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
What is Directed Hypergraph?

- Directional relations + Group relations = Directional group relations

- Chemical reactions
- Question & Answering

Chemical Reaction:

\[ \text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O} \]

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
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
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 $e_1$.
- **Ideally**, it should return to each of its sources with prob. $1/(\text{tail set size})$.

![Diagram showing the concept of HyperRec with optimally reciprocal hyperarc $e'_1$.](image-url)
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.
HyberRec: Proposed Reciprocity Measure

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

Ideal transition (return) probabilities:

| Node | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------|---|---|---|---|---|---|---|
| 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.

![Diagram showing transition probabilities]

**Ideal**

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| e₂| 0 | 0 | 0 | 1/2| 1/2| 0 | 0 |
| Ideal| 0 | 0 | 0 | 1/3| 1/3| 1/3| 0 |

- Observed transition (return) probabilities are shown with emojis for clarity.
HyperRec: Proposed Reciprocity Measure

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

| Probability Distance | Ideal  | Observed |
|----------------------|--------|----------|
| 0 0 0 0 1/3 1/3 1/3 0 | 0 0 0 1/2 1/2 0 0 |

Ideal edge $e_2$: Transition probability from node 2 to other nodes.

Observed edge $e_2$: Transition probability from node 2 to other nodes.
HyperRec: Proposed Reciprocity Measure

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

Ideal transition probabilities:

| Node | 0 | 0 | 0 | 1/3 | 1/3 | 1/3 | 0 |
|------|---|---|---|-----|-----|-----|---|

Diagram showing the transition (return) probabilities from node 3.
HyperRec: Proposed Reciprocity Measure

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

Ideal

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
\end{array}
\]

\[
e_2
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\end{array}
\]
HyperRec: Proposed Reciprocity Measure

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

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
Ideal & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
e_3 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\end{array}
\]
HyperRec: Proposed Reciprocity Measure

- Transition (return) probability from node 3 to other nodes.
HyperRec: Proposed Reciprocity Measure

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

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
0 & 0 & 0 & 1 & 1 & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
\end{array}
\]
HyperRec: Proposed Reciprocity Measure

- Hyperarc level HyperRec

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

- 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

\[ r(e_i) = \max_{R_i \subseteq E, R_i \neq \emptyset} \left( \frac{1}{|R_i|} \right)^\alpha \left( 1 - \frac{1}{|H_i|} \sum_{v \in H_i} \frac{L(p_v, p_v^*)}{L_{\text{max}}} \right) \]

• Hypergraph level HyperRec

\[ r(G) = \frac{1}{|E|} \sum_{e_i \in E} r(e_i) \]

Reciprocal set size penalty term

Using a reciprocal set that can maximize the hyperarc reciprocity

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

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

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

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

**Bitcoin (2014, 2015, 2016)**
- Node: Users
- Hyperarcs: Transactions
Observation 1: Hypergraph Level.
Real-world hypergraphs are more reciprocal than randomized hypergraphs.

|          | metabolic | email | citation | q&a | bitcoin |
|----------|-----------|-------|----------|-----|---------|
|          | AF1260b   | enron| data mining | math | 2014    |
|          | iJO1366   | eu   | software | server | 2015    |
|          |           |      |          |       | 2016    |
| Real World | r(G)     | 21.455 | 59.001 | 12.078 | 9.608 |
|           |          | 22.533 | 79.416 | 15.316 | 13.219 |
| Null     | r(G)     | 0.306  | 14.862 | 0.094  | 0.018  |
| D-Stat   |          | 0.625  | 0.539  | 0.147  | 0.002  |
|          |          | 0.642  | 0.807  | 0.355  | 0.124  |

** Null hypergraph is a randomized hypergraph that fills nodes uniformly at random where hyperarc size is preserved.
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 $\beta_1$ and $\beta_2$.

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 \(\equiv\) Can preserve reciprocal patterns.
- ReDi iteratively add nodes and create hyperarcs \(\propto\) group degree.
- Extent of reciprocity is being controlled by hyperparameters \(\beta_1\) and \(\beta_2\).

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.

Add two hyperarcs!
ReDi: Realistic Directed Hypergraph Generator

• Realistic $\equiv$ Can preserve reciprocal patterns.
• ReDi iteratively add nodes and create hyperarcs $\propto$ group degree.
• Extent of reciprocity is being controlled by hyperparameters $\beta_1$ and $\beta_2$.

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.
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \beta_1$) or random ($\propto 1 - \beta_1$).

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 ($\propto \beta_2$), and decide how many and which nodes will we bring.

7. Fill remaining positions $\propto$ the group degree, and add hyperarc.
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \beta_1$) or random ($\propto 1 - \beta_1$).

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 ($\propto \beta_2$), and decide how many and which nodes will we bring.

7. Fill remaining positions $\propto$ the group degree, and add hyperarc.
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\alpha \beta_1$) or random ($\alpha (1 - \beta_1)$).

4. (Random) Put a new node in either head or tail set, and fill the remaining positions $\alpha$ the group degree. Then add hyperarc.

5. (Reciprocal) Choose a hyperarc to be reciprocal.

6. Toss a coin ($\alpha \beta_2$), and decide how many and which nodes will we bring.

7. Fill remaining positions $\alpha$ the group degree, and add hyperarc.
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \beta_1$) or random ($\propto 1 - \beta_1$).

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 ($\propto \beta_2$), and decide how many and which nodes will we bring.

7. Fill remaining positions $\propto$ the group degree, and add hyperarc.
3. Toss a coin, and decide whether current hyperarc is reciprocal ($\propto \beta_1$) or random ($\propto 1 - \beta_1$).

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 ($\propto \beta_2$), and decide how many and which nodes will we bring.

7. Fill remaining positions $\propto$ 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 | Real World | r(G) | D-Stat |
|----------|-----------|-------|----------|-----|------------|------|--------|
| iAF1260b| 21.455    | 59.001| 12.078   | 9.608| 10.829     | 22.533| 0.098  |
| iJO1366 | 22.185    | 79.416| 15.316   | 13.219| 6.923      | 0.104 | 0.143  |
| enron   |           |       |          |     |            |       |        |
| eu      |           |       |          |     |            |       |        |
| data    |           |       |          |     |            |       |        |
| mining  |           |       |          |     |            |       |        |
| software|           |       |          |     |            |       |        |
| math    | 12.601    | 0.212 | 9.427    | 0.011| 10.267     |       |        |
| server  | 79.489    | 0.151 | 13.229   | 0.005| 6.587      |       |        |
| 2014    | 14.279    |       |          |     |            |       |        |
| 2015    | 0.045     |       |          |     |            |       |        |
| 2016    | 0.033     |       |          |     |            |       |        |
| bitcoin | 0.017     |       |          |     |            |       |        |

|          | r(G)      | D-Stat |
|----------|-----------|--------|
| Null     | 0.306     | 0.625  |
|          | 0.270     | 0.642  |
|          | 14.862    | 0.539  |
|          | 4.633     | 0.807  |
|          | 0.094     | 0.355  |
|          | 0.147     | 0.377  |
|          | 0.018     | 0.124  |
|          | 0.002     | 0.160  |
|          | 0.0001    | 0.147  |
|          | 0.000*    | 0.100  |
|          | 0.000*    | 0.050  |
ReDi: Realistic Directed Hypergraph Generator

ReDi: Realistic Directed Hypergraph Generator

ReDi can preserve the relationship between arc degree and 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.
