On the Persistence of Higher-Order Interactions in Real-World Hypergraphs

Hyunjin Choo

Kijung Shin
Hypergraph

- A **hypergraph** is a generalization of an ordinary graph
- A **hyperedge** joins an **arbitrary** number of nodes

- Sender and receivers of an email
- Co-authors of a publication
- Items co-purchased by a customer
Higher-Order Interaction (HOI)

• A higher-order interaction (HOI) is the co-appearance of a set of nodes in any hyperedge
  ➢ E.g.) If $A$, $B$, and $C$ publish a paper together, any of $\{A, B\}$, $\{A, C\}$, $\{B, C\}$, $\{A, B, C\}$ becomes a HOI
Persistence of HOIs

- HOIs can appear repeatedly over time
- **Persistence** of repeated HOIs can be used to measure the strength or robustness of group relations
Applications

• Predicting the persistence of HOIs has many **potential applications**
  • Recommending groups (e.g., Facebook groups) in social networks
  • Recommending multiple items together
  • Predicting missing recipients of emails

- Amy
- Bob
- Carl
- Dan

Jan.

- Amy
- Bob
- Carl
- Dan

Feb.

- Amy
- Bob
- Carl
- Dan

Mar.

- Amy
- Bob
- Carl
- Dan

Apr.

- Amy
- Bob
- Carl
- Dan

Missing?
Our Questions

1. How do HOIs in real-world hypergraphs persist over time?
2. What are the key factors governing the persistence?
3. How accurately can we predict the persistence?
Roadmap

• Introduction

• Observations <<
  ◦ Hypergraph-Level Analysis
  ◦ Group-Level Analysis
  ◦ Node-Level Analysis

• Predictions

• Conclusions
Datasets

Observations

Coauthorship

Email

Predictions

NDC 0777-3105-02

Tags

SDM 2022  On the Persistence of Higher-Order Interactions in Real-World Hypergraphs
# Datasets

<table>
<thead>
<tr>
<th>Domain</th>
<th>Dataset</th>
<th>Node</th>
<th>Hyperedge</th>
<th>Time Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coauthorship</td>
<td>DBLP</td>
<td>an author</td>
<td>authors</td>
<td>1 Year</td>
</tr>
<tr>
<td></td>
<td>Geology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>History</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact</td>
<td>High</td>
<td>a person</td>
<td>a group interaction</td>
<td>1 Day</td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td></td>
<td></td>
<td>6 Hours</td>
</tr>
<tr>
<td>Email</td>
<td>Enron</td>
<td>an email address</td>
<td>sender and all receivers</td>
<td>1 Month</td>
</tr>
<tr>
<td></td>
<td>Eu</td>
<td></td>
<td></td>
<td>2 Weeks</td>
</tr>
<tr>
<td>NDC</td>
<td>Classes</td>
<td>a class label</td>
<td>class labels applied to a drug</td>
<td>2 Years</td>
</tr>
<tr>
<td></td>
<td>Substances</td>
<td>a substance</td>
<td>substances in a drug</td>
<td></td>
</tr>
<tr>
<td>Tags</td>
<td>Math.sx</td>
<td>a tag</td>
<td>tags added to a question</td>
<td>1 Month</td>
</tr>
<tr>
<td></td>
<td>Ubuntu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threads</td>
<td>Math.sx</td>
<td>a user</td>
<td>users who participate in a thread</td>
<td>1 Month</td>
</tr>
<tr>
<td></td>
<td>Ubuntu</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Timestamped Hyperedges**

- For each HOI $S$,
  - $E(S)$: Set of hyperedges containing $S$
  - $E(S, t)$: Set of hyperedges at time $t$ containing $S$
  - Hyperedge $e_i$ is associated with the timestamp $t_i$

**Examples:**

- $S = \{v_1, v_2, v_3\}$
- $E(S) = \{e_1, e_2, e_3\}$
- $E(S, 1) = \{e_1, e_2\}$
- $E(S, 2) = \emptyset$
- $E(S, 3) = \{e_3\}$

**Timestamped Hyperedges:**

- $e_1 = \{v_1, v_2, v_3, v_4\}, \quad t_1 = 1$
- $e_2 = \{v_1, v_2, v_3, v_5, v_6\}, \quad t_2 = 1$
- $e_3 = \{v_1, v_2, v_3, v_7\}, \quad t_3 = 3$
Measure: Persistence of a HOI

- **Persistence** of a HOI $S$ over a time range $T$ is the number of time units in $T$ when $S$ co-appear in any hyperedge, i.e.,

$$ P(S, T) := \sum_{t \in T} I(S, t) $$

where $I(S, t) = \begin{cases} 1, & \text{if } |E(S, t)| \geq 1 \\ 0, & \text{otherwise} \end{cases}$

- $E(S, 1) = \{e_1, e_2\}$
- $E(S, 2) = \emptyset$
- $E(S, 3) = \{e_3\}$

$$ P(S, [1, 3]) = \sum_{t=1}^{3} I(S, t) = 1 + 0 + 1 = 2 $$
Roadmap

• Introduction
• Observations
  ◦ Hypergraph-Level Analysis <<
  ◦ Group-Level Analysis
  ◦ Node-Level Analysis
• Predictions
• Conclusions
# Persistence vs. Frequency

**Obs. 1:** Persistence of HOIs tends to follow a **power-law**.

- **DBLP ($|S| = 2$)**
- **DBLP ($|S| = 3$)**
- **DBLP ($|S| = 4$)**

<table>
<thead>
<tr>
<th>Size of HOIs</th>
<th>$R^2$ of Fitted Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average over all 13 datasets</td>
<td>0.90</td>
</tr>
</tbody>
</table>
Persistence vs. Size of HOIs

**Obs. 2:** As HOIs grow in size, their average persistence and the power-law exponents of fitted power-law distributions tend to decrease.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Average Persistence (Relative)</th>
<th>Power-Law Exponent (Relative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of HOIs</td>
<td>2 3 4</td>
<td>2 3 4</td>
</tr>
<tr>
<td>Average over all 13 datasets</td>
<td>1.00 0.72 0.63</td>
<td>1.00 0.71 0.59</td>
</tr>
</tbody>
</table>
Roadmap

• Introduction
• Observations
  ◦ Hypergraph-Level Analysis
  ◦ Group-Level Analysis <<
  ◦ Node-Level Analysis
• Predictions
• Conclusions
Group Features vs. Group Persistence

- We examined the relations between the structural group features and the persistence of HOIs (i.e., group persistence).
- We measured the **Pearson correlation coefficient (CC)** and **normalized mutual information (MI)** between the persistence and each structural feature to examine the relation between them.
  - Normalized mutual information scales from 0 (no mutual information) to 1 (perfect correlation).
Group Features: Definition

• Basic structural features of each HOI $S$:
  • $\#$: number of hyperedges including $S$
  • $\Sigma$: sum of sizes of hyperedges containing $S$
  • $\cup$: number of hyperedges overlapping $S$
  • $\Sigma \cup$: sum of sizes of hyperedges overlapping $S$
  • $\cap$: number of common neighbors of $S$
  • $\mathcal{H}$: entropy in sizes of hyperedges containing $S$

• Group structural features of each HOI $S$:
  ➢ (1) $\#$, (2) $\# / \cup$, (3) $\Sigma / (\Sigma \cup)$, (4) $\cap$, (5) $\# / \cap$, (6) $\Sigma / \cap$, (7) $\Sigma / \#$, (8) $\mathcal{H}$

  density of hyperedges containing $S$           avg. sizes of hyperedges containing $S$
Measure: Structural Features & Persistence

1) HOI $S$ appears in a hyperedge for the first time at time $t$
Measure: Structural Features & Persistence

1) HOI $S$ appears in a hyperedge for the first time at time $t$

2) Compute its structural features using only the hyperedges appearing between time $t + 1$ and $t + T_s$
Measure: Structural Features & Persistence

1) HOI $S$ appears in a hyperedge for the first time at time $t$

2) Compute its structural features using only the hyperedges appearing between time $t + 1$ and $t + T_s$

3) Measure its persistence between time $t + T_s + 1$ and $t + T_s + T_p$

• We set $T_s = 5$ and $T_p = 10$
## Group Features vs. Group Persistence

**Obs. 3:** Persistence of each HOI $S$ is positively correlated with (a) the **number of hyperedges containing $S$** and (b) the **entropy in the sizes of hyperedges containing $S$**.

<table>
<thead>
<tr>
<th>Size of HOIs</th>
<th># of HOIs</th>
<th>MI</th>
<th># U</th>
<th>Σ U</th>
<th>∩</th>
<th># Σ</th>
<th>∑</th>
<th>∑ #</th>
<th>$\mathcal{H}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.13</td>
<td>0.11</td>
<td>0.14</td>
<td>0.05</td>
<td>0.10</td>
<td>0.12</td>
<td>0.10</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>0.06</td>
<td>0.08</td>
<td>0.05</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.11</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td>0.10</td>
<td>0.07</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Avg.</td>
<td>0.12</td>
<td>0.08</td>
<td>0.10</td>
<td>0.05</td>
<td>0.08</td>
<td>0.11</td>
<td>0.08</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.36</td>
<td>0.09</td>
<td>0.09</td>
<td>0.17</td>
<td>0.19</td>
<td>0.26</td>
<td>-0.08</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.31</td>
<td>0.10</td>
<td>0.10</td>
<td>0.05</td>
<td>0.16</td>
<td>0.20</td>
<td>-0.09</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.30</td>
<td>0.13</td>
<td>0.13</td>
<td>-0.01</td>
<td>0.17</td>
<td>0.20</td>
<td>-0.10</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Avg.</td>
<td>0.32</td>
<td>0.10</td>
<td>0.11</td>
<td>0.07</td>
<td>0.17</td>
<td>0.22</td>
<td>-0.09</td>
<td>0.27</td>
<td></td>
</tr>
</tbody>
</table>
Obs. 3: Persistence of each HOI $S$ is positively correlated with (a) the number of hyperedges containing $S$ and (b) the entropy in the sizes of hyperedges containing $S$. 
Node Features: Definition

- We examine the relations between the persistence of each HOI (i.e., group persistence) and the structural features of individual nodes involved in the HOI

- Structural features of each node \( v \) in the clique expansion:
  a. \textbf{degree} \( d(v) \)
  b. \textbf{weighted degree} \( w(v) \)
  c. \textbf{core number} \( c(v) \)
  d. \textbf{PageRank} \( r(v) \)
  e. \textbf{average degree of neighbors} \( \bar{d}(v) \)
  f. \textbf{average weighted degree of neighbors} \( \bar{w}(v) \)
  g. \textbf{local clustering coefficient} \( l(v) \)
  h. \textbf{number of occurrences of} \( v \) \( o(v) \)
Clique Expansion: Definition

• The **clique expansion** of a hypergraph is a pairwise graph between nodes

• It is obtained by replacing each hyperedge with the clique with the nodes in the hyperedge
Node Features vs. Group Persistence

**Obs. 4:** Persistence of each HOI S is negatively correlated with the average (weighted) degree of neighbors of each node involved in the HOI.

<table>
<thead>
<tr>
<th>Size of HOIs</th>
<th>d</th>
<th>w</th>
<th>c</th>
<th>r</th>
<th>(\bar{d})</th>
<th>(\bar{w})</th>
<th>l</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.09</td>
<td>0.04</td>
<td>0.17</td>
<td>0.16</td>
<td>0.17</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.04</td>
<td>0.07</td>
<td>0.05</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.07</td>
<td>-0.12</td>
<td>-0.14</td>
<td>-0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>3</td>
<td>-0.02</td>
<td>0.06</td>
<td>-0.05</td>
<td>0.03</td>
<td>-0.11</td>
<td>-0.12</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>-0.07</td>
<td>0.03</td>
<td>-0.09</td>
<td>0.03</td>
<td>-0.14</td>
<td>-0.14</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Avg.</td>
<td>-0.01</td>
<td>0.06</td>
<td>-0.05</td>
<td>0.04</td>
<td>-0.12</td>
<td>-0.13</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Node Features vs. Group Persistence

**Obs. 4:** Persistence of each HOI $S$ is negatively correlated with the *average (weighted) degree of neighbors* of each node involved in the HOI.

- **DBLP ($|S| = 2$)**
- **DBLP ($|S| = 3$)**
- **DBLP ($|S| = 4$)**
- **Eu ($|S| = 2$)**
- **Eu ($|S| = 3$)**
- **Eu ($|S| = 4$)**
Roadmap

• Introduction
• Observations
  ◦ Hypergraph-Level Analysis
  ◦ Group-Level Analysis
  ◦ **Node-Level Analysis **
• Predictions
• Conclusions
Node Features vs. Node Persistence

- We explore the relations between the structural features of each node and its $k$-node persistence

- **$k$-node persistence** of a node $v$: average persistence of the HOIs of size $k \in \{2,3,4\}$ that the node $v$ is involved in

- For each node $v$, let $t_v$ be the time when $v$ is involved in any HOI of size $k$ for the first time
  
  ➢ Measure the structural node features of $v$ using only the hyperedges appearing between time $t_v + 1$ and $t_v + T_S$

  1. First appearance of a HOI of size $k$ containing $v$
  2. Observe structural features $t_v + T_s$
  3. Measure $k$-node persistence $t_v + T_s + T_p$
Node Features vs. Node Persistence

**Obs. 5:** The *weighted degree* and *number of occurrences* of each node are positively correlated with the $k$-node persistence of HOIs that the node is involved in.

<table>
<thead>
<tr>
<th>Size of HOIs</th>
<th>$d$</th>
<th>$w$</th>
<th>$c$</th>
<th>$r$</th>
<th>$\bar{d}$</th>
<th>$\bar{w}$</th>
<th>$l$</th>
<th>$o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
<td>0.43</td>
<td>0.28</td>
<td>0.53</td>
<td>0.49</td>
<td>0.51</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
<td>0.37</td>
<td>0.24</td>
<td>0.44</td>
<td>0.42</td>
<td>0.44</td>
<td>0.37</td>
<td>0.34</td>
</tr>
<tr>
<td>4</td>
<td>0.26</td>
<td>0.31</td>
<td>0.21</td>
<td>0.36</td>
<td>0.35</td>
<td>0.36</td>
<td>0.31</td>
<td>0.30</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.30</td>
<td>0.37</td>
<td>0.24</td>
<td>0.44</td>
<td>0.42</td>
<td>0.43</td>
<td>0.37</td>
<td>0.35</td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
<td>0.22</td>
<td>0.14</td>
<td>0.08</td>
<td>0.00</td>
<td>-0.07</td>
<td>-0.02</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
<td>0.16</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.04</td>
<td>-0.08</td>
<td>-0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.12</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.07</td>
<td>0.17</td>
<td>0.06</td>
<td>0.04</td>
<td>-0.03</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Node Features vs. Node Persistence

**Obs. 5:** The **weighted degree** and **number of occurrences** of each node are positively correlated with the \( k \)-node persistence of HOIs that the node is involved in.

- **DBLP (|S| = 2)**
- **DBLP (|S| = 3)**
- **DBLP (|S| = 4)**
- **Eu (|S| = 2)**
- **Eu (|S| = 3)**
- **Eu (|S| = 4)**
Roadmap

• Introduction

• Observations
  ◦ Hypergraph-Level Analysis
  ◦ Group-Level Analysis
  ◦ Node-Level Analysis

• Predictions

• Conclusions
Prediction Experiments

- **Exp. 1: Predictability.** How accurately can we predict the persistence of HOIs using the structural features?
- **Exp. 2: Feature Importance.** Which structural features are important in predicting the persistence?
- **Exp. 3: Effect of Observation Periods.** How does the period of observation for measuring the structural features affect the prediction accuracy?
Problem 1: Persistence Prediction

• Given:
  – a **HOI $S$** that appears for the first time at time $t$,
  – all hyperedges appearing in the past
    – between time $t + 1$ and $t + T_s$

• Predict:
  – **persistence of $S$ in the near future**
    – between $t + T_s + 1$ and $t + T_s + T_p$
Problem 2: $k$-Node Persistence Prediction

• Given:
  – a node $v$ involved in a HOI of size $k$ for the first time at time $t$,
  – all hyperedges appearing in the past
    – between time $t + 1$ and $t + T_s$

• Predict:
  – $k$-node persistence of $v$ in the near future
    – between $t + T_s + 1$ and $t + T_s + T_p$
Prediction Methods

• We use all 16 structural features (8 group and 8 node features) as input features into four regression models:
  1) multiple linear regression (LR)
  2) random forest regression (RF)
  3) linear support vector regression (SVR)
  4) multi-layer perceptron regressor (MLP)

✓ Baseline: mean (k-node) persistence in the training set

• Training set: 2/3 of the HOIs and their persistence and 4/5 of the nodes and their k-node persistence

• Test set: the remaining ones
Evaluation Methods

• We evaluate the predictive performance of the models using two metrics:
  
  ➢ **Coefficients of determination \((R^2)\):** measures how well the predictions approximate the real data

  ➢ **Root mean squared error \((RMSE)\):** between predicted and real \((k\)-node) persistence

• A **higher \(R^2\)** and **lower \(RMSE\)** indicate better performance
Exp. 1: Predictability

**Obs. 6:** The **structural features are useful** for predicting the persistence, especially when the size of the HOI is large.

<table>
<thead>
<tr>
<th>Target</th>
<th>Prediction of Persistence of HOIs</th>
<th>Prediction of $k$-Node Persistence of Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>$R^2^*$</td>
<td>RMSE**</td>
</tr>
<tr>
<td>Size of HOIs</td>
<td>2 3 4</td>
<td>2 3 4</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00 0.00 0.00</td>
<td>1.29 0.73 0.60</td>
</tr>
<tr>
<td>SVR</td>
<td>0.17 0.13 0.10</td>
<td>1.12 0.63 0.48</td>
</tr>
<tr>
<td>LR</td>
<td>0.28 0.22 0.23</td>
<td>1.05 0.58 0.45</td>
</tr>
<tr>
<td>MLP</td>
<td>0.34 0.31 0.37</td>
<td>0.95 0.53 0.42</td>
</tr>
<tr>
<td>RF</td>
<td><strong>0.61</strong> <strong>0.62</strong> <strong>0.68</strong></td>
<td><strong>0.83</strong> <strong>0.38</strong> <strong>0.24</strong></td>
</tr>
</tbody>
</table>

*The higher, the better. **The lower, the better.
Measure: Feature Importance

• We use the **Gini importance** to measure the importance of each structural feature for random forest

• We compute the **rankings** of the features based on the importance
### Exp. 2: Feature Importance

**Obs. 7:** In predicting the persistence, the number of hyperedges containing $S$ (i.e., #), and the average (weighted) degree of the neighbors of each node in $S$ (i.e., $\bar{w}$ and $\bar{d}$) are most useful.

<table>
<thead>
<tr>
<th>Size of HOIs</th>
<th>#</th>
<th>$\frac{#}{\Sigma U}$</th>
<th>$\Sigma \frac{#}{\Sigma U}$</th>
<th>$\cap$</th>
<th>$\frac{#}{\cap}$</th>
<th>$\Sigma \frac{#}{\cap}$</th>
<th>$\Sigma$</th>
<th>$\mathcal{H}$</th>
<th>$d$</th>
<th>$w$</th>
<th>$c$</th>
<th>$r$</th>
<th>$\bar{d}$</th>
<th>$\bar{w}$</th>
<th>$l$</th>
<th>$o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><strong>2.8</strong></td>
<td>10.7</td>
<td>8.6</td>
<td>13.1</td>
<td>13.3</td>
<td>9.0</td>
<td>9.2</td>
<td>8.7</td>
<td>9.9</td>
<td>8.6</td>
<td>8.8</td>
<td>5.9</td>
<td>4.9</td>
<td>4.3</td>
<td>6.4</td>
<td>11.9</td>
</tr>
<tr>
<td>3</td>
<td><strong>5.4</strong></td>
<td>9.2</td>
<td>9.2</td>
<td>11.8</td>
<td>11.2</td>
<td>9.6</td>
<td>9.8</td>
<td>7.9</td>
<td>11.2</td>
<td>9.1</td>
<td>8.4</td>
<td>5.7</td>
<td><strong>5.1</strong></td>
<td><strong>4.3</strong></td>
<td>6.4</td>
<td>12.0</td>
</tr>
<tr>
<td>4</td>
<td><strong>5.3</strong></td>
<td>9.3</td>
<td>9.9</td>
<td>10.3</td>
<td>10.6</td>
<td>8.3</td>
<td>8.7</td>
<td>7.0</td>
<td>9.5</td>
<td>7.3</td>
<td>9.2</td>
<td>7.7</td>
<td>7.7</td>
<td>6.3</td>
<td>8.0</td>
<td>11.0</td>
</tr>
<tr>
<td>Avg.</td>
<td><strong>4.5</strong></td>
<td>9.7</td>
<td>9.2</td>
<td>11.7</td>
<td>11.7</td>
<td>9.0</td>
<td>9.2</td>
<td>7.9</td>
<td>10.2</td>
<td>8.3</td>
<td>8.8</td>
<td>6.4</td>
<td><strong>5.9</strong></td>
<td><strong>5.0</strong></td>
<td>6.9</td>
<td>11.6</td>
</tr>
</tbody>
</table>

**Feature Importance Ranking**
Exp. 2: Feature Importance

Obs. 8: In predicting the $k$-node persistence, its PageRank (i.e., $r$) and the average (weighted) degree of its neighbors (i.e., $\bar{w}$ and $\bar{d}$) are most useful.

<table>
<thead>
<tr>
<th>Size of HOIs</th>
<th>$d$</th>
<th>$w$</th>
<th>$c$</th>
<th>$r$</th>
<th>$\bar{d}$</th>
<th>$\bar{w}$</th>
<th>$l$</th>
<th>$o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.7</td>
<td>4.3</td>
<td>7.2</td>
<td>3.2</td>
<td>3.4</td>
<td>2.9</td>
<td>5.3</td>
<td>3.2</td>
</tr>
<tr>
<td>3</td>
<td>6.6</td>
<td>4.1</td>
<td>7.3</td>
<td>2.7</td>
<td>3.5</td>
<td>2.7</td>
<td>5.0</td>
<td>4.3</td>
</tr>
<tr>
<td>4</td>
<td>6.1</td>
<td>4.0</td>
<td>6.6</td>
<td>2.6</td>
<td>3.5</td>
<td>3.1</td>
<td>5.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Avg.</td>
<td>6.4</td>
<td>4.1</td>
<td>7.0</td>
<td>2.8</td>
<td>3.5</td>
<td>2.9</td>
<td>5.2</td>
<td>4.1</td>
</tr>
</tbody>
</table>

Feature Importance Ranking
Exp. 2: Effect of Number of Features

Obs. 9: About a half of the considered structural features based on their importance yields similar performance.
Exp. 3: Effect of Observation Periods

**Obs. 10:** Observing HOIs for longer periods of time enables us to better predict their persistence.

<table>
<thead>
<tr>
<th>Target</th>
<th>Persistence of HOIs</th>
<th>$k$-Node Persistence of Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE$^*$ of RF</td>
<td>Improvement (in %)</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_s$</td>
<td>2** 3 4</td>
<td>2 3 4</td>
</tr>
<tr>
<td>1</td>
<td>0.96 0.48 0.32</td>
<td>31.6 42.3 50.7</td>
</tr>
<tr>
<td>3</td>
<td>0.88 0.42 0.28</td>
<td>34.1 45.4 55.0</td>
</tr>
<tr>
<td>5</td>
<td><strong>0.83</strong> 0.38 0.24</td>
<td><strong>36.0</strong> 47.7 <strong>59.4</strong></td>
</tr>
</tbody>
</table>

*The lower, the better. **The size of HOIs (i.e., $|S|$).*
Roadmap

• Introduction

• Observations
  ◦ Hypergraph-Level Analysis
  ◦ Group-Level Analysis
  ◦ Node-Level Analysis

• Predictions

• Conclusions <<
Conclusions

• We empirically examined the persistence of HOIs at hypergraph-, group-, and node- levels in 13 real-world hypergraphs to answer the following questions:

✓ How is the persistence of HOIs distributed?
✓ Which structural features govern the persistence of HOIs?
✓ How accurately can we forecast the persistence of HOIs?

• Github link: https://github.com/jin-choo/persistence
On the Persistence of Higher-Order Interactions in Real-World Hypergraphs

Hyunjin Choo

Kijung Shin