Reciprocity in Directed Hypergraphs: Measures, Findings, and Generators

Sunwoo Kim[1], Minyoung Choe[1], Jaemin Yoo[3], and Kijung Shin[1, 2]
[1]Kim Jaechul Graduate School of AI, KAIST, [2]School of Electrical Engineering, KAIST
[3]Heinz College of Information Systems and Public Policy, Carnegie Mellon University
kswoo97@kaist.ac.kr, minyoung.choe@kaist.ac.kr, jaeminyoo@cmu.edu, kijungs@kaist.ac.kr

***Abstract*—Group interactions are prevalent in a variety of areas. Many of them, including email exchanges, chemical reactions, and bitcoin transactions, are directional, and thus they are naturally modeled as directed hypergraphs, where each hyperarc consists of the set of source nodes and the set of destination nodes. For directed graphs, which are a special case of directed hypergraphs, reciprocity has played a key role as a fundamental graph statistic in revealing organizing principles of graphs and in solving graph learning tasks. For general directed hypergraphs, however, even no systematic measure of reciprocity has been developed.**

In this work, we investigate the reciprocity of 11 real-world hypergraphs. To this end, we first introduce eight axioms that any reasonable measure of reciprocity should satisfy. Second, we propose HYPERREC, a principled measure of hypergraph reciprocity that satisfies all the axioms. Third, we develop FERRET, a fast and exact algorithm for computing the measure, whose search space is up to $10^{147}\times$ smaller than that of naive computation. Fourth, using them, we examine 11 real-world hypergraphs and discover patterns that distinguish them from random hypergraphs. Lastly, we propose REDI, an intuitive generative model for directed hypergraphs exhibiting the patterns. The code and the datasets are available at `https://github.com/kswoo97/hyprec`.

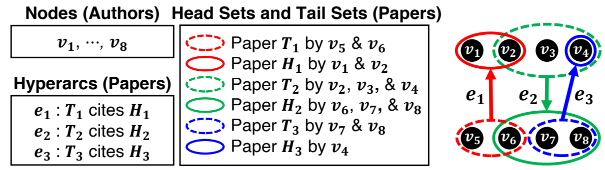
***Index Terms*—Reciprocity, Directed Hypergraph, Generator**

I. INTRODUCTION

Beyond pairwise interactions, understanding and modeling group-wise interactions in complex systems have recently received considerable attention [1]–[4]. A *hypergraph*, which is a generalization of a graph, has been used widely as an appropriate abstraction for such group-wise interactions. Each hyperedge in a hypergraph is a set of any number of nodes, and thus it naturally represents a group-wise interaction.

Many group-wise interactions are directional, and they are modeled as a *directed hypergraph*, where each hyperarc consists of the set of source nodes and the set of destination nodes. Examples of directional group-wise interactions include email exchanges (from senders to receivers), chemical reactions [5], road networks [6], and bitcoin transactions [7]; and they are modeled as directed hypergraphs for various applications [5], [6]. See Figure 1 for an example of hypergraph modeling.

Reciprocity [8], [9], which quantifies how mutually nodes are linked, has been used widely as a basic statistic of directed graphs, which are a special case of directed hypergraphs where every arc has exactly one source node and one destination node. Reciprocity helps understanding of a graph, especially potential organizing principles of it, and has proved useful for various tasks, including anomaly detection [10], and analysis of the spread of a computer virus through emails [8].



(a) Example Citation Dataset (b) Model

Fig. 1. A citation dataset modeled as a directed hypergraph with 8 nodes and 3 hyperarcs. Nodes correspond to authors. Hyperarcs correspond to citations. The head set and tail set of each hyperarc correspond to sets of papers.

However, reciprocity has remained unexplored for directed hypergraphs, and to the best of our knowledge, even no principled measure of reciprocity has been defined for directed hypergraphs. One straightforward approach is to compute the reciprocity after replacing a given directed hypergraph into a directed graph by clique expansion (i.e., replacing each hyperarc with the directed bi-clique from its source nodes to its destination nodes), as suggested in [11]. However, clique expansion may incur considerable information loss [5], [12], [13]. Thus, multiple directed hypergraphs whose reciprocity should differ, if they are determined by a proper measure, may become indistinguishable after being clique-expanded.

In this work, we investigate the reciprocity of real-world hypergraphs based on the first principled notion of reciprocity for directed hypergraphs. Our contributions toward this goal are summarized as follows:

- **Principled Reciprocity Measure:** We design HYPERREC, a probabilistic measure of hypergraph reciprocity. We prove that HYPERREC satisfies eight axioms that any reasonable measure of hypergraph reciprocity should satisfy.
- **Fast and Exact Search Algorithm:** The size of search space for computing HYPERREC is exponential in the number of hyperarcs. We develop FERRET, a fast and exact algorithm for computing HYPERREC, whose search space is up to $10^{147}\times$ smaller than that of naive computation.
- **Observations in Real-world Hypergraphs:** Using HYPERREC and FERRET, we investigate 11 real-world directed hypergraphs, and discover two reciprocal patterns pervasive in them, which are verified using a null hypergraph model.
- **Realistic Generative Model:** To confirm our understanding of the patterns, we develop REDI, a directed-hypergraph generator based on simple mechanisms on individual nodes. Our experiments demonstrate that REDI yields directed hypergraphs with realistic reciprocal patterns.

II. Basic Concepts and Related Work

We introduce some basic concepts and related studies.

A. Basic Concepts

A ***directed hypergraph*** $G = (V, E)$ consists of a set of nodes $V = \{v_1, \cdots, v_{|V|}\}$ and a set of ***hyperarcs*** $E = \{e_1, \cdots, e_{|E|}\} \subseteq \{\langle H, T\rangle : H \subseteq V, T \subseteq V\}$. For each hyperarc $e_i = \langle H_i, T_i\rangle \in E$, H_i indicates the ***head set*** and T_i indicates the ***tail set***. In Figure 1, the hyperarc $e_1 = \langle H_1, T_1\rangle \in E$ is represented as an arrow that heads to $H_1 = \{v_1, v_2\}$ from $T_1 = \{v_5, v_6\}$. It is assumed typically and also in this work that, in every hyperarc, the head set and the tail set are disjoint (i.e., $H_i \cap T_i = \emptyset, \forall i = 1, \cdots, |E|$). The ***in-degree*** $d_{in}(v) = |\{e_i \in E : v \in H_i\}|$ of a node $v \in V$ is the number of hyperarcs that include v as a head. Similarly, the ***out-degree*** $d_{out}(v) = |\{e_i \in E : v \in T_i\}|$ of $v \in V$ is the number of hyperarcs that include v as a tail.

From now on, we will use the term ***hypergraph*** to indicate a *directed hypergraph* and use the term *undirected hypergraph* to indicate an undirected one. We will also use the term ***arc*** to indicate a hyperarc when there is no ambiguity.

B. Related Work

Reciprocity of Directed Graphs: Reciprocity of directed graphs (i.e., a special case of directed hypergraphs where all head sets and tail sets are of size one) is a tendency of two nodes to be mutually linked [8], [9]. This is formally defined as $|E^{\leftrightarrow}|/|E|$, where $|E|$ is the number of edges in a graph, and $|E^{\leftrightarrow}|$ is the number of edges whose opposite directional arc exists. The notion was extended to weighted graphs [10], [14], and using them, the relationship between degree and reciprocity was investigated [10]. Moreover, the preferential attachment model [15] was extended by adding a parameter that controls the probability of creating a reciprocal edge for generating reciprocal graphs [16].

Patterns and Generative Models of Hypergraphs: Hypergraphs have been used widely for modeling group-wise interactions in complex systems, and considerable attention has been paid to the structural properties of real-world hypergraphs, with focuses on node degrees [2], [3], singular values [2], [3], diameter [2], [3], density [3], the occurrences of motifs [17], [18], the repetition of hyperedges [19], [20], and the overlap of hyperedges [4]. Many of these patterns can be reproduced by hypergraph generative models that are based on intuitive mechanisms [2]–[4], [19]. Such models can be used for anonymization and graph upscaling in addition to testing our understanding of the patterns [21]. All the above studies are limited to undirected hypergraphs, while this paper focuses on directed hypergraphs.

Directed Hypergraphs and Reciprocity: Directed hypergraphs have been used for modeling chemical reactions [5], knowledge bases [22], road networks [6], bitcoin transactions [7], etc. To the best of our knowledge, there has been only one attempt to measure the reciprocity of directed hypergraphs [11], where a hypergraph G is transformed into a weighted digraph $\bar{G}$ by *clique expansion*, for the digraph [8], [9] reciprocity measure is applied. However, as discussed in Section I, clique expansion may cause substantial information loss [12], and thus multiple directed hypergraphs whose reciprocities should differ, if they are determined by a proper measure, can be transformed into the same directed graphs by clique expansion.

III. Directed Hypergraph Reciprocity

In this section, we present eight necessary properties of an appropriate hypergraph reciprocity measure in the form of axioms. Then, we present our reciprocity measure, namely HyperRec, which satisfies all the axioms. Lastly, we propose an algorithm Ferret for fast computation of HyperRec.

A. Framework and Axioms

We present our framework for measuring hypergraph reciprocity. Then, we suggest eight axioms that any reasonable reciprocity measure must satisfy.

Framework for Hypergraph Reciprocity: Given a hypergraph G, we measure its reciprocity at two levels:

- How much each arc (i.e., group interaction) is reciprocal.
- How much the entire hypergraph G is reciprocal.

For a ***target arc***, which we measure reciprocity for, multiple arcs should be involved in measuring its reciprocity inevitably. For example, in Figure 1, arc e_2's head set and tail set overlap with e_1 and e_3's tail set and head set, respectively, and thus we should consider both e_1 and e_3 in measuring e_2's reciprocity. In graphs, however, only the arc with the opposite direction is involved in the reciprocity of an arc. This unique characteristic of hypergraphs poses challenges in measuring reciprocity. The ***reciprocal set*** R_i of a target arc e_i is the set of ***reciprocal arcs*** that we use to compute the reciprocity of e_i, We use $r(e_i, R_i)$ to denote the ***reciprocity of an arc*** e_i, where the domain is $E \times 2^E$.[1] In graphs, a traditional reciprocity measure [8] is defined as the proportion of arcs between nodes that point both ways, and if we assign 1 to such an arc and 0 to the others as reciprocity, the proportion is equivalent to the average reciprocity of arcs. Similarly, we regard, as the ***reciprocity of a hypergraph*** G, the average reciprocity of arcs, i.e.,

$$r(G) := \frac{1}{|E|}\sum\nolimits_{i=1}^{|E|} r(e_i, R_i). \tag{1}$$

Axioms: What are the characteristics required for $r(e_i, R_i)$ and $r(G)$? We introduce eight axioms that any reasonable measure of $r(e_i, R_i)$ (**Axioms** 1-5) and $r(G)$ (**Axioms** 6-8) should satisfy. In **Axioms** 1-4, we compare the reciprocity of two target arcs e_i and e_j whose reciprocal sets are R_i and R_j, respectively. Moreover, in **Axiom** 2-4, we commonly assume two target arcs e_i and e_j are of equal size (i.e., $|H_i| = |H_j|$ and $|T_i| = |T_j|$). Here, we say two arcs e_i and $e_k \in R_i$ ***inversely overlap*** if and only if $H_i \cap T_k \neq \emptyset$ and $T_i \cap H_k \neq \emptyset$. Below, the statements in **Axioms** 1-4 are limited to the examples in Figure 2 for simplicity. The statements in **Axioms** 1-4, however, are generalized and formalized in [23].

[1] Note that all arcs in R_i are used in computing the reciprocity of e_i, and thus it does not correspond to a search space.

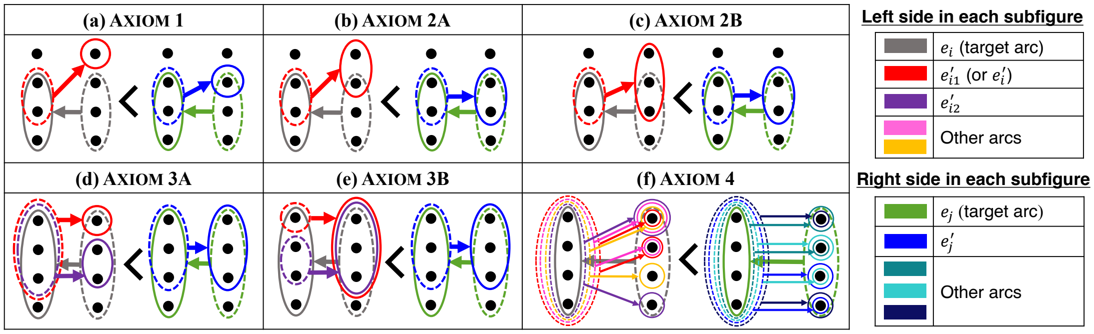


Fig. 2. Examples for **Axioms** 1-4. In each subfigure, the reciprocity of the arc e_i on the left side should be smaller than that of the arc e_j on the right side. This inequality holds by HyperRec (see Section III-B) in all subfigures. Specifically, if $\alpha = 1$, $r(e_i)$ & $r(e_j)$ are 0.0000 & 0.3605 in (a), 0.2697 & 0.5394 in (b), 0.4444 & 0.5394 in (c), 0.3167 & 0.6466 in (d), 0.3233 & 0.6466 in (e), and 0.2347 & 0.2500 in (f).

Axiom 1 (Existence of Inverse Overlap)**.** *In Figure 2(a), $r(e_i, R_i) < r(e_j, R_j)$ should hold. Roughly, an arc with at least one inverse-overlapping reciprocal arc is more reciprocal than an arc with no inverse-overlapping reciprocal arcs.*

Axiom 2 (Degree of Inverse Overlap)**.** *In Figures 2(b-c), $r(e_i, R_i) < r(e_j, R_j)$ should hold. Roughly, an arc that inversely overlaps with reciprocal arcs to a greater extent (with a larger intersection and/or with a smaller difference) is more reciprocal.*

Axiom 3 (Number of Reciprocal Arcs)**.** *In Figures 2(d-e), $r(e_i, R_i) < r(e_j, R_j)$ should hold. Roughly, an arc requiring fewer reciprocal arcs to inversely overlap to the same extent is more reciprocal.*

Axiom 4 (Bias)**.** *In Figure 2(f), $r(e_i, R_i) < r(e_j, R_j)$ should hold. Roughly, when two arcs inversely overlap **equally** with their reciprocal sets, an arc whose reciprocal arcs are equally reciprocal to all nodes in the arc is more reciprocal than one with reciprocal arcs **biased** towards some nodes in the arc.*

Axiom 5 (Upper and Lower Bounds)**.** *The reciprocity of any arc should be within a fixed range. Specifically, for every $e_i \in E$ and $R_i \in 2^E$, $r : E \times 2^E \mapsto [0, 1]$ should hold.*

Now, we present the axioms defined at the hypergraph level.

Axiom 6 (Inclusion of Graph Reciprocity)**.** *The graph reciprocity [8] should be included as a special case. That is, if G is a graph (i.e., $|H_i| = |T_i| = 1, \forall i \in \{1, \cdots, |E|\}$), then the following equality should hold:*

$$r(G) = |E^{\leftrightarrow}|/|E|, \tag{2}$$

where $E^{\leftrightarrow}$ is the set of arcs between nodes that point each other in both directions.

Axiom 7 (Upper and Lower Bounds)**.** *The reciprocity of any hypergraph should be within a fixed range. Specifically, for any hypergraph G, $r : G \mapsto [0, 1]$ should hold.*

Axiom 8 (Surjection of Reciprocity)**.** *The maximum reciprocity, which is 1 by* **Axiom** *7, should be reachable from any hypergraph by adding specific arcs. That is, for every $G = (V, E)$, there exist $G^* = (V, E^*)$ with $E^* \supseteq E$ such that $r(G^*) = 1$.*

B. Proposed Measure of Hypergraph Reciprocity: HyperRec

We propose HyperRec, a principled hypergraph-reciprocity measure based on transition probability.

Transition Probability: For a target arc $e_i = \langle H_i, T_i \rangle$ and its reciprocal arcs R_i, the ***transition probability*** $p_h(v)$ from a head set node $v_h \in H_i$ to each node v is the probability of a random walker transiting from v_h to v when she moves to a uniform random tail-set node of a uniform random arc among the reciprocal arcs incident to v_h. There might be some head set nodes that are not incident to any reciprocal arc. We assume that, from such a node, the random walker always transits to the virtual ***sunken node*** $v_{sunk} \notin V$. An example of how the transition probability is computed is given in [23].

Then, for each head set node $v_h \in H_i$ of a target arc e_i, a ***transition probability distribution*** over $V \cup \{v_{sunk}\}$ is defined, and we use p_h to denote it. We also denote an ***optimal transition probability distribution*** by p_h^*, which is a transition probability distribution when the perfectly reciprocal arc $e_i^* = \langle H_i^* = T_i, T_i^* = H_i \rangle$ is assumed as the reciprocal arc of e_i, i.e., $R_i = \{e_i^*\}$. The following equality always holds:

$$p_h^*(v) = \begin{cases} \frac{1}{|T_i|} & \text{if } v \in T_i, \\ 0 & \text{otherwise.} \end{cases}$$

Proposed Measure: Based on the above concepts, we propose HyperRec (Hypergraph Reciprocity) as a principled measure of hypergraph reciprocity. We notice that reciprocal arcs in a graph lead to paths of length two that start and end at the same node. Thus, intuitively, in a hypergraph, a target arc should become more reciprocal if its reciprocal arcs allow for heading back to the head-set nodes of the target arc more "accurately". In order to measure numerically the accuracy for a target arc e_i, we compare the transition probability distribution p_h from each head-set node $v_h \in H_i$ with the optimal distribution p_h^*.

While any distance function $\mathcal{L}$ can be used to quantify the difference between p_h and p_h^*, we use the *Jensen-Shannon Divergence* (JSD) [24] since it is a symmetric measure that can handle zero mass in both distributions. Based on $\mathcal{L}$, we define HyperRec of an arc e_i whose reciprocal set is R_i as

$$r(e_i, R_i) := \left(\frac{1}{|R_i|}\right)^{\alpha} \left(1 - \frac{\sum_{v_h \in H_i} \mathcal{L}(p_h, p_h^*)}{|H_i| \cdot \mathcal{L}_{\max}}\right), \tag{3}$$

where $\alpha \in (0, 1]$ is a constant controlling the degree of penalization of a large reciprocal set, and $\mathcal{L}_{\max}$ is the maximum value of the distance measure $\mathcal{L}$, which is $\log 2$ for the JSD. Note that $r(e_i, R_i)$ becomes larger if $\mathcal{L}(p_h, p_h^*)$ becomes small, implying that an arc is more reciprocal if its transition distribution becomes closer to the optimal distribution.

Composing Reciprocal Sets: The value of $r(e_i, R_i)$ is dependent on how we select the reciprocal set R_i from the set E of all arcs. For each target arc e_i, we propose to choose non-empty $R_i \subseteq E$ that maximizes the reciprocity $r(e_i, R_i)$ of e_i, i.e.,

$$R_i := \underset{R_i' \subseteq E, R_i' \neq \emptyset}{\operatorname{argmax}} \; r(e_i, R_i'). \tag{4}$$

In summary, according to HYPERREC, the reciprocity of an arc $e_i \in E$ is

$$r(e_i) := \max_{R_i \subseteq E, R_i \neq \emptyset} r(e_i, R_i), \tag{5}$$

and by Eq. (1), the reciprocity of G is $r(G) := \frac{1}{|E|} \sum_{i=1}^{|E|} r(e_i)$.

Axiomatic Analysis: HYPERREC satisfies all proposed **AXIOMS** regardless of the value of $\alpha > 0$, as stated in Theorem 1.

Theorem 1 (Soundness of HYPERREC)**.** HYPERREC always satisfies **AXIOMS** *1-8.*

Proof. The numerical values for the examples in Figure 2, which can be found in the caption, imply **AXIOMS** 1-4. Proofs of **AXIOMS** 5-8 can be found in [23].

C. *Exact and Rapid Search for Reciprocal Sets*

We propose FERRET (**F**ast and **E**xact Algo**r**ithm for Hypergraph R**e**ciproci**t**y), an approach for rapidly searching for the reciprocal set R_i of Eq. (4). We prove the exactness of FERRET and demonstrate its efficiency in real-world hypergraphs.

Procedure: Pseudocode of FERRET is given in [23]. For each arc e_i, we first retrieve the set Ω_i of inverse-overlapped arcs (see Section III-A for the definition) and check whether e_i is (1) non-reciprocal, (2) perfectly reciprocal, or (3) partially reciprocal. Reciprocity for the first two cases is 0 and 1, respectively. For a partially reciprocal case, we group the arcs in Ω_i using a mapping table Φ_i where the key of each arc $e_k \in \Omega_i$ is the head-set and tail-set nodes of e_i that it covers (i.e., $\langle H_i', T_i' \rangle$ where $H_i' \leftarrow H_i \cap T_k$ and $T_i' \leftarrow T_i \cap H_k$). For each group with the same key $\langle H_i', T_i' \rangle$, we choose an arc with the minimum number of head set nodes. Then, we create a new search space Ψ_i containing only the chosen arcs. After that, every subset R_i of Ψ_i is considered to maximize Eq. (3), and we return the maximum value as the reciprocity $r(e_i)$ of e_i.

Theoretical Properties and Evaluation: As stated in Theorem 2, FERRET finds the best reciprocal set, as in Eq. (4), i.e., it computes the reciprocity of each arc exactly, as in Eq. (5).

Theorem 2 (Exactness of FERRET)**.** *For every* $e_i \in E$, $\max_{R_i \subseteq E} r(e_i, R_i)$ *is identical to the* $\max_{R_i \subseteq \Psi_i} r(e_i, R_i)$*.*

Proof. Proof can be found in [23]. □

After the reduction above, the size of the search space for R_i becomes $O(2^{|\Psi_i|})$ in general, and as desribed in [23], it becomes $O(|\Psi_i|)$ for the case where every arc's tail set size is 1 (i.e., $|T_i| = 1, \forall i \in \{1, \cdots, |E|\}$). Although the complexity is still exponential, we demonstrate that the search space is reasonably small and thus a search can be performed within a reasonable time period (spec., at most 3.5 hours) for all considered real-world hypergraphs. Further analysis of the search space and running time can be found in [23].

IV. DATASETS AND OBSERVATIONS

In this section, we investigate the reciprocal patterns of real-world hypergraphs using HYPERREC and FERRET. After introducing used real-world hypergraph datasets and null hypergraphs, we discuss our observations at two different levels: hypergraphs and arcs. The significance of the patterns are verified by a comparison with the null hypergraphs.

A. *Datasets*

Datasets: We use 11 real-world hypergraphs from five different domains. Refer to [23] for the sources, preprocessing methods, and basic statistics of the hypergraphs.

- **Metabolic** (iAF1260b and iJO1366): Each network models chemical reactions among various genes. Nodes correspond to genes, and arcs indicate reactions.
- **Emails** (email-enron and email-eu): Each node is an email account, and each arc consists of two ordered sets of senders and receivers of an email.
- **Citations** (DBLP-data mining and DBLP-software). Each node is a researcher, and each head set and tail set indicates a paper. Arcs represent citations, as in Figure 1.
- **Question and Answering** (math-overflow and stack-exchange server fault). Each node is a user, and each arc corresponds to a post. The questioner of a post becomes the head of an arc and the answerers compose its tail set.
- **Bitcoin Transactions** (bitcoin-2014, 2015, 2016). Each node is an address in bitcoin transactions, and each arc is a transaction among users.

B. *Observations*

We investigate the reciprocal patterns of real-world hypergraphs at two different levels: hypergraphs and arcs. In order to demonstrate that discovered characteristics are distinguishable from random behavior, we measure the same statistics and patterns in randomized hypergraphs, which we call ***null hypergraphs***. Details of the null hypergraphs can be found in [23]. Due to the space limit, we report the results in only one dataset from each domain. Results that are not shown in this paper can be found in the [23].

L1. Hypergraph Level: Since hypergraph reciprocity $r(G) = \frac{1}{|E|} \sum_{e_i \in E} r(e_i)$ is robust to the choice of α, as shown in [23], we fix α to a value near zero for the investigation below. As shown in Table I, the hypergraph reciprocity is several orders of magnitude greater in real-world hypergraphs than in the corresponding null hypergraphs. Moreover, the differences are statistically significant, as shown in [23].

TABLE I

OBSERVATIONS 1 AND THE SUPERIORITY OF REDI. RECIPROCITY IN (A) REAL-WORLD HYPERGRAPHS, (B) NULL HYPERGRAPHS, (C) THOSE GENERATED BY REDI (SECTION V), AND (D) THOSE GENERATED BY A BASELINE GENERATOR IS REPORTED. AS THE ARC-LEVEL DIFFERENCE, WE REPORT THE D-STATISTIC (THE LOWER THE BETTER) BETWEEN EACH DISTRIBUTION OF ARC-LEVEL RECIPROCITY AND THAT IN THE CORRESPONDING REAL-WORLD HYPERGRAPH. VALUES BELOW 10^{-6} ARE ALL MARKED WITH *. IN EACH COLUMN, THE HYPERGRAPH RECIPROCITY CLOSEST TO THAT IN THE REAL-WORLD HYPERGRAPH AND THE MINIMUM D-STATISTIC ARE UNDERLINED. NOTE THAT REAL-WORLD HYPERGRAPHS ARE MORE RECIPROCAL THAN NULL HYPERGRAPHS, AND OUR PROPOSED GENERATOR, REDI, SUCCESSFULLY REPRODUCES THE RECIPROCITY IN REAL-WORLD 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
REDI (Section V)	$r(G)$	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
Baseline (Section V)	$r(G)$	0.412	0.851	23.846	31.190	0.048	0.004	1.622	0.002	0.002	0.002	0.001
	D-Stat	0.625	0.623	0.403	0.535	0.328	0.367	0.103	0.160	0.147	0.099	0.050

Observation 1. *Real-world hypergraphs are more reciprocal than randomized hypergraphs.*

L2. Arc Level: As shown empirically in [23], arc-level reciprocity is also robust to the choice of α. Thus, we fix α to a value near zero for the investigation below. At the arc level, we examine the relations between the degree of arcs and their reciprocity. We define head set out-degree ($d_{H,out}(e_i)$) and tail set in-degree ($d_{T,in}(e_i)$) as follows:

$$d_{H,out}(e_i) = \frac{1}{|H_i|}\sum_{v \in H_i} d_{out}(v) \quad d_{T,in}(e_i) = \frac{1}{|T_i|}\sum_{v \in T_i} d_{in}(v) \tag{6}$$

Refer to Section II-A for the definitions of $d_{out}(v)$ and $d_{in}(v)$. Then, we compare the distributional difference of these statistics (i.e., Eqs. (6)) between the arcs of zero reciprocity and those of non-zero reciprocity. As shown in Figure 3(a), the degrees at arcs with non-zero reciprocity tend to be greater than those at arcs with zero reciprocity. This is intuitive since arcs where their head sets are frequently being pointed and tail sets are frequently pointing others tend to have higher chance to be reciprocal. Such tendency, however, is not clear in null hypergraphs.

Observation 2. *Arcs with non-zero reciprocity tend to have higher head set out-degree and tail set in-degree than arcs with zero reciprocity.*

V. DIRECTED HYPERGRAPH GENERATION: REDI

In this section, we propose REDI (**Re**ciprocal and **Di**rectional Hypergraph Generator), a realistic generative model of directed hypergraphs. We first describe REDI. Then, we demonstrate its successful reproduction of the reciprocal properties of real-world hypergraphs examined in Section IV. In addition to testing our understanding of the patterns, REDI can also be used for anonymization, graph upscaling, etc [21].

A. Model Description

High-level Introduction to REDI: Given some basic hypergraph statistics and three hyperparameter values, REDI generates a directed hypergraph with realistic structural and reciprocal patterns. REDI is largely based on HYPERPA [2], an extension of the preferential attachment model [15] to hypergraphs. In HYPERPA, each new node forms hyperedges with groups of nodes that are drawn with probability proportional to the degree of groups (i.e., the number of hyperedges containing each group). Introducing the degree of groups, instead of the degree of individual nodes, tends to lead to more realistic higher-order structures of generated graphs [2]. REDI extends HYPERPA, which only can generate undirected hypergraphs, to generate directed hypergraphs and especially those with realistic reciprocal patterns. In a nutshell, REDI stochastically creates reciprocal arcs while controlling the number of reciprocal arcs and their degree of reciprocity.

Details of REDI: Pseudocode of REDI is provided in [23]. It requires three hyperparameters: (a) a proportion $\beta_1 \in [0,1]$ of reciprocal arc, (b) their extent $\beta_2 \in [0,1]$ of reciprocity, and (c) the number N of initial nodes. In addition, REDI requires the following statistics that it preserves in expectation: (a) the number n of nodes, (b) the distributions f_{HD} and f_{TD} of the head-set and tail-set sizes, and (c) the distribution f_{NP} of the number of new arcs per node.

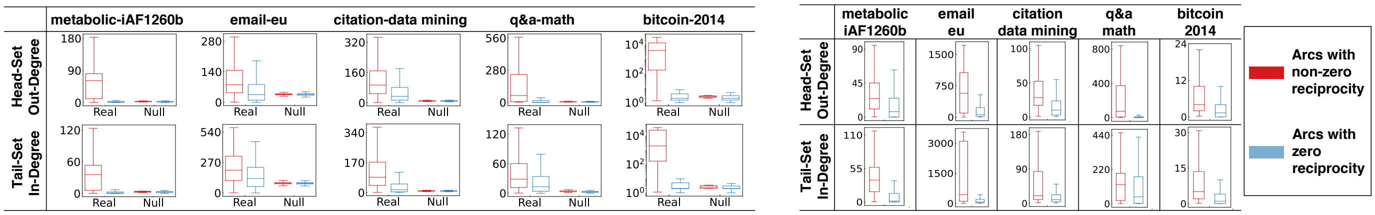
At each step, REDI introduces a new node v_i and creates k arcs where k is sampled from f_{NP}. Before creating a new arc, we decide whether it to be reciprocal (with prob. β_1) or ordinary. After deciding the size of a new arc according to the sizes sampled from f_{HD} and f_{TD}, we decide whether to include v into the head set (with prob. 0.5) or the tail set.

If a new arc is decided to be ordinary, we include v_i in either the head set or the tail set according to the choice made beforehand. Subsequently, we fill the new arc with nodes sampled based on in- and out-degrees of groups (i.e., the number of arcs that include the group in their head set and tail set, respectively). Note that the head set and the tail set should be disjoint for both reciprocal and ordinary arcs.

If a new arc is decided to be reciprocal, we choose an opponent arc e_o uniformly at random among those with v_i (or among all existing arcs if no arc contains v_i). Then, we decide how many nodes are brought from the opponent arc's head set and tail set by binomial sampling with probability $\beta_2 \in [0,1]$. After sampling nodes from the opponent arc with probability proportional to their degree, we fill the new arc with v_i and those sampled based on in- and out-degrees of groups.

B. Evaluation of REDI

We evaluate how well REDI can reproduces the reciprocal patterns of real-world hypergraphs discussed in Section IV. For each real-world hypergraph, we generate 5 hypergraphs using



(a) Observation 2 of real-world datasets

(b) Observation 2 of REDI

Fig. 3. (a) Observation 2. In real-world hypergraphs, the (I) head set out-degree and the (II) tail set in-degree tend to be larger at **arcs with non-zero reciprocity** than at **arcs with zero reciprocity**, while there is no such trend in null hypergraphs. (b) Hypergraphs generated by REDI exhibits Observation 2, which is a pervasive pattern in real-world hypergraphs, as shown in Figure 3(a).

their statistics and report the average of generated statistics.[2] In addition, we introduce a naive preferential attachment model, as a **baseline model** for comparison, to clarify the necessity of the reciprocal edge generation step. The baseline model is identical to REDI, except only for that it always decides to create ordinary arcs, i.e., $\beta_1 = \beta_2 = 0$.[3]

Reproducibility of Observation 1: We measure the reciprocity of generated hypergraphs at the hypergraph and arc levels and compare it with that of real-world hypergraphs. As shown in Table I, REDI generates hypergraphs whose reciprocity is very close to that in the corresponding real-world hypergraphs both at the hypergraph and arc levels. The baseline model fails to reproduce high enough reciprocity in most cases.

Reproducibility of Observation 2: Moreover, as shown in Figure 3(b), in the hypergraphs generated by REDI, arcs with non-zero reciprocity tend to have higher (I) head set out-degree and (II) tail set in-degree than arcs with zero reciprocity, just as in the real-world hypergraphs.

VI. CONCLUSION

In this paper, we perform a systematic and extensive study of reciprocity in real-world hypergraphs. We propose HYPERREC, a probabilistic measure of reciprocity that guarantees all eight desirable properties (Theorem 1). Our algorithmic contribution is to develop FERRET, which enables rapid yet exact computation of HYPERREC (Theorem 2). Using both, we discover several unique reciprocal patterns (Table I and Figures 3(a)) that distinguish real-world hypergraphs from random hypergraphs. Lastly, we design REDI, a simple yet powerful generator that yields realistic directed hypergraphs (Table I and Figures 3(b)).

Acknowledgements: This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1C1C1008296) and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00871, Development of AI Autonomy and Knowledge Enhancement for AI Agent Collaboration) (No. 2019-0-00075, Artificial Intelligence Graduate School Program (KAIST)).

[2]The search space of β_1 is (a) $[0.05, 0.1, \cdots, 0.6]$ for the small datasets where $|V| \leq 10^4$, and (b) $[0.001, 0.0015, \cdots, 0.005]$ for the dense large datasets where $|V| > 10^4$ and $|E|/|V| \geq 3$, and (c) $[0.01, 0.02, \cdots 0.15]$ for the other sparse large datasets. The search space of β_2 is fixed to $\in [0.1, 0.1, \cdots, 0.5]$ for all datasets.

[3]As discussed in detail [23], some minor changes are made in both REDI and the baseline model, when the statistics from the bitcoin and q&a datasets are given as their inputs.

REFERENCES

[1] C. Comrie and J. Kleinberg, "Hypergraph ego-networks and their temporal evolution," in *ICDM*, 2021.

[2] M. T. Do, S.-e. Yoon, B. Hooi, and K. Shin, "Structural patterns and generative models of real-world hypergraphs," in *KDD*, 2020.

[3] Y. Kook, J. Ko, and K. Shin, "Evolution of real-world hypergraphs: Patterns and models without oracles," in *ICDM*, 2020.

[4] G. Lee, M. Choe, and K. Shin, "How do hyperedges overlap in real-world hypergraphs?-patterns, measures, and generators," in *WWW*, 2021.

[5] N. Yadati, V. Nitin, M. Nimishakavi, P. Yadav, A. Louis, and P. Talukdar, "Nhp: Neural hypergraph link prediction," in *CIKM*, 2020.

[6] X. Luo, J. Peng, and J. Liang, "Directed hypergraph attention network for traffic forecasting," *IET Intelligent Transport Systems*, vol. 16, no. 1, pp. 85–98, 2022.

[7] S. Ranshous, C. A. Joslyn, S. Kreyling, K. Nowak, N. F. Samatova, C. L. West, and S. Winters, "Exchange pattern mining in the bitcoin transaction directed hypergraph," in *FC*, 2017.

[8] M. E. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Physical Review E*, vol. 66, no. 3, p. 035101, 2002.

[9] D. Garlaschelli and M. I. Loffredo, "Fitness-dependent topological properties of the world trade web," *Physical review letters*, vol. 93, no. 18, p. 188701, 2004.

[10] L. Akoglu, P. O. Vaz de Melo, and C. Faloutsos, "Quantifying reciprocity in large weighted communication networks," in *PAKDD*, 2012.

[11] N. Pearcy, J. J. Crofts, and N. Chuzhanova, "Hypergraph models of metabolism," *International Journal of Biological, Veterinary, Agricultural and Food Engineering*, vol. 8, no. 8, pp. 752–756, 2014.

[12] Y. Dong, W. Sawin, and Y. Bengio, "Hnhn: Hypergraph networks with hyperedge neurons," *arXiv preprint arXiv:2006.12278*, 2020.

[13] S.-e. Yoon, H. Song, K. Shin, and Y. Yi, "How much and when do we need higher-order information in hypergraphs? a case study on hyperedge prediction," in *WWW*, 2020.

[14] T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, "Reciprocity of weighted networks," *Scientific reports*, vol. 3, no. 1, pp. 1–9, 2013.

[15] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[16] D. Cirkovic, T. Wang, and S. Resnick, "Preferential attachment with reciprocity: Properties and estimation," *arXiv preprint arXiv:2201.03769*, 2022.

[17] G. Lee, J. Ko, and K. Shin, "Hypergraph motifs: Concepts, algorithms, and discoveries," *PVLDB*, vol. 13, no. 12, pp. 2256–2269, 2020.

[18] G. Lee and K. Shin, "Thyme+: Temporal hypergraph motifs and fast algorithms for exact counting," in *ICDM*, 2021.

[19] A. R. Benson, R. Kumar, and A. Tomkins, "Sequences of sets," in *KDD*, 2018.

[20] H. Choo and K. Shin, "On the persistence of higher-order interactions in real-world hypergraphs," in *SDM*, 2022.

[21] J. Leskovec, *Dynamics of large networks.* Carnegie Mellon University, 2008.

[22] N. Yadati, T. Gao, S. Asoodeh, P. Talukdar, and A. Louis, "Graph neural networks for soft semi-supervised learning on hypergraphs," in *PAKDD*, 2021.

[23] S. Kim, M. Choe, J. Yoo, and K. Shin, "Reciprocity in directed hypergraphs: Measures, findings, and generators (extended version)," *arXiv preprint arXiv:2210.05328*, 2022.

[24] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information theory*, vol. 37, no. 1, pp. 145–151, 1991.