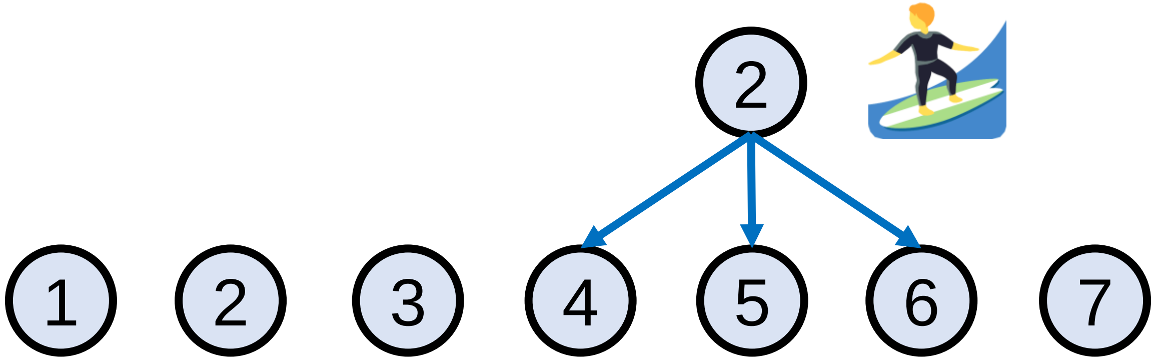
HyperRec: Proposed Reciprocity Measure

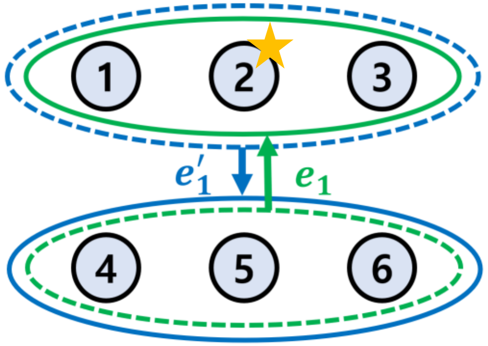
- There are observed hyperarcs (e_2, e_3).
- We use these hyperarcs (Reciprocal set) to measure the reciprocity of e_1.
- HyperRec measures how far observed hyperarcs are from the ideal case.
- Distance: Transition (return) probability distance between them.



HyperRec: Proposed Reciprocity Measure

- Transition (return) probability from node 2 to other nodes.



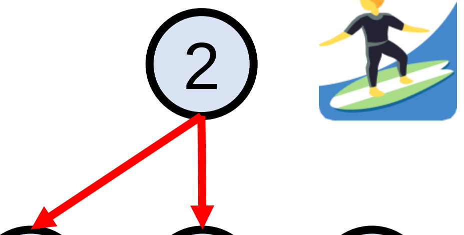


Ideal

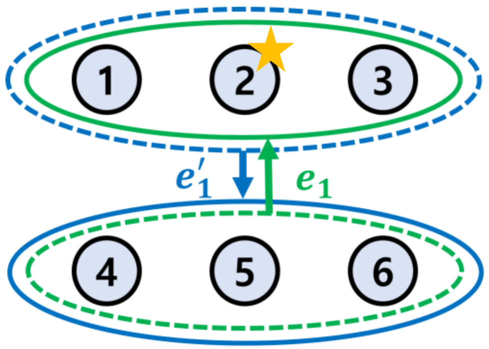
Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0

HyperRec: Proposed Reciprocity Measure

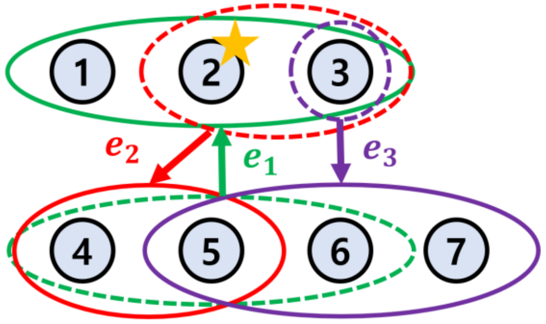
- Transition (return) probability from node 2 to other nodes.



	1	2	3	4	5	6	7
Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
e_2	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0



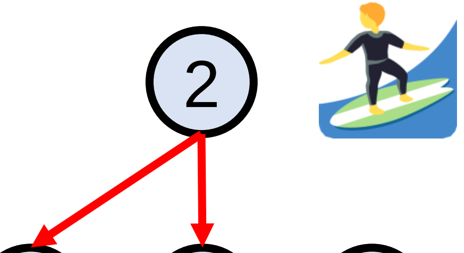
Ideal



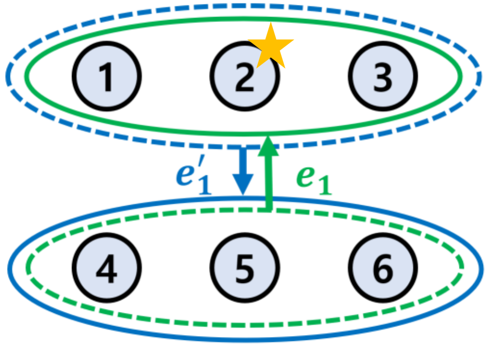
Observed

HyperRec: Proposed Reciprocity Measure

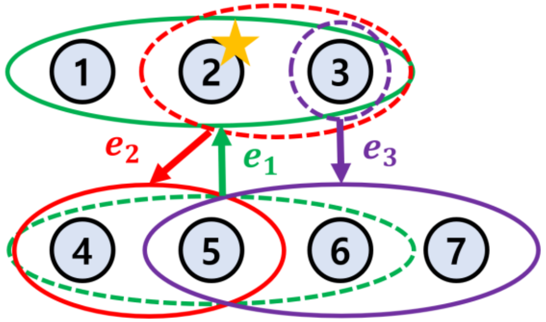
- Transition (return) probability from node 2 to other nodes.



Probability Distance		1	2	3	4	5	6	7
	Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
	e_2	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0



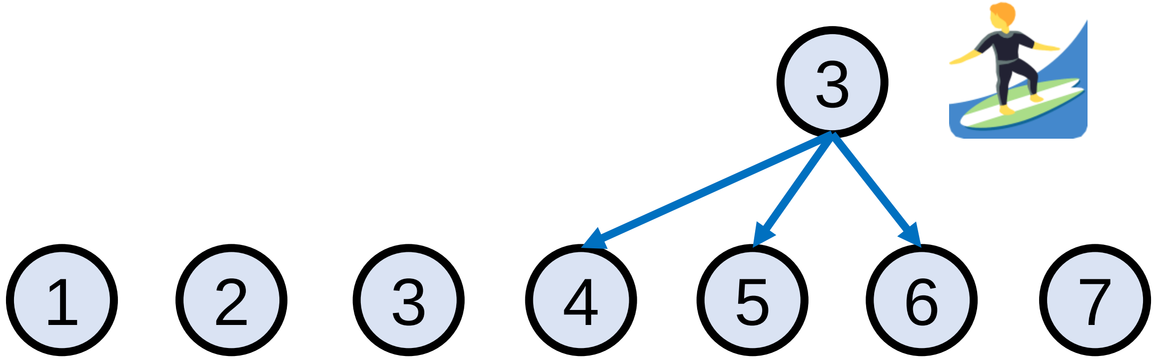
Ideal



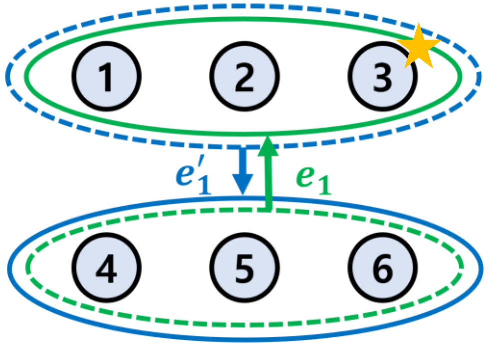
Observed

HyperRec: Proposed Reciprocity Measure

- Transition (return) probability from node 3 to other nodes.



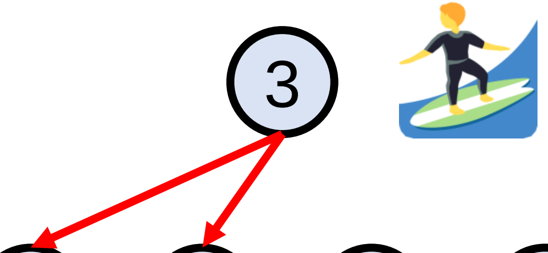
Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0



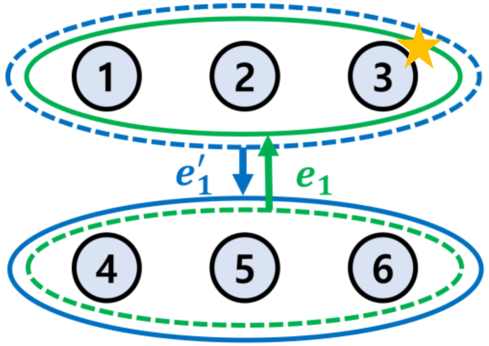
Ideal

HyperRec: Proposed Reciprocity Measure

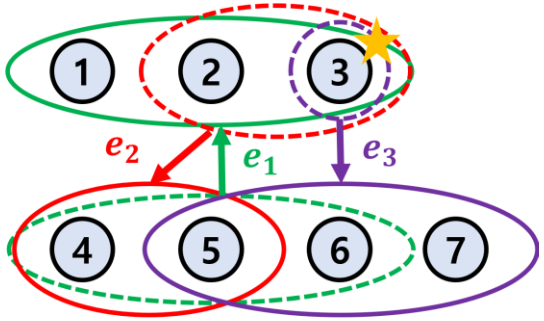
- Transition (return) probability from node 3 to other nodes.



	1	2	3	4	5	6	7
Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
e_2	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0



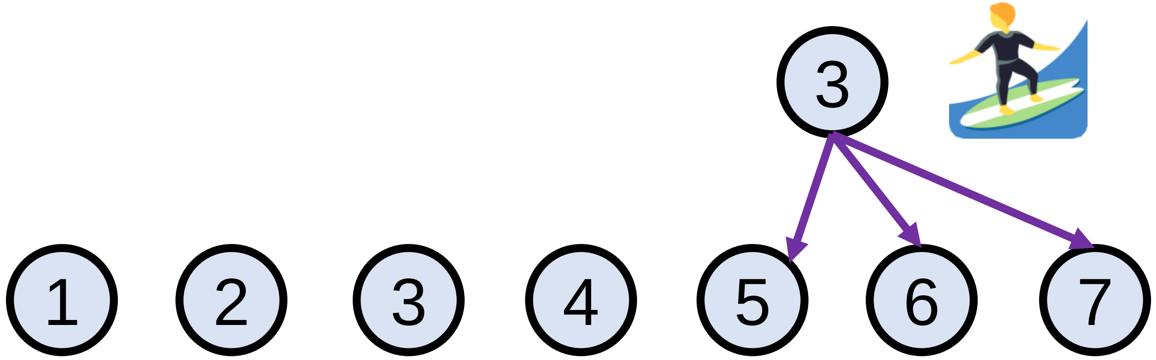
Ideal



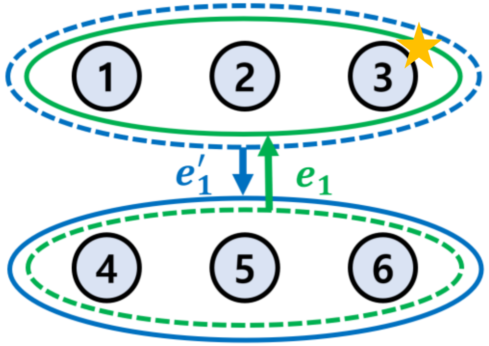
Observed

HyperRec: Proposed Reciprocity Measure

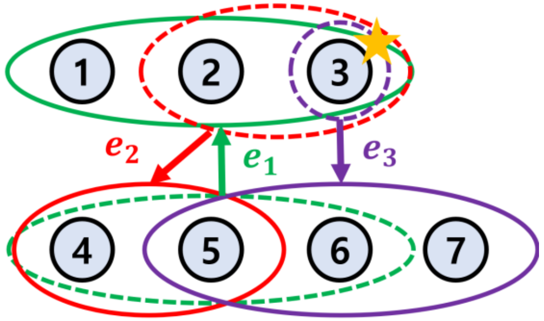
- Transition (return) probability from node 3 to other nodes.



	1	2	3	4	5	6	7
Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
e_3	0	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$



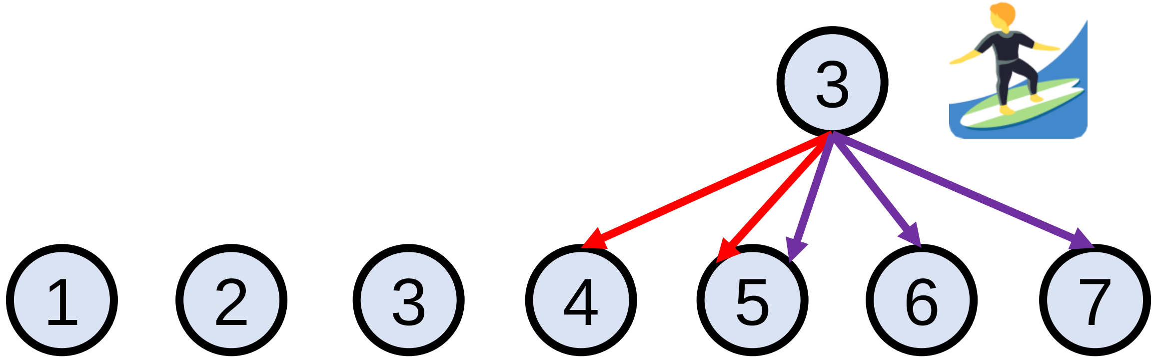
Ideal



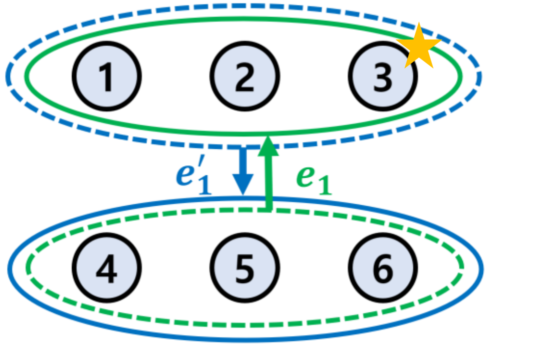
Observed

HyperRec: Proposed Reciprocity Measure

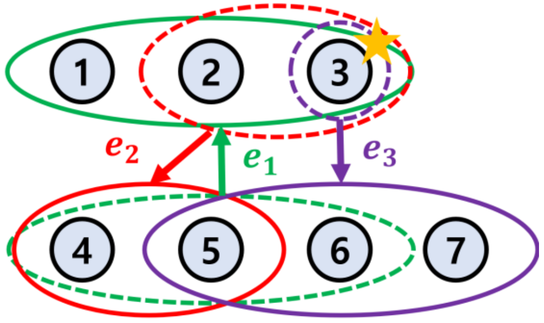
- Transition (return) probability from node 3 to other nodes.



	1	2	3	4	5	6	7
Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
e_2	0	0	0	$\frac{1}{2}\cdot\frac{1}{2}$	$\frac{1}{2}\cdot\frac{1}{2}$	0	0
e_3	0	0	0	0	$\frac{1}{2}\cdot\frac{1}{3}$	$\frac{1}{2}\cdot\frac{1}{3}$	$\frac{1}{2}\cdot\frac{1}{3}$



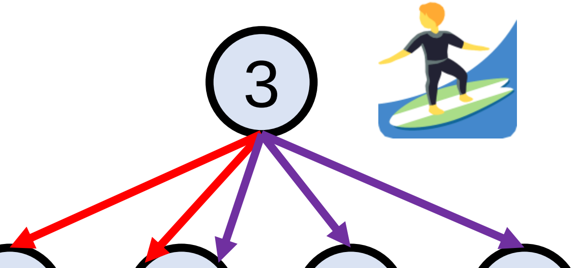
Ideal



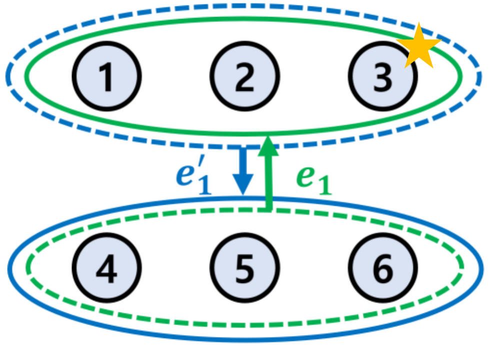
Observed

HyperRec: Proposed Reciprocity Measure

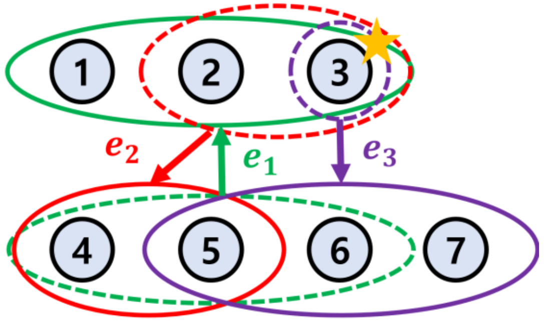
- Transition (return) probability from node 3 to other nodes.



		1	2	3	4	5	6	7
Probability	Ideal	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
Distance	e_2	0	0	0	$\frac{1}{2} \cdot \frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2}$	0	0
	+ e_3	0	0	0	0	$\frac{1}{2} \cdot \frac{1}{3}$	$\frac{1}{2} \cdot \frac{1}{3}$	$\frac{1}{2} \cdot \frac{1}{3}$



Ideal



Observed

HyperRec: Proposed Reciprocity Measure

- Hyperarc level HyperRec

$$r(e_i, R_i) = \left(\frac{1}{|R_i|}\right)^{\alpha} \left(1 - \frac{1}{|H_i|} \sum_{v \in \boxed{H_i}} \frac{\boxed{\mathcal{L}}(\boxed{p_v}, \boxed{p_v^*})}{\mathcal{L}_{max}}\right)$$

Head set of e_i

Probability distance measure (e.g., Jensen Shannon Divergence)

Optimal transition probability distribution

Transition probability distribution

HyperRec: Proposed Reciprocity Measure

Details

- Hyperarc level HyperRec

$$r(e_i) = \boxed{\max_{R_i \subseteq E, R_i \neq \emptyset}} \left(\frac{1}{|R_i|}\right)^{\boxed{\alpha}} \left(1 - \frac{1}{|H_i|} \sum_{v \in H_i} \frac{\mathcal{L}(p_v, p_v^*)}{\mathcal{L}_{max}}\right)$$

Reciprocal set size penalty term

Using a reciprocal set that can maximize the hyperarc reciprocity

- Hypergraph level HyperRec

$$r(G) = \frac{1}{\boxed{|E|}} \sum_{\boxed{e_i \in E}} r(e_i)$$

Average for every hyperarc

Roadmap

- Overview
- Reciprocity Measure
- Observations
- Generators
- Conclusions

Observation

- What are the real-world hypergraphs' reciprocal patterns?

Metabolic (iAF1260b & iJO1366)
- Node: Genes
- Hyperarcs: Reactions

Q&A (Math & Server)
- Node: Users
- Hyperarcs: Posts

Email (Enron & EU)
- Node: Users
- Hyperarcs: Mails

Citation (Data Mining & Software)
- Node: Authors
- Hyperarcs: Citations

Bitcoin (2014, 2015, 2016)
- Node: Users
- Hyperarcs: Transactions

Observation

Observation 1: Hypergraph Level.

Real-world hypergraphs are more reciprocal than randomized hypergraphs.

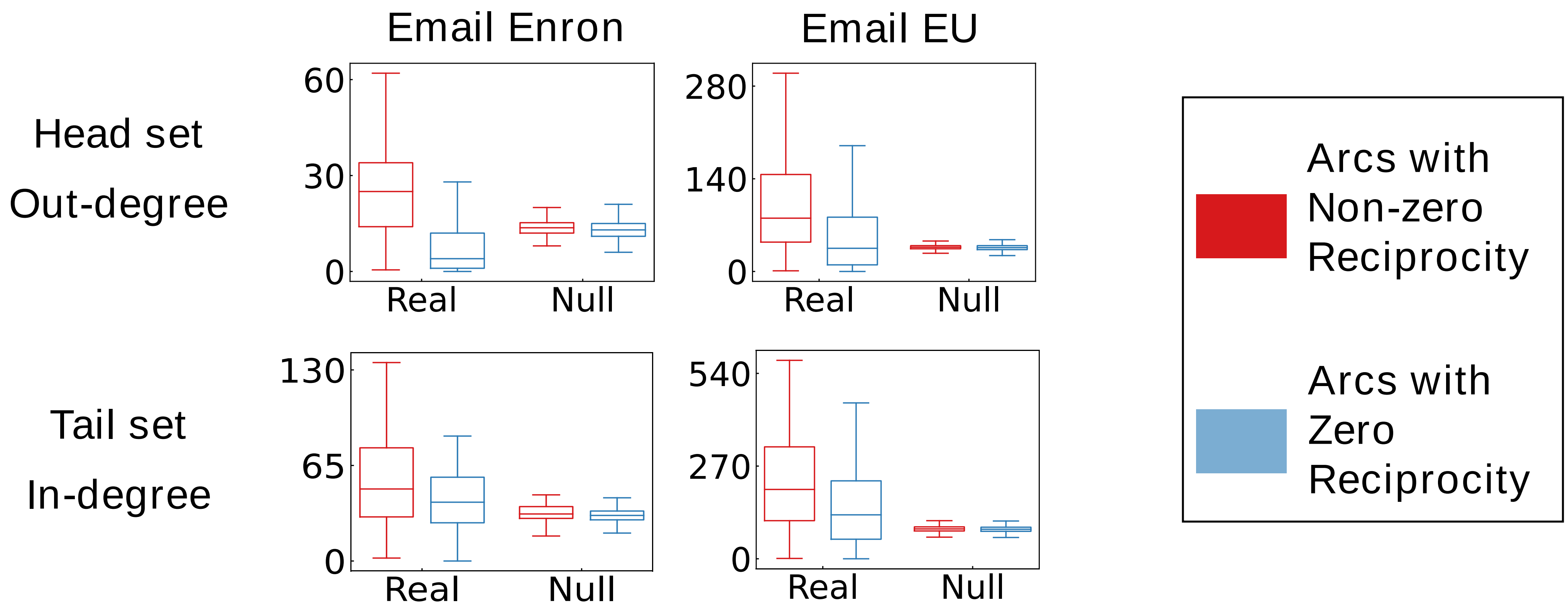
		metabolic		email		citation		q&a		bitcoin		
		iAF1260b	iJO1366	enron	eu	data mining	software	math	server	2014	2015	2016
Real World	$r(G)$	21.455	22.533	59.001	79.416	12.078	15.316	9.608	13.219	10.829	6.923	3.045
Null	$r(G)$	0.306	0.270	14.862	4.633	0.094	0.147	0.018	0.002	0.0001	0.000*	0.000*
	D-Stat	0.625	0.642	0.539	0.807	0.355	0.377	0.124	0.160	0.147	0.100	0.050

** Null hypergraph is a randomized hypergraph that fills nodes uniformly at random where hyperarc size is preserved.

Observation

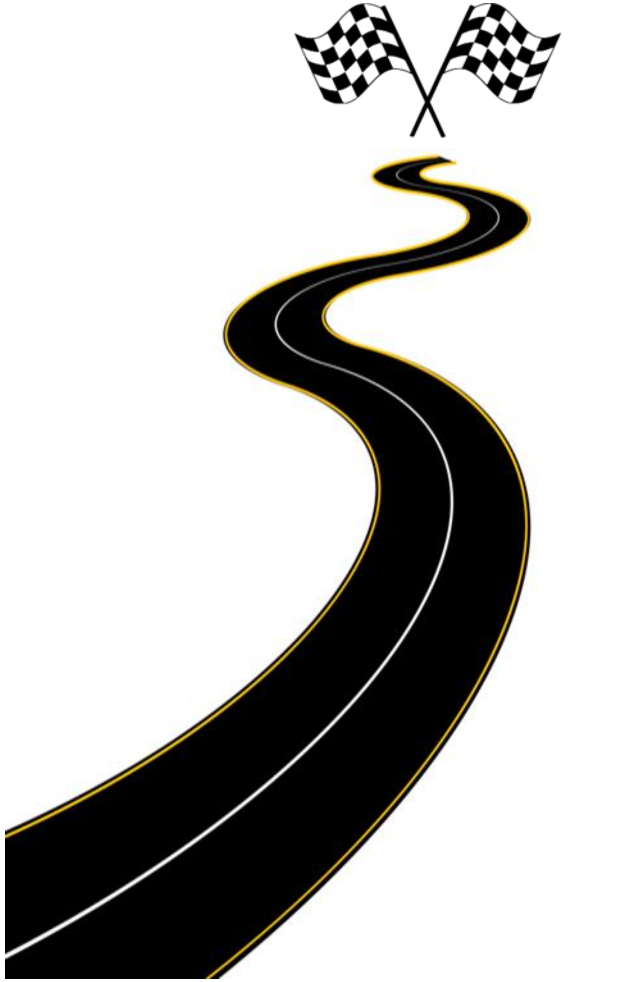
Observation 2: Hyperarc Level.

Arcs with non-zero reciprocity tend to have higher head set out-degree and tail set in-degree.



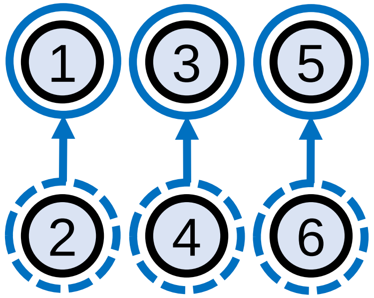
Roadmap

- Overview
- Reciprocity Measure
- Observations
- Generators
- Conclusions



ReDi: Realistic Directed Hypergraph Generator

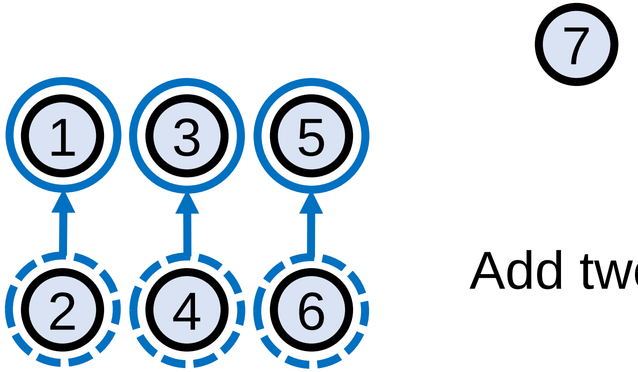
- Realistic ≒ Can preserve reciprocal patterns.
- ReDi iteratively add nodes and create hyperarcs ∝ group degree.
- Extent of reciprocity is being controlled by hyperparameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$.



0. Initialize with normal graph edges.
1. Add one node and decide the number of hyperarcs to be added with node 7.
2. Choose new hyperarcs' size.

ReDi: Realistic Directed Hypergraph Generator

- Realistic ≒ Can preserve reciprocal patterns.
- ReDi iteratively add nodes and create hyperarcs ∝ group degree.
- Extent of reciprocity is being controlled by hyperparameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$.

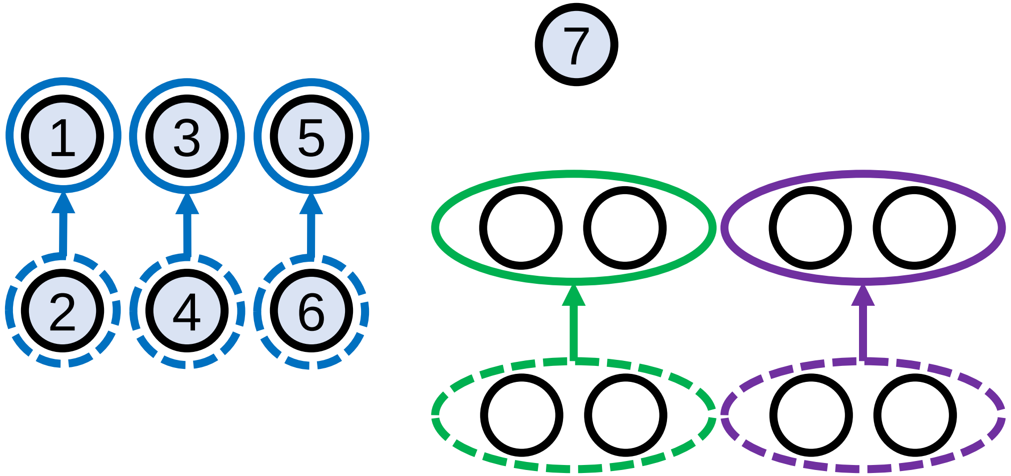


Add two hyperarcs!

0. Initialize with normal graph edges.
1. Add one node and decide the number of hyperarcs to be added with node 7.
2. Choose new hyperarcs’ size.

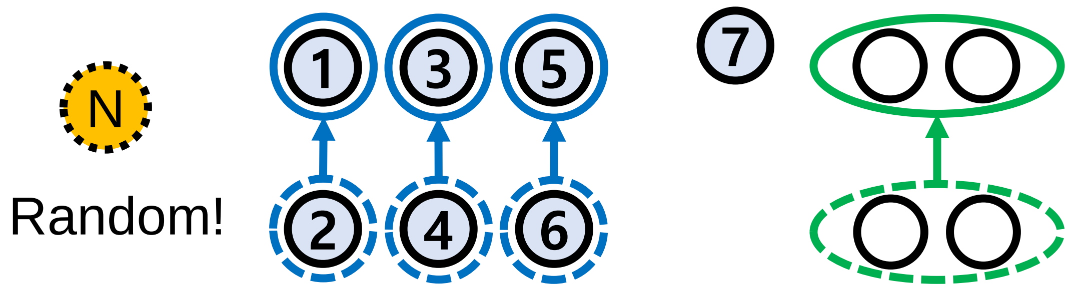
ReDi: Realistic Directed Hypergraph Generator

- Realistic ≒ Can preserve reciprocal patterns.
- ReDi iteratively add nodes and create hyperarcs ∝ group degree.
- Extent of reciprocity is being controlled by hyperparameters $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$.



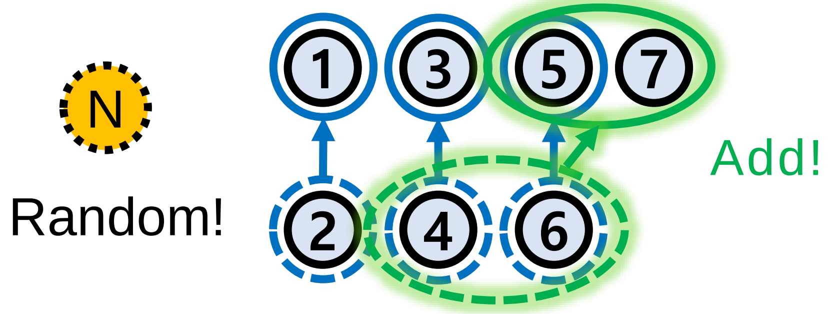
0. Initialize with normal graph edges.
1. Add one node and decide the number of hyperarcs to be added with node 7.
2. Choose new hyperarcs' size.

ReDi: Realistic Directed Hypergraph Generator



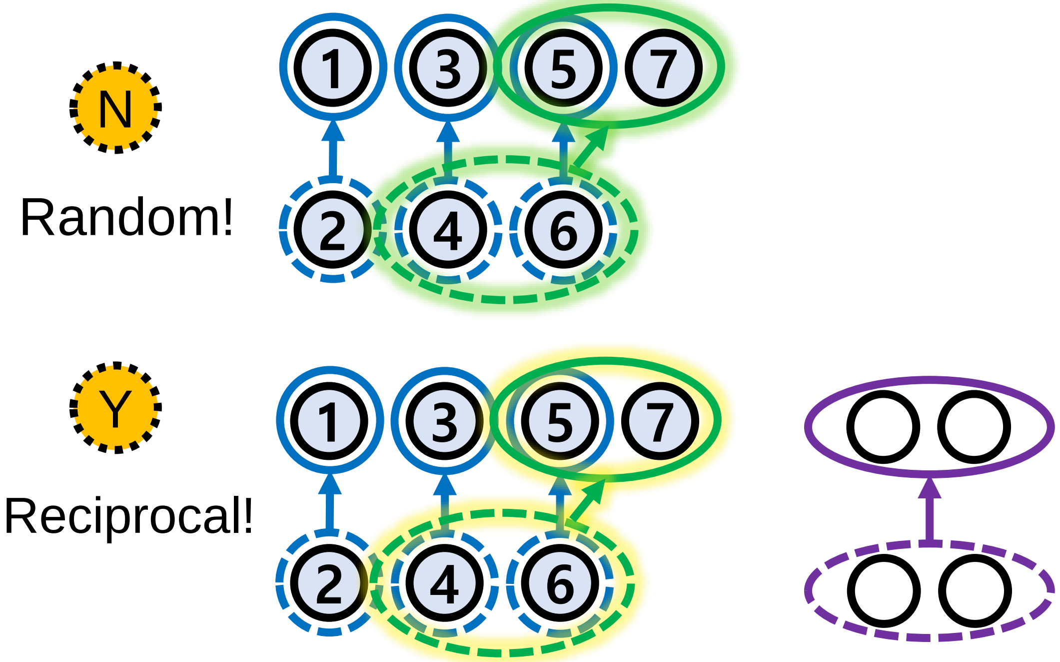
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \boldsymbol{\beta_1}$) or random ($\propto \mathbf{1} - \boldsymbol{\beta_1}$).
4. (Random) Put a new node in either head or tail set, and fill the remaining positions ∝ the group degree. Then add hyperarc.
5. (Reciprocal) Choose a hyperarc to be reciprocal.
6. Toss a coin ($\propto \boldsymbol{\beta_2}$), and decide how many and which nodes will we bring.
7. Fill remaining positions ∝ the group degree, and add hyperarc.

ReDi: Realistic Directed Hypergraph Generator



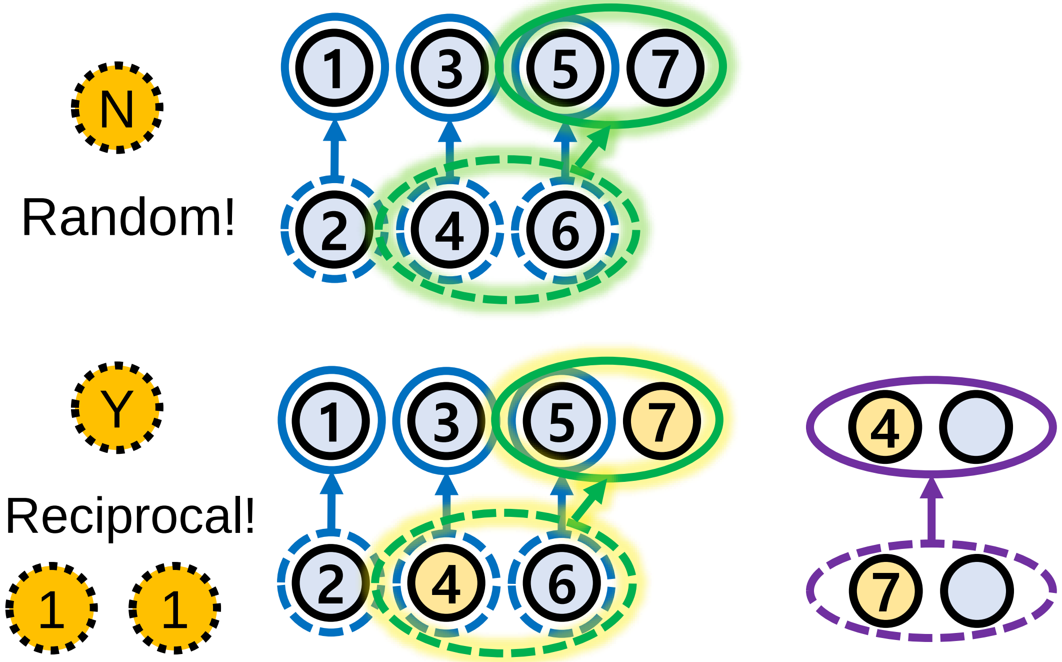
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \boldsymbol{\beta_1}$) or random ($\propto \mathbf{1} - \boldsymbol{\beta_1}$).
4. (Random) Put a new node in either head or tail set, and fill the remaining positions ∝ the group degree. Then add hyperarc.
5. (Reciprocal) Choose a hyperarc to be reciprocal.
6. Toss a coin ($\propto \boldsymbol{\beta_2}$), and decide how many and which nodes will we bring.
7. Fill remaining positions ∝ the group degree, and add hyperarc.

ReDi: Realistic Directed Hypergraph Generator



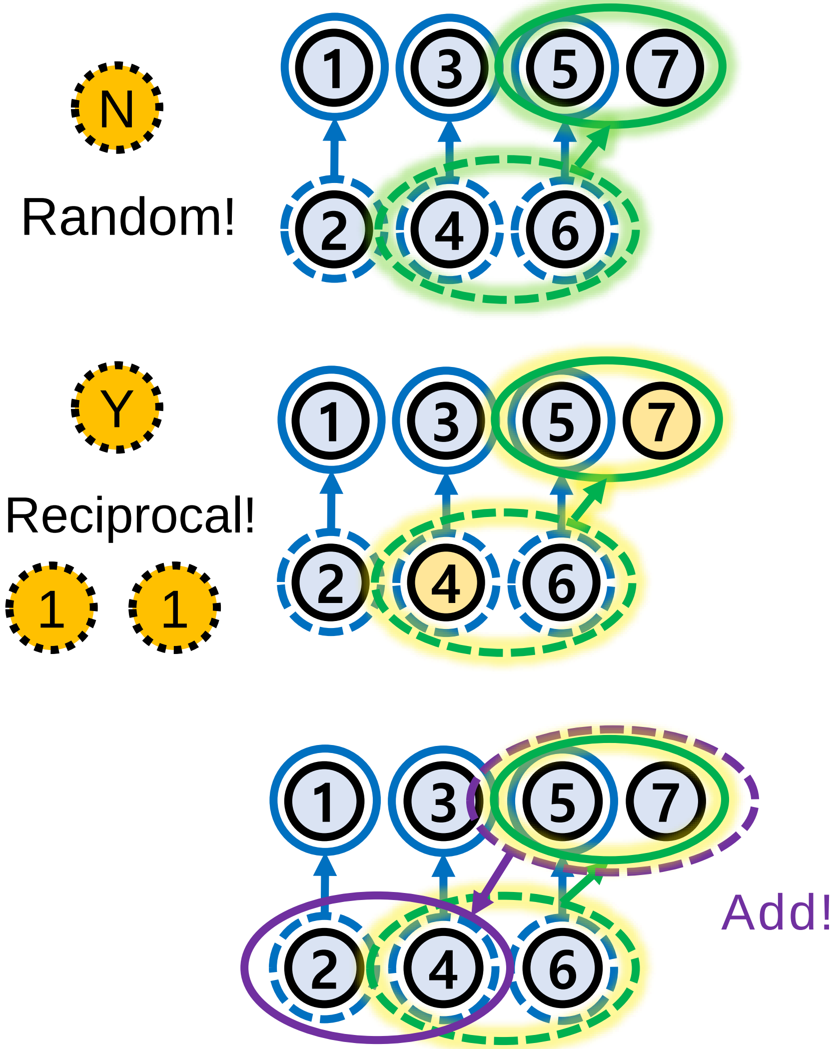
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \boldsymbol{\beta_1}$) or random ($\propto \mathbf{1} - \boldsymbol{\beta_1}$).
4. (Random) Put a new node in either head or tail set, and fill the remaining positions $\propto$ the group degree. Then add hyperarc.
5. (Reciprocal) Choose a hyperarc to be reciprocal.
6. Toss a coin ($\propto \boldsymbol{\beta_2}$), and decide how many and which nodes will we bring.
7. Fill remaining positions $\propto$ the group degree, and add hyperarc.

ReDi: Realistic Directed Hypergraph Generator



3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \boldsymbol{\beta_1}$) or random ($\propto \mathbf{1} - \boldsymbol{\beta_1}$).
4. (Random) Put a new node in either head or tail set, and fill the remaining positions ∝ the group degree. Then add hyperarc.
5. (Reciprocal) Choose a hyperarc to be reciprocal.
6. Toss a coin ($\propto \boldsymbol{\beta_2}$), and decide how many and which nodes will we bring.
7. Fill remaining positions ∝ the group degree, and add hyperarc.

ReDi: Realistic Directed Hypergraph Generator



3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \boldsymbol{\beta_1}$) or random ($\propto \mathbf{1} - \boldsymbol{\beta_1}$).
4. (Random) Put a new node in either head or tail set, and fill the remaining positions ∝ the group degree. Then add hyperarc.
5. (Reciprocal) Choose a hyperarc to be reciprocal.
6. Toss a coin ($\propto \boldsymbol{\beta_2}$), and decide how many and which nodes will we bring.
7. Fill remaining positions ∝ the group degree. Then add hyperarc.

ReDi: Realistic Directed Hypergraph Generator

Reproducibility of observation 1 of ReDi:

ReDi can preserve the reciprocity value of hypergraphs.

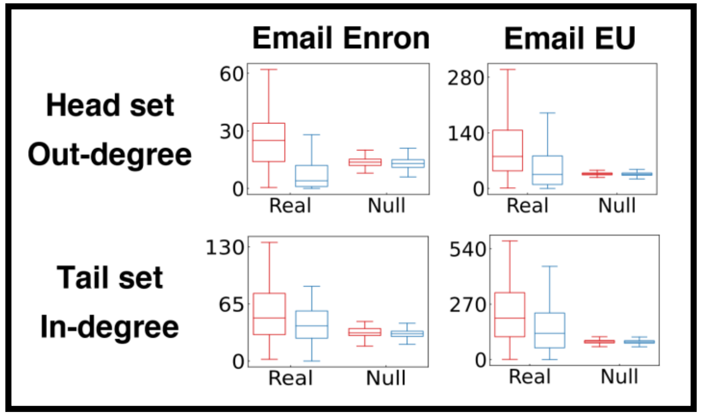
		metabolic		email		citation		q&a		bitcoin		
		iAF1260b	iJO1366	enron	eu	data mining	software	math	server	2014	2015	2016
Real World	$r(G)$	21.455	22.533	59.001	79.416	12.078	15.316	9.608	13.219	10.829	6.923	3.045
ReDi (Section 5)	$r(G)$	21.727	22.185	59.161	79.489	12.601	14.279	9.427	13.229	10.267	6.587	3.497
	D-Stat	0.098	0.104	0.053	0.043	0.212	0.151	0.011	0.005	0.045	0.033	0.017
Null	$r(G)$	0.306	0.270	14.862	4.633	0.094	0.147	0.018	0.002	0.0001	0.000*	0.000*
	D-Stat	0.625	0.642	0.539	0.807	0.355	0.377	0.124	0.160	0.147	0.100	0.050

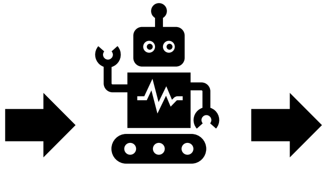
ReDi: Realistic Directed Hypergraph Generator

Reproducibility of observation 2 of ReDi:

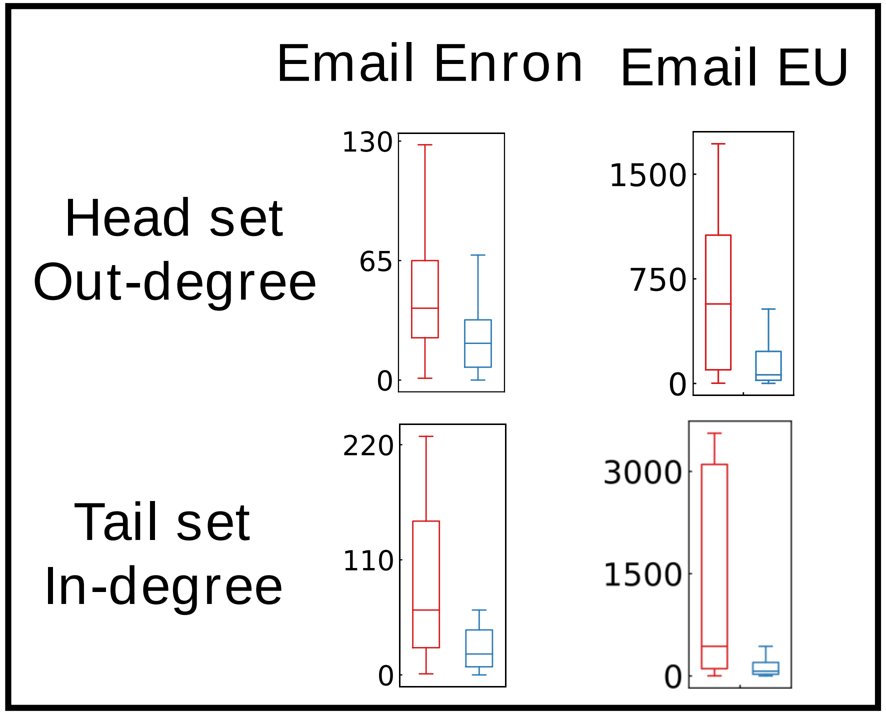
ReDi can preserve the relationship between arc degree and reciprocity.

Real-world





Generated



Arcs with Non-zero Reciprocity

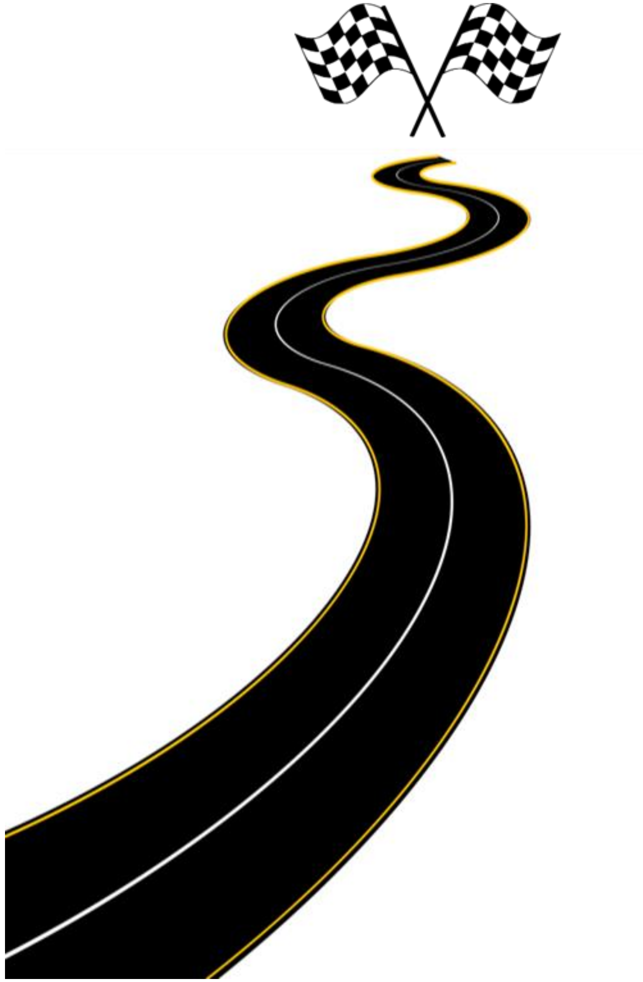
Arcs with Zero Reciprocity

Observation 2: Hyperarc Level.

Arcs with non-zero reciprocity tend to have higher head set out-degree and tail set in-degree.

Roadmap

- Overview
- Reciprocity Measure
- Observations
- Generators
- Conclusions



Conclusions

Our contributions in this work:

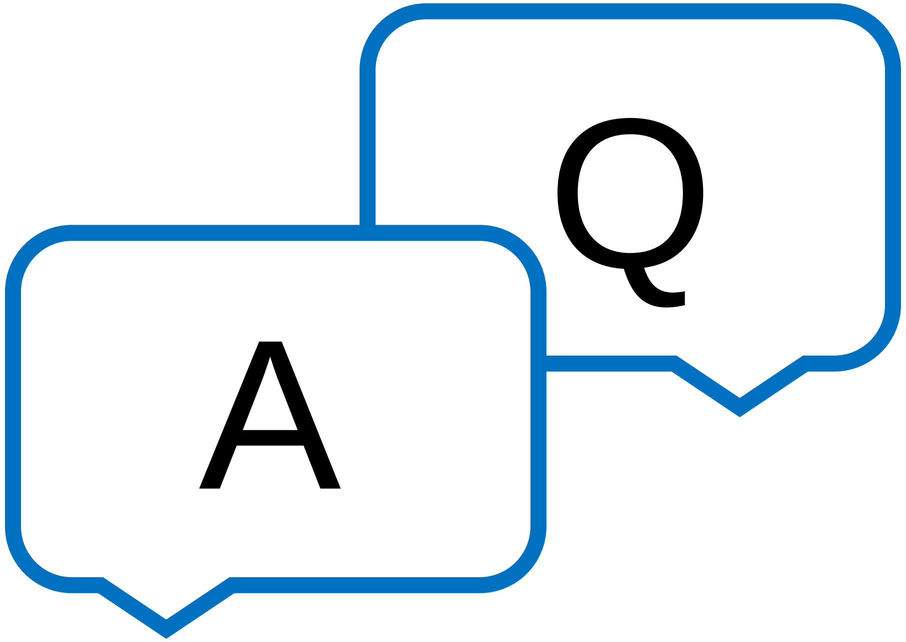
- Principled Reciprocal Measure: HyperRec
- Observations in Real-world Directed Hypergraphs
  1. Hypergraph level
  2. Hyperarc level
- Realistic Generative Model: ReDi

The code and datasets used in the paper are available at

https://github.com/kswoo97/hyprec

Skipped Details

- AXIOMs: Properties that a proper reciprocity measure should have.
- Reciprocal set: Composing a reciprocal set of the target arc.
- Components of HyperRec: Formal expression of HyperRec.
- Searching algorithm (Ferret): Finding a proper reciprocal set fast & accurately.
- Soundness of HyperRec: Theoretical analysis that a HyperRec can satisfy all the axioms.
- Exactness of Ferret: Theoretical analysis that a Ferret's output is accurate.
- Additional experimental results: Additional experiments and full dataset results regarding observations and generations.


Q
A