# Uniqueness of Edge Entropy on the Bookkeeping Graph: An Axiomatization under Double-Entry Structure*

Pierre Jinghong Liang
Accounting AI Research Lab
Tepper School of Business
Carnegie Mellon University
liangj@andrew.cmu.edu

July 2026

**Abstract**

Double-entry bookkeeping accumulates a period's journal entries into a directed, weighted graph on the chart of accounts, whose off-diagonal mass records the flow between accounts. We ask what it means to summarize the structure of that flow with a single scalar, and answer the question axiomatically. On the natural domain—zero-diagonal nonnegative matrices taken modulo simultaneous row-column reordering, the quotient induced by the arbitrariness of chart-of-accounts ordering—we show that three conditions (normalization, continuity, and a node-merge grouping identity) determine a unique measure: the Shannon entropy of the edge distribution. Following Faddeev [1956], reordering invariance is placed in the domain rather than imposed as an axiom; a minimality argument then shows that this invariance does the essential work in forcing the four aggregation channels of a node merge to share a single information function, so no symmetry axiom is needed. The characterization is a genuine extension of the classical probability-simplex result rather than a restatement of it: the grouping axiom's structure encodes features specific to double-entry with no counterpart on the simplex—the directional debit-credit distinction, the additive aggregation of flows when accounts are combined, and the elimination and rescaling of intercompany flows under consolidation. A self-loop-permissive variant preserves intercompany flows as merged-node self-edges and yields the decomposition $H^{\circlearrowleft} = (1-\tau)\, H^{\text{edge}} + H_2(\tau, 1-\tau)$, whose gap isolates the consolidation convention's contribution. We close with a proposed application: a node-merge entropy statistic that scores the informational quality of a classification scheme by comparing within-classification merges against random ones.

*This paper serves as a self-contained expanded treatment of the axiomatic foundation for an entropy-based measure applied to edges of a bookkeeping graph, a concept used in the working paper "Accounting Graph Entropy (AGE)" by Liang, Pyo, Zhang, and Zhang [Liang et al., 2026], written at the Accounting AI Research Lab at the Tepper School of Business at Carnegie Mellon University. All errors are mine.

# 1 Introduction and Motivation

Double-entry bookkeeping records each transaction as a set of balanced debits and credits posted across accounts. Accumulated over a reporting period, these postings form a directed, weighted graph on the chart of accounts: accounts are the nodes, and each journal flow is a directed edge whose weight is the cumulative amount crediting one account and debiting another. The diagonal is empty (i.e., an atomic account does not transact with itself) so the object of interest is a nonnegative matrix $W$ with zero diagonal, the off-diagonal entry $w_{ij}$ recording the gross flow from account $i$ to account $j$. That double-entry carries this linear-algebraic structure is classical [Arya et al., 2000]; we take the graph as given and ask a measurement question about it.

We want to summarize the structure of this flow with a single scalar, and two features of the setting constrain what such a summary may depend on. First, the order in which accounts are listed is an artifact of record-keeping, not information: reordering the accounts (i.e., permuting the rows and columns of $W$ simultaneously) must leave the summary unchanged. Second, accounts are aggregates, and the summary must behave coherently when they are combined. Merging two accounts into one aggregates their flows additively, and does so in both directions at once: it sums the merged node's outflows across its row and its inflows down its column in a single operation. The natural domain for the scalar-summarization question is therefore not $W$ itself but its equivalence class under simultaneous row-column reordering, with this additive node-merge as the operation under which behavior is constrained.

Our main result is that these requirements, made precise, admit essentially one answer. On the domain of zero-diagonal nonnegative matrices modulo simultaneous reordering, exactly one measure satisfies three conditions: a normalization, continuity, and a grouping identity governing the additive node-merge. This unique measure is the Shannon entropy of the edge distribution, $H^{\text{edge}}(W) = -\sum_{i \neq j} \pi_{ij} \log \pi_{ij}$. The conditions are individually unremarkable; that together they leave no freedom is the content of the theorem. Following Faddeev [1956], we place reordering invariance in the domain rather than among the axioms, and a minimality argument shows that this invariance does the essential work in forcing the four aggregation channels of a node merge (i.e., outflow, inflow, bilateral, and recursive) to share one information function. No symmetry axiom is imposed; the shared function is a consequence of the domain rather than a separate axiom.

This is a substantive extension of the classical characterization on the probability simplex, not its restatement. The off-diagonal entries of $W$ do form a distribution, and one might expect the simplex result to transfer verbatim; it does not, because the admissible operation differs. The node-merge grouping identity is strictly more restrictive than arbitrary grouping on the simplex, and its structure encodes features specific to double-entry that have no vector-setting counterpart: the directional debit-credit distinction, preserved because the domain quotients only by reordering and never identifies $w_{ij}$ with $w_{ji}$; the additive aggregation of flows when accounts are combined; and the elimination and rescaling of intercompany flows when merged accounts are consolidated, reflecting the intercompany-elimination convention of consolidated reporting. These structural properties are developed in Liang [2026]; here they enter through the axioms and are recovered, in Section 5, as the accounting content the characterization turns out to require.

The paper proceeds as follows. Section 2 provides the graph, the domain, and the definition of edge entropy. Section 3 defines the node-merge operation and derives its grouping identity. Section 4 states the three axioms and proves that edge entropy is their unique solution. Section 5 identifies where the structural features of double-entry enter the characterization—at the level of the domain and the merge operation rather than the axioms. Section 6 develops the self-loop-permissive variant and the consolidation bridge identity. Section 7 proposes an application to scoring classification schemes, and Section 8 positions the result against the classical characterization and the single-sided

ACE measure [Li et al., 2025]. Section 9 delimits what the theorem does and does not establish and collects extensions.

# 2 Preliminaries

## 2.1 The Bookkeeping Graph

Let $\mathcal{G} = (V, E, W)$ denote a directed weighted graph on $m$ nodes, with:

- $V = \{v_1, v_2, \ldots, v_m\}$ the set of nodes, interpreted as accounts in the chart of accounts;
- $E \subseteq V \times V$ the set of directed edges, with $e_{ij}$ representing a journal-entry flow crediting account $i$ and debiting account $j$;
- $W = (w_{ij}) \in \mathbb{R}_{\geq 0}^{m \times m}$ the weighted adjacency matrix, with $w_{ij} \geq 0$ the cumulative dollar weight of flow from $i$ to $j$, and $w_{ii} = 0$ for all $i$ (no self-transactions on atomic accounts).

Denote by $\mathcal{W}_m^0$ the set of $m \times m$ non-negative matrices with zero diagonal. Total edge mass is $S(W) = \sum_{i \neq j} w_{ij}$, assumed positive. The normalized edge distribution is $\pi_{ij} = w_{ij}/S(W)$ for $i \neq j$.[1]

## 2.2 Node Reordering and the Domain of Edge Entropy

The chart-of-accounts ordering is an artifact of recordkeeping practice, not informational content: a balance sheet that lists Cash before Receivables conveys the same information as one that lists them in the reverse order. The natural object of analysis is therefore not the matrix $W$ itself but its equivalence class under *simultaneous row-column permutation*—the operation that reorders nodes consistently in both the inflow (column) and outflow (row) dimensions of the adjacency matrix. (In graph-theoretic terms this is node *relabeling*; the quotient defined below is the set of isomorphism classes of weighted directed graphs.)

Formally, define the equivalence relation $\sim$ on $\mathcal{W}_m^0$ by

$$W \sim W' \quad \Longleftrightarrow \quad \exists\, \sigma \in S_m \text{ such that } w'_{ij} = w_{\sigma(i),\sigma(j)} \text{ for all } i, j.$$

The quotient $\widetilde{\mathcal{W}}_m^0 := \mathcal{W}_m^0 / \sim$ is the space of node-labeled-up-to-reordering adjacency matrices.

This is the right domain on which to define entropy measures of the bookkeeping graph for two reasons. First, simultaneous row-column permutation respects the directed structure of journal-entry flows: it preserves which entries lie above the diagonal, which lie below, and which are bilateral. Arbitrary permutation of the $m(m-1)$ off-diagonal cells—which would let one swap $w_{ij}$ with $w_{ji}$ freely—would destroy this directionality and is therefore not an admissible symmetry on the bookkeeping graph. Second, simultaneous row-column permutation is precisely the symmetry generated by node-reordering: it quotients out the arbitrary chart-of-accounts ordering while preserving every other structural feature.

[1] A reader familiar with the linear-algebraic representation of double-entry [Arya et al., 2000] may ask why we build on the adjacency matrix rather than the signed incidence matrix on which that literature rests. The reason is intrinsic to the object we study. An information measure requires a nonnegative, normalizable quantity, and the incidence matrix—whose entries carry transaction direction as sign, and whose transaction rows sum to zero—admits no such normalization. The adjacency matrix carries direction as *position*, the ordered pair $i \to j$, rather than as sign; every entry is then nonnegative and yields the edge distribution $\pi$ on which $H^{\text{edge}}$ is defined. The two representations answer different questions—the signed incidence matrix supports transaction-level inference, redundancy, and error correction through its four fundamental subspaces; the adjacency matrix supports the informational summary of flow structure—and no comparison between them is intended here.

We therefore frame all subsequent results as statements about functions on $\widetilde{\mathcal{W}}_m^0$, treating node-reordering invariance as a property of the domain rather than an axiom.

### 2.3 Edge Entropy

**Definition 1** (Shannon edge entropy)**.** For $W \in \mathcal{W}_m^0$ with $S(W) > 0$, the Shannon edge entropy [Shannon, 1948] is

$$H_m^{\text{edge}}(W) \;=\; -\sum_{i \neq j} \pi_{ij} \log \pi_{ij}, \tag{1}$$

where the sum is over ordered pairs with $i \neq j$, and log denotes $\log_2$ throughout the paper.

By construction, $H_m^{\text{edge}}$ is a function on the quotient $\widetilde{\mathcal{W}}_m^0$: it depends only on the multiset of edge weights, not on the row-column labeling. We do *not* claim invariance under arbitrary permutation of the $m(m-1)$ off-diagonal cells, which would be a stronger and accounting-inadmissible symmetry. The relevant invariance is exactly node-reordering, built into the domain.

## 3 The Node-Merge Operation and the Grouping Identity

The natural grouping operation in the adjacency-matrix setting is not pairwise merging of edges (which lacks accounting interpretation) but merging of two nodes into one, with the incident edges aggregated accordingly.

### 3.1 Node-Merge: Zero-Diagonal Convention

**Definition 2** (Node-merge, zero-diagonal)**.** For $W \in \mathcal{W}_m^0$ and $a, b \in \{1, \ldots, m\}$ with $a \neq b$, the *node-merge* operation $W \mapsto W^{(a,b)}$ produces a matrix in $\mathcal{W}_{m-1}^0$ as follows. Let $c$ denote the merged node. Then:

$$\begin{aligned}
w_{cj}^{(a,b)} &= w_{aj} + w_{bj} \quad \text{for } j \notin \{a, b\}, \\
w_{ic}^{(a,b)} &= w_{ia} + w_{ib} \quad \text{for } i \notin \{a, b\}, \\
w_{ij}^{(a,b)} &= w_{ij} \quad \text{for } i, j \notin \{a, b\}, \\
w_{cc}^{(a,b)} &= 0.
\end{aligned}$$

The bilateral mass $w_{ab} + w_{ba}$ is eliminated; total edge mass becomes $S' = S(W) - w_{ab} - w_{ba}$.

The zero-diagonal convention $w_{cc}^{(a,b)} = 0$ corresponds to a pervasive account-aggregation operation in conventional bookkeeping practice, including *intercompany elimination* in consolidated financial reporting: transactions between entities that become a single reporting unit post-merger vanish from the consolidated statements.[2]

[2]Intercompany elimination is the instance with an institutional name, but the same netting arises wherever accounts are aggregated, at any level of the reporting hierarchy and with no consolidation involved. Merging Cash and Inventory into a single Current Assets node internalizes cash purchases of inventory—a flow between the merged accounts—which the zero-diagonal convention nets out of the aggregated graph; rolling sub-ledger accounts (raw materials, work-in-progress, finished goods) into a single Inventory account likewise internalizes the transfers among them (Section 6). Node-merge is thus the graph form of account aggregation generally—sub-ledger to account, account to line item, entity to consolidated group—which is what qualifies it, rather than arbitrary cell grouping, as the accounting-native grouping operation. The additive aggregation underlying the merge is the linearity law of the companion paper [Liang, 2026].

Define, for the merge $(a, b) \to c$:

$$\begin{aligned} \tau &= \pi_{ab} + \pi_{ba} \qquad \text{(bilateral mass fraction)}, \\ p_j^{\text{out}} &= \pi_{aj} + \pi_{bj} \qquad \text{(row-merge mass for external node } j\text{)}, \\ p_i^{\text{in}} &= \pi_{ia} + \pi_{ib} \qquad \text{(column-merge mass for external node } i\text{)}. \end{aligned}$$

We denote by $H_2(x, 1-x) = -x \log_2 x - (1-x) \log_2 (1-x)$ the binary Shannon entropy, abbreviated $H_2(x)$ where the second argument is clear.

**Theorem 3** (Node-merge grouping identity, zero-diagonal)**.** *For any $W \in \mathcal{W}_m^0$ with $S(W) > 0$ and any pair $a \neq b$,*

$$\begin{aligned} H_m^{edge}(W) \;=\; & (1-\tau)\, H_{m-1}^{edge}(W^{(a,b)}) \;+\; H_2(\tau, 1-\tau) \\ & + \sum_{j \notin \{a,b\}} p_j^{out}\, H_2\Big(\tfrac{\pi_{aj}}{p_j^{out}}, \tfrac{\pi_{bj}}{p_j^{out}}\Big) \\ & + \sum_{i \notin \{a,b\}} p_i^{in}\, H_2\Big(\tfrac{\pi_{ia}}{p_i^{in}}, \tfrac{\pi_{ib}}{p_i^{in}}\Big) \\ & + \tau\, H_2\big(\tfrac{\pi_{ab}}{\tau}, \tfrac{\pi_{ba}}{\tau}\big)\,. \end{aligned} \tag{2}$$

*Proof.* Partition the $m(m-1)$ off-diagonal cells of $W$ into four disjoint classes:

1. *Untouched external cells*: $(i, j)$ with $i, j \notin \{a, b\}$ and $i \neq j$. These pass through the merge unchanged.
2. *Row-merge pairs*: for each $j \notin \{a, b\}$, the pair $\{(a, j), (b, j)\}$ collapses to a single cell $(c, j)$ of mass $p_j^{\text{out}}$.
3. *Column-merge pairs*: for each $i \notin \{a, b\}$, the pair $\{(i, a), (i, b)\}$ collapses to a single cell $(i, c)$ of mass $p_i^{\text{in}}$.
4. *Bilateral pair*: $\{(a, b), (b, a)\}$, which is *removed* from the distribution under the zero-diagonal convention.

For any partition of a probability distribution $\{\pi_k\}$ into groups $G_1, \ldots, G_K$ with group masses $P_\ell = \sum_{k \in G_\ell} \pi_k$, the standard grouping identity gives

$$-\sum_k \pi_k \log \pi_k \;=\; -\sum_\ell P_\ell \log P_\ell \;+\; \sum_\ell P_\ell \left( -\sum_{k \in G_\ell} \tfrac{\pi_k}{P_\ell} \log \tfrac{\pi_k}{P_\ell} \right). \tag{3}$$

Apply (3) with the four classes above as groups. The untouched external cells are singleton groups: each contributes its raw entropy to the group-mass term and nothing to the within-group term. The row-merge, column-merge, and bilateral classes contribute their group masses to the former and mass-weighted binary entropies to the latter. Written out in full for this instance, (3) reads

$$\begin{aligned} H_m^{\text{edge}}(W) \;=\; & \underbrace{ -\sum_{\substack{i,j \notin \{a,b\} \\ i \neq j}} \pi_{ij} \log \pi_{ij} \;-\; \sum_{j \notin \{a,b\}} p_j^{\text{out}} \log p_j^{\text{out}} \;-\; \sum_{i \notin \{a,b\}} p_i^{\text{in}} \log p_i^{\text{in}} \;-\; \tau \log \tau }_{\text{group-mass entropy}} \\ & + \underbrace{ \sum_{j \notin \{a,b\}} p_j^{\text{out}}\, H_2\Big(\tfrac{\pi_{aj}}{p_j^{\text{out}}}, \tfrac{\pi_{bj}}{p_j^{\text{out}}}\Big) \;+\; \sum_{i \notin \{a,b\}} p_i^{\text{in}}\, H_2\Big(\tfrac{\pi_{ia}}{p_i^{\text{in}}}, \tfrac{\pi_{ib}}{p_i^{\text{in}}}\Big) \;+\; \tau\, H_2\big(\tfrac{\pi_{ab}}{\tau}, \tfrac{\pi_{ba}}{\tau}\big) }_{\text{within-group entropy}}. \end{aligned} \tag{4}$$

The within-group component already matches the last three terms of (2). It remains to show that the group-mass entropy equals $(1-\tau)\, H_{m-1}^{\text{edge}}(W^{(a,b)}) + H_2(\tau, 1-\tau)$.

Using the post-merge renormalization $\pi_{ij}^{(a,b)} = \pi_{ij}/(1-\tau)$ for external cells and $\pi_{cj}^{(a,b)} = p_j^{\text{out}}/(1-\tau)$, $\pi_{ic}^{(a,b)} = p_i^{\text{in}}/(1-\tau)$, and noting that the surviving pre-merge masses sum to $1-\tau$,

$$H_{m-1}^{\text{edge}}(W^{(a,b)}) \;=\; \frac{1}{1-\tau}\left[-\sum_{\substack{i,j\notin\{a,b\}\\ i\neq j}} \pi_{ij}\log\pi_{ij} - \sum_j p_j^{\text{out}}\log p_j^{\text{out}} - \sum_i p_i^{\text{in}}\log p_i^{\text{in}}\right] + \log(1-\tau).$$

Multiplying through by $(1-\tau)$ shows that the first three pieces of the group-mass entropy in (4) equal $(1-\tau)\, H_{m-1}^{\text{edge}}(W^{(a,b)}) - (1-\tau)\log(1-\tau)$. Combining with the fourth piece via the binary-entropy identity

$$-\tau\log\tau \;-\; (1-\tau)\log(1-\tau) \;=\; H_2(\tau, 1-\tau)$$

yields group-mass entropy $= (1-\tau)\, H_{m-1}^{\text{edge}}(W^{(a,b)}) + H_2(\tau, 1-\tau)$, and hence (2). □

**Example 4** (Three-node sanity check)**.** Let $m = 3$ with $\pi_{12} = \pi_{21} = 0.1$, $\pi_{13} = \pi_{31} = 0.2$, $\pi_{23} = \pi_{32} = 0.2$. Total is 1 by construction.

Direct computation: $H_3^{\text{edge}}(W) = -[2(0.1)\log_2 0.1 + 4(0.2)\log_2 0.2] \approx 2.522$ bits.

Merge $\{2,3\}$: $\tau = 0.4$, $p_1^{\text{out}} = 0.3$, $p_1^{\text{in}} = 0.3$. Post-merge graph has two edges, both renormalized to probability 0.5, so $H_2^{\text{edge}}(W^{(2,3)}) = 1$ bit.

Row-split binary: $H_2(1/3, 2/3) \approx 0.918$. Same for column-split. Bilateral binary: $H_2(1/2, 1/2) = 1$.

Right-hand side of (2):

$$0.6 \cdot 1 + H_2(0.4, 0.6) + 0.3 \cdot 0.918 + 0.3 \cdot 0.918 + 0.4 \cdot 1 \approx 0.6 + 0.971 + 0.275 + 0.275 + 0.4 = 2.521.$$

Agreement to three decimals. □

# 4 Axiomatic Characterization

We now consider an arbitrary sequence of functions $\{J_m\}_{m\geq 2}$, where each $J_m : \widetilde{\mathcal{W}}_m^0 \to \mathbb{R}$ assigns a real number to each node-reordering-equivalence class of zero-diagonal adjacency matrices, and ask: under what axioms must $J_m$ coincide with $H_m^{\text{edge}}$?

Because the domain $\widetilde{\mathcal{W}}_m^0$ is the quotient under simultaneous row-column permutation, node-reordering invariance is built into the formulation rather than imposed as a separate axiom. This parallels the Faddeev [1956] treatment of the classical case, in which entropy is defined on the space of probability distributions on $m$ atoms (rather than on ordered tuples in $\mathbb{R}^m$), so that symmetry is a property of the domain rather than an axiom about the function. The Cover and Thomas [2006] presentation derives symmetry from the grouping axiom; we adopt the cleaner Faddeev convention because it isolates more sharply the axiom that does the substantive work—the grouping axiom—from the structural choice of domain.

## 4.1 The Axioms

Denote by $J_2^{\text{bin}}(x) := J_2(\pi)$ where $\pi$ is the two-edge distribution with probabilities $x$ and $1-x$. (On $\widetilde{\mathcal{W}}_2^0$, the two off-diagonal cells are exchanged by node-reordering, so $J_2^{\text{bin}}(x) = J_2^{\text{bin}}(1-x)$ automatically.)

**Axiom 1** (Normalization)**.** $J_2^{\mathrm{bin}}(1/2) = 1$.

**Axiom 2** (Continuity)**.** $J_2^{\mathrm{bin}}$ is continuous on $(0,1)$.

**Axiom 3** (Node-merge grouping)**.** For any $m \geq 3$, any $W \in \mathcal{W}_m^0$ with $S(W) > 0$, and any pair $a \neq b$,

$$J_m(W) \;=\; (1-\tau)\, J_{m-1}(W^{(a,b)}) \;+\; \underbrace{J_2^{\mathrm{bin}}(\tau)}_{\text{bilateral-elimination}} \;+\; \underbrace{\sum_{j \notin \{a,b\}} p_j^{\mathrm{out}}\, J_2^{\mathrm{bin}}\!\left(\tfrac{\pi_{aj}}{p_j^{\mathrm{out}}}\right)}_{\text{row-split}} \;+\; \underbrace{\sum_{i \notin \{a,b\}} p_i^{\mathrm{in}}\, J_2^{\mathrm{bin}}\!\left(\tfrac{\pi_{ia}}{p_i^{\mathrm{in}}}\right)}_{\text{column-split}} \;+\; \underbrace{\tau\, J_2^{\mathrm{bin}}\!\left(\tfrac{\pi_{ab}}{\tau}\right)}_{\text{recursive base}}, \tag{5}$$

with a *single* binary information function $J_2^{\mathrm{bin}}$ occupying all four named roles, and terms with zero group mass read as zero.

Identity (5) is the grouping identity of Theorem 3 with $J$'s in place of $H^{\mathrm{edge}}$'s. At $m = 2$ the graph consists of the bilateral pair alone ($\tau = 1$) and the identity degenerates to its final term, which is why that slot is the recursion's base.

Three remarks on the axioms.

*First*, the axiomatization is structurally parallel to Faddeev's: normalization, continuity, and a single grouping axiom. Relative to the Cover–Thomas axiom set, monotonicity is not assumed; normalization plays its scale-fixing role, as in Faddeev's treatment, while grouping takes the node-merge form.

*Second*, no symmetry axiom appears in the list. Node-reordering invariance is built into the domain (Section 2); no further symmetry between row-cells and column-cells, or between $\pi_{ij}$ and $\pi_{ji}$, is imposed. This is essential: imposing such a symmetry on the matrix would conflate the credit side with the debit side at the matrix-cell level, foreclosing on debit-credit duality [Ijiri, 1975] as a substantive constraint on the entropy measure.

*Third*, the single-function clause of Axiom 3 is stated for definiteness. In principle, one could admit four distinct binary functions—$\phi_{\mathrm{bil}}, \phi_{\mathrm{row}}, \phi_{\mathrm{col}}, \phi_{\mathrm{base}}$—in the four roles and still obtain a well-defined grouping rule. One might therefore expect the single-function clause to carry substantive axiomatic content, specifically encoding debit-credit duality at the axiom level. Proposition 9 below establishes that this expectation is wrong: under Axioms 1 and 2 and the weakened Axiom 3 admitting four distinct role functions (with the regularity hypotheses stated there), node-reordering invariance forces the four functions to coincide. The single-function clause of Axiom 3 is therefore redundant: no separate axiomatic commitment to duality is needed.

### 4.2 A Sufficient-Determination Lemma

The following lemma separates two tasks in the uniqueness argument: pinning down the functional form of $J_2^{\mathrm{bin}}$, and propagating that determination to $J_m$ on the full domain $\widetilde{\mathcal{W}}_m^0$. It is valid for any $J_2^{\mathrm{bin}}$, not just $H_2$.

**Lemma 5** (Sufficiency of $J_2^{\mathrm{bin}}$)**.** *Suppose $\{J_m\}_{m\geq 2}$ satisfies Axioms 1, 2, 3, and suppose $J_2^{bin} : (0,1) \to \mathbb{R}$ is determined. Then $J_m(W)$ is determined for every $m \geq 2$ and every $W \in \widetilde{\mathcal{W}}_m^0$ with $S(W) > 0$.*

*Proof.* Induction on $m$.

*Base* ($m = 2$). For $W \in \widetilde{\mathcal{W}}_2^0$, the edge distribution is $(x, 1-x)$ for some $x \in (0,1)$, and $J_2(W) = J_2^{\text{bin}}(x)$ by the definition of $J_2^{\text{bin}}$. Determined.

*Inductive step* ($m > 2$). Fix $W \in \widetilde{\mathcal{W}}_m^0$ with $S(W) > 0$, and choose any pair $a \neq b$. Identity (5) expresses $J_m(W)$ through $J_{m-1}(W^{(a,b)})$—determined by the inductive hypothesis—and evaluations of $J_2^{\text{bin}}$ at explicit arguments in $(0,1)$, each determined by hypothesis. Hence $J_m(W)$ is determined. □

**Remark 6** (Consistency of Axiom 3)**.** Axiom 3 prescribes a value for $J_m(W)$ for each choice of pair $(a, b)$. For Axiom 3 to be satisfiable at all, all pair choices must prescribe the same value. This consistency is guaranteed once $J_2^{\text{bin}} = H_2$ by Theorem 3, which establishes that $H_m^{\text{edge}}$ satisfies the identity for every $(a, b)$. The Sufficiency Lemma therefore determines $J_m(W)$ consistently: the right-hand side of Axiom 3, evaluated with $J_2^{\text{bin}} = H_2$ and with $J_{m-1} = H_{m-1}^{\text{edge}}$ by induction, equals $H_m^{\text{edge}}(W)$ independent of the pair choice.

### 4.3 The Uniqueness Theorem

**Theorem 7** (Uniqueness of edge entropy)**.** *Suppose* $\{J_m\}_{m\geq 2}$ *satisfies Axioms 1–3. Then for every* $m \geq 2$ *and every* $W \in \mathcal{W}_m^0$ *with* $S(W) > 0$*,*

$$J_m(W) \;=\; H_m^{edge}(W).$$

*Proof.* The proof proceeds in three steps: (i) derive a functional equation on $J_2^{\text{bin}}$ by applying Axiom 3 to a three-node family in two distinct ways; (ii) solve the functional equation using Axioms 1 and 2; (iii) apply Lemma 5 to extend the determination to all of $\widetilde{\mathcal{W}}_m^0$.

*Step 1: Functional equation from* $m = 3$. We derive the functional equation on the direction-symmetric 3-node family $\pi_{ij} = \pi_{ji}$. The restriction to this family is a dimensional convenience, not an accounting commitment: the family's 2-parameter freedom matches the argument dimensionality of the Faddeev equation, and its direction-symmetry sends both the recursion's post-merge argument and the recursive-base argument to $1/2$, where normalization immediately fixes them. The accounting content—debit-credit duality—does not enter at this step; it enters through the domain's node-reordering invariance, as made explicit in Proposition 9 below. Lemma 5 establishes that the determination of $J_2^{\text{bin}}$ via the symmetric-family equation extends to $J_m$ on all $W \in \widetilde{\mathcal{W}}_m^0$, symmetric or not.

Consider the three-node family parameterized by $(a, b, c) \in \mathbb{R}_{\geq 0}^3$ with $2a + 2b + 2c = 1$, defined by

$$\pi_{12} = \pi_{21} = a, \quad \pi_{13} = \pi_{31} = b, \quad \pi_{23} = \pi_{32} = c.$$

Write $J_3(\pi)$ for the common value of $J_3$ on matrices with edge distribution $\pi$, and $\phi := J_2^{\text{bin}}$, a function of a single argument in $(0,1)$, with $\phi(x) = \phi(1-x)$ by the domain symmetry and $\phi(1/2) = 1$ by Axiom 1.

*Merging* $\{2,3\}$*:* $\tau = 2c$, $p_1^{\text{out}} = a + b$, $p_1^{\text{in}} = a + b$. The post-merge graph has two edges, each of mass $a+b$ renormalized by $1-2c = 2(a+b)$ to probability $(a+b)/(2(a+b)) = 1/2$, so the recursive term evaluates to $\phi(1/2) = 1$. The bilateral split is $c/(2c) = 1/2$, likewise contributing $\phi(1/2) = 1$.

Applying Axiom 3, with terms in the order of (2):

$$J_3(\pi) = (1-2c)\cdot 1 \;+\; \phi(2c) \;+\; 2(a+b)\,\phi\Big(\tfrac{a}{a+b}\Big) \;+\; 2c\cdot 1, \tag{6}$$

where the row-split and column-split contributions coincide on this direction-symmetric family (each equals $(a+b)\,\phi(a/(a+b))$; hence the factor 2).

*Merging* $\{1,2\}$*:* $\tau' = 2a$, and the analogous computation gives

$$J_3(\pi) = (1-2a)\cdot 1 \;+\; \phi(2a) \;+\; 2(b+c)\,\phi\Big(\tfrac{b}{b+c}\Big) \;+\; 2a\cdot 1. \tag{7}$$

Equating (6) and (7) and simplifying (using $a+b+c=1/2$):

$$\phi(2c) + (1-2c)\,\phi\Big(\tfrac{a}{a+b}\Big) \;=\; \phi(2a) + (1-2a)\,\phi\Big(\tfrac{b}{b+c}\Big). \tag{8}$$

Substitute $x := 2a, y := 2c$, so $a = x/2, c = y/2, b = (1-x-y)/2$. Then $a/(a+b) = x/(1-y)$ and $b/(b+c) = (1-x-y)/(1-x)$. Using $\phi(1-z) = \phi(z)$, $\phi((1-x-y)/(1-x)) = \phi(y/(1-x))$. Equation (8) becomes

$$\phi(y) + (1-y)\,\phi\Big(\tfrac{x}{1-y}\Big) \;=\; \phi(x) + (1-x)\,\phi\Big(\tfrac{y}{1-x}\Big), \tag{9}$$

valid for all $(x,y)$ with $x,y>0$ and $x+y<1$. As $(a,b,c)$ range over the open 2-simplex, the substitutions $x=2a, y=2c$ range over the full open triangle $\{(x,y) : x,y>0, x+y<1\}$, which is the standard domain of Faddeev's equation.

*Step 2: Solve* (9). Equation (9) is Faddeev's (1956) fundamental functional equation for binary entropy. The unique continuous solution satisfying $\phi(1/2)=1$ and $\phi(x)=\phi(1-x)$ is

$$\phi(x) \;=\; -x\log_2 x - (1-x)\log_2(1-x) \;=\; H_2(x,1-x).$$

Standard references are Aczél and Daróczy [1975, Ch. 3] and Khinchin [1957]; the Appendix (Theorem 16) states the theorem in the form used here and records its reformulation as a grouping identity.

*Step 3: Extend to $m \geq 2$ via the Sufficiency Lemma.* By Step 2, $J_2^{\mathrm{bin}} = H_2$ on $(0,1)$. Apply Lemma 5: $J_m(W)$ is determined for every $m \geq 2$ and every $W \in \widetilde{\mathcal{W}}_m^0$.

To identify the determined value as $H_m^{\mathrm{edge}}(W)$, proceed by induction. For $m=2$: $J_2(W) = J_2^{\mathrm{bin}}(\pi_{12}) = H_2(\pi_{12}) = H_2^{\mathrm{edge}}(W)$ directly. For $m>2$: the inductive hypothesis $J_{m-1} = H_{m-1}^{\mathrm{edge}}$ together with $J_2^{\mathrm{bin}} = H_2$ makes the right-hand side of Axiom 3's identity coincide arithmetically with the right-hand side of Theorem 3's identity applied to $H_m^{\mathrm{edge}}(W)$. By the consistency of Axiom 3 (guaranteed by Theorem 3 as satisfiability witness), either side equals $H_m^{\mathrm{edge}}(W)$. Hence $J_m(W) = H_m^{\mathrm{edge}}(W)$ for all $m$ and all $W$. This completes the proof. □

### 4.4 Axiomatic Minimality

The single-function clause of Axiom 3 requires that one binary function $J_2^{\mathrm{bin}}$ governs all four binary-entropy roles on the right-hand side of the grouping identity: bilateral-elimination, row-split, column-split, and the recursive base. A natural question is whether this clause is necessary for uniqueness, or whether the other axioms together with the domain symmetry already force the four roles to be served by a single function. The following result shows the latter; for a systematic treatment of functional equations involving several unknown functions, see Kannappan [2009].

**Remark 8** (Role symmetry from the domain)**.** For three of the four roles, symmetry of the role function need not be assumed—it follows from well-definedness on the quotient. Exchanging the labels of $a$ and $b$ is a node reordering: it fixes the equivalence class of $W$, hence every $J$-value in the

weakened identity, while replacing the row-split arguments $\pi_{aj}/p_j^{\text{out}}$ by $\pi_{bj}/p_j^{\text{out}} = 1-\pi_{aj}/p_j^{\text{out}}$, and likewise for the column-split and recursive-base arguments. Choosing families in which only one role carries a non-degenerate split isolates the roles one at a time and forces $\phi_{\text{row}}(u) = \phi_{\text{row}}(1-u)$, $\phi_{\text{col}}(u) = \phi_{\text{col}}(1-u)$, and $\phi_{\text{base}}(u) = \phi_{\text{base}}(1-u)$. The bilateral-elimination argument $\tau$, by contrast, is invariant under every node reordering—no reordering exchanges $\tau$ with $1-\tau$—so the symmetry of $\phi_{\text{bil}}$ is a genuine hypothesis of Proposition 9, as is the normalization $\phi(1/2) = 1$ of each role function (a per-role strengthening of Axiom 1). Under the single-function Axiom 3 neither is an issue: the shared function is $J_2^{\text{bin}}$, whose symmetry is inherited from the two-node quotient and whose normalization is Axiom 1 itself.

**Proposition 9** (Axiomatic Minimality)**.** *Let $\{J_m\}_{m\geq 2}$ satisfy Axioms 1, 2, and the weakened form of Axiom 3 that admits possibly distinct continuous normalized binary functions $\phi_{bil}, \phi_{row}, \phi_{col}, \phi_{base}$ on $(0,1)$ in the four named roles of (5), each with finite one-sided limits at the endpoints, and with $\phi_{bil}$ symmetric ($\phi_{bil}(u) = \phi_{bil}(1-u)$); the symmetry of the remaining three role functions need not be assumed (Remark 8). Then*

$$\phi_{bil} = \phi_{row} = \phi_{col} = \phi_{base} = H_2.$$

*Consequently, the single-function clause of Axiom 3 is redundant: the weakened axiom admitting four binary functions is equivalent, under Axioms 1 and 2, to Axiom 3 as stated.*

*Proof.* We prove the coincidence of the four functions in four steps. Step 1 derives a two-function functional equation on the symmetric 3-node family. Step 2 uses the boundary limit $y \to 0^+$ and node-reordering symmetry to collapse the two-function equation to the classical Faddeev equation, yielding $\phi_{\text{bil}} = H_2$ and the sum $\Phi := \phi_{\text{row}} + \phi_{\text{col}} = 2H_2$. Step 3 uses a direction-asymmetric 3-node subfamily to force $\phi_{\text{row}} = \phi_{\text{col}}$ individually, hence both equal $H_2$. Step 4 uses fully asymmetric 3-node graphs to force $\phi_{\text{base}} = H_2$, pinning down at the same time the binary function $J_2^{\text{bin}}$ that the recursion bottoms out to.

*Step 1: Two-function equation on the symmetric family.* Apply the weakened Axiom 3 to the direction-symmetric 3-node family of the proof of Theorem 7. The symmetric family cannot separate $\phi_{\text{row}}$ from $\phi_{\text{col}}$, as the row-split and column-split arguments coincide at $a/(a+b)$. Equating merges $\{2,3\}$ and $\{1,2\}$ yields, after the substitutions $x = 2a$, $y = 2c$,

$$\phi_{\text{bil}}(y) + \tfrac{1-y}{2}\,\Phi\Big(\tfrac{x}{1-y}\Big) \;=\; \phi_{\text{bil}}(x) + \tfrac{1-x}{2}\,\Phi\Big(\tfrac{y}{1-x}\Big) \tag{10}$$

for all $(x,y)$ with $x, y > 0$ and $x + y < 1$, where $\Phi := \phi_{\text{row}} + \phi_{\text{col}}$.

*Step 2: $\phi_{bil} = \Phi/2 = H_2$.* Pass to the limit $y \to 0^+$ in (10). Let $\lambda := \phi_{\text{bil}}(0^+)$ and $\mu := \Phi(0^+)$, both finite by hypothesis. The limit gives

$$\lambda + \tfrac{1}{2}\Phi(x) \;=\; \phi_{\text{bil}}(x) + \tfrac{(1-x)\,\mu}{2},$$

so

$$\phi_{\text{bil}}(x) \;=\; \big(\lambda - \tfrac{\mu}{2}\big) + \tfrac{1}{2}\Phi(x) + \tfrac{\mu x}{2}.$$

Applying the symmetry hypothesis $\phi_{\text{bil}}(x) = \phi_{\text{bil}}(1-x)$ and the derived symmetry $\Phi(x) = \Phi(1-x)$ (Remark 8):

$$\tfrac{\mu x}{2} = \tfrac{\mu(1-x)}{2}, \quad \text{i.e.,} \quad \mu(2x-1) = 0 \text{ for all } x \in (0,1),$$

which forces $\mu = 0$. Then $\phi_{\text{bil}}(x) = \lambda + \frac{1}{2}\Phi(x)$. Evaluating at $x = 1/2$ with $\Phi(1/2) = \phi_{\text{row}}(1/2) + \phi_{\text{col}}(1/2) = 1 + 1 = 2$:

$$1 = \phi_{\text{bil}}(1/2) = \lambda + 1, \quad \text{so } \lambda = 0.$$

Therefore $\phi_{\text{bil}} = \Phi/2$. Substituting $\Phi = 2\phi_{\text{bil}}$ into (10):

$$\phi_{\text{bil}}(y) + (1-y)\,\phi_{\text{bil}}\Big(\tfrac{x}{1-y}\Big) \;=\; \phi_{\text{bil}}(x) + (1-x)\,\phi_{\text{bil}}\Big(\tfrac{y}{1-x}\Big),$$

which is the classical Faddeev equation. By Aczél and Daróczy [1975, Ch. 3], the unique continuous symmetric solution with $\phi_{\text{bil}}(1/2) = 1$ is $\phi_{\text{bil}} = H_2$. Hence $\Phi = 2H_2$.

*Step 3: $\phi_{row} = \phi_{col}$.* Write $\phi_{\text{row}} = H_2 + \eta$, $\phi_{\text{col}} = H_2 - \eta$ for a continuous symmetric $\eta$ on $(0,1)$ with $\eta(1/2) = 0$ and finite one-sided limits at the endpoints. Consider the direction-asymmetric 3-node family

$$\pi_{12} = \alpha, \quad \pi_{21} = \beta, \quad \pi_{13} = \pi_{31} = \pi_{23} = \pi_{32} = \gamma,$$

with $\alpha+\beta+4\gamma = 1$ and $\alpha \neq \beta$ (asymmetry between nodes 1 and 2 only). Apply the weakened Axiom 3 with $\phi_{\text{bil}} = H_2$ (Step 2) and $\phi_{\text{row}} = H_2 + \eta$, $\phi_{\text{col}} = H_2 - \eta$ to the merges $\{2,3\}$ and $\{1,3\}$; on this family the recursive-base slot enters both merges only at argument $1/2$, where the normalization hypothesis gives $\phi_{\text{base}}(1/2) = 1$, and the recursive terms coincide between the two merges (the post-merge edge distributions agree) and cancel. The $H_2$-occupied contributions coincide between the two merge expressions by the symmetry of the family—the two merges see identical bilateral fractions, identical recursive terms, and row/column split sums that agree term by term—so they cancel, leaving the residual $\eta$-constraint

$$(\beta+\gamma)\,\eta\Big(\tfrac{\beta}{\beta+\gamma}\Big) \;=\; (\alpha+\gamma)\,\eta\Big(\tfrac{\alpha}{\alpha+\gamma}\Big)$$

for all admissible $(\alpha, \beta, \gamma)$.

Using $\eta(u) = \eta(1-u)$: $\eta(\beta/(\beta+\gamma)) = \eta(\gamma/(\beta+\gamma))$. Setting $v_1 := \alpha+\gamma$, $v_2 := \beta+\gamma$, so $v_1 + v_2 = 1-2\gamma$ and $v_1, v_2 > \gamma$:

$$v_2\,\eta(\gamma/v_2) \;=\; v_1\,\eta(\gamma/v_1). \tag{11}$$

Define $g_\gamma(v) := v\,\eta(\gamma/v)$ on $v \in (\gamma, 1-3\gamma)$. Constraint (11) says $g_\gamma(v_1) = g_\gamma(v_2)$ whenever $v_1 + v_2 = 1-2\gamma$; i.e., $g_\gamma$ is symmetric around the midpoint $(1-2\gamma)/2$ of its domain.

Pass to the boundary $v \to \gamma^+$:

$$\lim_{v\to\gamma^+} g_\gamma(v) = \gamma\cdot\eta(1^-) = \gamma\cdot\eta(0^+),$$

using $\eta(1^-) = \eta(0^+)$ by $\eta$'s symmetry. The symmetric partner $v = 1-3\gamma$ gives $g_\gamma(1-3\gamma) = (1-3\gamma)\,\eta(\gamma/(1-3\gamma))$. Equality of the two endpoint values:

$$\gamma\cdot\eta(0^+) \;=\; (1-3\gamma)\,\eta\Big(\tfrac{\gamma}{1-3\gamma}\Big).$$

Let $t := \gamma/(1-3\gamma)$. As $\gamma$ ranges over $(0, 1/4)$, $t$ ranges over $(0,1)$, and the substitution $\gamma = t/(1+3t)$ yields

$$\eta(t) \;=\; t\cdot\eta(0^+) \qquad \text{for all } t \in (0,1).$$

Now apply $\eta(t) = \eta(1-t)$: $t\,\eta(0^+) = (1-t)\,\eta(0^+)$ for all $t$, which forces $\eta(0^+) = 0$. Hence $\eta \equiv 0$, so $\phi_{\text{row}} = \phi_{\text{col}} = H_2$.

*Step 4: $\phi_{base} = H_2$ and $J_2^{bin} = H_2$.* Steps 1–3 never evaluate $\phi_{\text{base}}$ away from $1/2$: every family used so far has a direction-symmetric bilateral pair in the merged position. Fully asymmetric 3-node graphs supply the missing constraint. Write $F := \phi_{\text{base}} - H_2$ for the residual of the recursive-base slot and $G := J_2^{\text{bin}} - H_2$ for the residual of the binary function that the recursion bottoms out to ($J_2$ itself; under the single-function clause the two coincide by definition, under the weakened

axiom they are a priori distinct). Both are continuous and symmetric on $(0,1)$ and vanish at $1/2$ (Axiom 1 for $G$, the normalization hypothesis for $F$). Encode them as 1-homogeneous symmetric functions on the positive quadrant:

$$D_F(m_1,m_2) \;:=\; (m_1+m_2)\,F\Big(\tfrac{m_1}{m_1+m_2}\Big), \qquad D_G(m_1,m_2) \;:=\; (m_1+m_2)\,G\Big(\tfrac{m_1}{m_1+m_2}\Big).$$

Consider a general 3-node distribution $\pi_{12}=p$, $\pi_{21}=q$, $\pi_{13}=r$, $\pi_{31}=s$, $\pi_{23}=v$, $\pi_{32}=w$, all positive and summing to one. Apply the weakened Axiom 3 to the merges $\{1,2\}$ and $\{1,3\}$, and subtract from each the corresponding instance of Theorem 3 for $H_3^{\text{edge}}$. Every slot occupied by $H_2$—bilateral-elimination, row-split, and column-split, by Steps 2–3—cancels, leaving only the recursion and recursive-base residuals:

$$D_G(r+v,\,s+w) \;+\; D_F(p,q) \;=\; D_G(p+w,\,q+v) \;+\; D_F(r,s), \tag{12}$$

where $(r+v,\,s+w)$ and $(p+w,\,q+v)$ are the post-merge edge-mass pairs of the two merges and $(p,q)$, $(r,s)$ their bilateral pairs. By 1-homogeneity the sum-to-one normalization is immaterial, so (12) holds for all positive mass 6-tuples.

Two specializations conclude. First, letting $v,w\to 0^+$ gives $D_G(r,s)-D_F(r,s)=D_G(p,q)-D_F(p,q)$: the difference $D_G-D_F$ is constant on the positive quadrant, and 1-homogeneity forces the constant to be zero, so $D_F=D_G=:D$—in particular $\phi_{\text{base}}=J_2^{\text{bin}}$, the base-case identification. Second, by the symmetry of $D$, $D(p+w,\,q+v)=D\big((q,p)+(v,w)\big)$, so (12) states that the increment $D\big(x+(v,w)\big)-D(x)$ is independent of the base pair $x$. Taking $x=(v,w)$ and using homogeneity, $D(2u)-D(u)=D(u)$, so the increment equals $D(v,w)$ itself and $D$ is additive: $D(x+u)=D(x)+D(u)$. With continuity, $D$ is linear, $D(m_1,m_2)=c_1m_1+c_2m_2$; symmetry gives $c_1=c_2=:\kappa$, hence $F(t)=D(t,1-t)=\kappa$ for all $t$, and $F(1/2)=0$ forces $\kappa=0$. Therefore $F=G\equiv 0$: $\phi_{\text{base}}=H_2$ and $J_2^{\text{bin}}=H_2$.

This completes the proof that all four binary functions coincide with $H_2$. Applying Lemma 5 with $J_2^{\text{bin}}=H_2$ (now determined under the weakened axioms), $J_m=H_m^{\text{edge}}$ on all of $\widetilde{\mathcal{W}}_m^0$. □

**Remark 10** (What carries the argument)**.** The proof of Proposition 9 invokes the symmetry $\phi(u)=\phi(1-u)$ of the role functions—derived from the domain quotient for three roles, hypothesized for $\phi_{\text{bil}}$ (Remark 8)—at four load-bearing steps: (i) forcing $\mu=0$ in the $y\to 0$ limit of Step 2; (ii) rewriting the $\eta$-constraint in terms of $\eta(\gamma/v)$ via $\eta(u)=\eta(1-u)$ in Step 3; (iii) forcing $\eta(0^+)=0$ at the final line of Step 3; (iv) the base-pair swap $D(p+w,\,q+v)=D\big((q,p)+(v,w)\big)$ in Step 4, which converts the two-merge comparison into base-point independence of the increments of $D$. Each use is essential. Without node-reordering symmetry, the two-function equation admits additional solutions (an anti-symmetric one-parameter family survives), and the $\eta$-constraint admits non-trivial solutions. Domain-level node-reordering invariance is therefore the mechanical source of the minimality result.

# 5 The Two-Level Locus of Accounting Content

The axiomatization above is structurally parallel to Faddeev [1956]. Its accounting content is located at two distinct levels, each of which does identifiable mathematical work in the uniqueness argument.

## 5.1 Domain-Level Content

Two features of the domain $\widetilde{\mathcal{W}}_m^0$ carry accounting content.

*Chart-of-accounts arbitrariness,* encoded as node-reordering invariance via the quotient under simultaneous row-column permutation. The operational motivation is the one given in Section 2: the chart-of-accounts ordering is a recordkeeping convenience, not informational content. Mathematically, node-reordering invariance forces $\phi(u) = \phi(1-u)$ for every binary information function on a 2-node graph. This symmetry is what Proposition 9's proof uses at four load-bearing steps to force the four binary functions governing the four roles of the grouping identity to coincide. The domain-level accounting content therefore does the work that one might naively attribute to an axiomatic duality requirement.

*Atomic-level conservation,* encoded as the zero-diagonal structure ($w_{ii} = 0$ for all $i$). This captures the accounting fact that atomic journal entries cannot debit and credit the same account—the balance constraint Dr = Cr would force an atomic self-entry to have zero weight. Restricting attention to $\mathcal{W}_m^0$ is the mathematical encoding of this constraint. It restricts which partitions the grouping rule may address (only those induced by node-merges, not arbitrary atom-mergers), giving the grouping identity its node-merge rather than atom-merge form.

### 5.2 Operation-Level Content

*The aggregation-netting convention* (intercompany elimination being its institutionally named instance), encoded in the zero-diagonal merge definition. When two nodes merge, the bilateral mass $w_{ab} + w_{ba}$ is destroyed, and the remaining distribution is renormalized. This operational choice generates two features of the grouping identity: the $(1-\tau)$ factor multiplying the recursive entropy $H_{m-1}^{\text{edge}}(W^{(a,b)})$, and the $H_2(\tau, 1-\tau)$ bilateral-elimination term.

These features are operational consequences, not axiomatic commitments. The self-loop-permissive variant developed in Section 6 demonstrates this directly: under a mass-preserving merge convention, the $(1-\tau)$ factor disappears and the $H_2(\tau, 1-\tau)$ term vanishes. The choice of convention reflects reporting practice (intercompany elimination, and aggregation netting more generally), and changing the practice changes the decomposition. The underlying entropy measure $H^{\text{edge}}$ is robust to this choice; the convention selects how the measure decomposes under aggregation.

### 5.3 Summary

| Accounting content | Enters at | Mechanism | Mathematical role |
|---|---|---|---|
| Chart-of-accounts arbitrariness | Domain | Quotient under row-col permutation: $\phi(u) = \phi(1-u)$ | Forces coincidence of four binary functions |
| Atomic-level conservation | Domain | Inclusion in $\mathcal{W}_m^0$: zero diagonal | Restricts grouping to node-merge partitions |
| Aggregation netting (incl. intercompany elimination) | Operation (merge def.) | Mass-destroying merge convention | Generates $(1-\tau)$ factor and $H_2(\tau, 1-\tau)$ term |

No accounting-specific content enters at the axiom level. Axioms 1 and 2 are generic Faddeev-style requirements (normalization and continuity), and Axiom 3's substantive content—the grouping identity's form—is the mathematical articulation of how the zero-diagonal merge operation interacts with the edge distribution. The single-function clause of Axiom 3 is redundant, as Proposition 9 establishes: it follows from the domain symmetry rather than imposing independent content.

The consequence for interpretation is that the three laws of bookkeeping are correctly read as operational and structural facts about the accounting signal, not as independent constraints on the

information measure. Duality manifests mathematically through chart-of-accounts arbitrariness at the domain level; conservation manifests through the zero-diagonal structure at the domain level; and the intercompany-elimination convention manifests through the merge operation's mass-destroying form. The uniqueness result tells us that Shannon edge entropy is the unique continuous normalized function on $\widetilde{\mathcal{W}}_m^0$ compatible with node-merge grouping under the zero-diagonal merge convention. It does not require—and the proof shows would be wrong to suggest—any additional axiomatic commitment to the three laws.

# 6 The Self-Loop Variant and the Consolidation Bridge

## 6.1 Motivation

Real accounting nodes are aggregations of finer sub-ledger entries. The "Inventory" account on a published balance sheet, for example, aggregates raw materials, work-in-progress, and finished goods; the internal transfer from WIP to finished goods is a genuine journal entry whose credit and debit both touch what appears on the consolidated statement as a single "Inventory" node. Enforcing zero diagonal on all aggregated nodes loses this information.

The self-loop-permissive variant preserves such flows by allowing the merged node to carry a non-zero self-edge equal to the bilateral mass that the zero-diagonal convention would otherwise destroy.

## 6.2 Definition and Grouping Identity

Let $\mathcal{W}_m^{\circlearrowleft}$ denote the set of $m \times m$ non-negative matrices (without the zero-diagonal constraint), with total mass $S^{\circlearrowleft}(W) = \sum_{i,j} w_{ij}$ and edge distribution $\pi_{ij}^{\circlearrowleft} = w_{ij}/S^{\circlearrowleft}(W)$. Define the self-loop-permissive edge entropy as

$$H_m^{\circlearrowleft}(W) \;=\; -\sum_{i,j} \pi_{ij}^{\circlearrowleft} \log \pi_{ij}^{\circlearrowleft}.$$

The self-loop-permissive node-merge operation $W \mapsto W_{\circlearrowleft}^{(a,b)}$ differs from Definition 2 only in its treatment of the diagonal:

$$w_{\circlearrowleft,cc}^{(a,b)} \;=\; w_{aa} + w_{bb} + w_{ab} + w_{ba}.$$

Total mass is preserved: $S^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) = S^{\circlearrowleft}(W)$.

**Proposition 11** (Self-loop-permissive grouping identity)**.** *For $W \in \mathcal{W}_m^{\circlearrowleft}$ with $S^{\circlearrowleft}(W) > 0$ and $w_{aa} = w_{bb} = 0$ (other diagonal entries may be nonzero; they pass through the merge as untouched cells), with $\tau$, $p_j^{out}$, and $p_i^{in}$ defined as in Section 3 but with respect to $\pi^{\circlearrowleft}$:*

$$\begin{aligned} H_m^{\circlearrowleft}(W) = H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) &+ \sum_{j \notin \{a,b\}} p_j^{out}\, H_2\Big(\tfrac{\pi_{aj}}{p_j^{out}}, \tfrac{\pi_{bj}}{p_j^{out}}\Big) \\ &+ \sum_{i \notin \{a,b\}} p_i^{in}\, H_2\Big(\tfrac{\pi_{ia}}{p_i^{in}}, \tfrac{\pi_{ib}}{p_i^{in}}\Big) \\ &+ \tau\, H_2\big(\tfrac{\pi_{ab}}{\tau}, \tfrac{\pi_{ba}}{\tau}\big)\,. \end{aligned} \tag{13}$$

The proof is an easier version of Theorem 3: because total mass is preserved, no renormalization occurs, the $(1-\tau)$ factor disappears, and the $H_2(\tau, 1-\tau)$ bilateral-elimination term is absent. The grouping identity in the self-loop setting is structurally closer to the standard probability-vector Cover–Thomas grouping.

**Remark 12** (Axiomatization in the self-loop-permissive setting)**.** We expect a uniqueness theorem analogous to Theorem 7, and a minimality result analogous to Proposition 9, to hold for $H^{\circlearrowleft}$ under correspondingly modified axioms, by the same functional-equation strategy. We have not carried out the details; the statement should be read as a conjecture.

### 6.3 The Consolidation Bridge

The zero-diagonal and self-loop-permissive edge entropies are linked by an identity with a clean accounting interpretation.

**Corollary 13** (Consolidation bridge)**.** *For $W \in \mathcal{W}_m^0$ with $S(W) > 0$, let $W^{(a,b)}$ and $W_{\circlearrowleft}^{(a,b)}$ denote the two post-merge matrices. Then*

$$H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) \;=\; (1-\tau)\, H_{m-1}^{edge}(W^{(a,b)}) \;+\; H_2(\tau, 1-\tau). \tag{14}$$

*Proof.* Direct computation: in the self-loop-permissive post-merge matrix, the diagonal cell $(c, c)$ has mass $\tau$ and the off-diagonal cells have total mass $1-\tau$, distributed as the zero-diagonal post-merge distribution scaled by $(1-\tau)$. Applying (3) with the binary partition {self-loop, external} gives the result. □

Equation (14) decomposes the self-loop-permissive entropy into two interpretable pieces: a binary question ("is this dollar intercompany?") and the conditional entropy of the external structure given the dollar is external. The *consolidation entropy gap* is defined as

$$\Delta_{\text{cons}} \;:=\; H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) - H_{m-1}^{\text{edge}}(W^{(a,b)}) \;=\; H_2(\tau, 1-\tau) - \tau\, H_{m-1}^{\text{edge}}(W^{(a,b)}).$$

This quantity is *sign-indefinite*: elimination can lower or raise measured entropy. From the right-hand expression, $\Delta_{\text{cons}} > 0$ when the intercompany fraction $\tau$ is moderate and the post-merge external graph is low-entropy, while $\Delta_{\text{cons}} < 0$ whenever $H_{m-1}^{\text{edge}}(W^{(a,b)}) > H_2(\tau, 1-\tau)/\tau$—in particular, whenever the external graph is sufficiently entropy-rich. The bridge identity (14) exhibits the two competing forces: eliminating the self-loop discards the binary internal-versus-external question, losing $H_2(\tau, 1-\tau)$, but promotes the conditional external entropy from weight $1-\tau$ to weight one, gaining $\tau\, H_{m-1}^{\text{edge}}(W^{(a,b)})$. The gap $\Delta_{\text{cons}}$ is exactly the first effect net of the second.

**Remark 14** (Elimination is not aggregation)**.** A tempting intuition runs: zeroing out the diagonal removes a state, so the remaining distribution is "more concentrated" and entropy must fall. The intuition is wrong, and the reason is renormalization. Elimination does not redistribute the bilateral mass to the surviving cells—it *destroys* that mass, and the surviving distribution must be renormalized to sum to one. Renormalization is a conditioning operation, and conditioning can push entropy in either direction. An extreme case makes the point: let the merged node's self-loop carry mass fraction $\tau \to 1$ while the external cells are uniform. Pre-elimination, $H_{m-1}^{\circlearrowleft} = (1-\tau)\, H_{m-1}^{\text{edge}} + H_2(\tau, 1-\tau) \to 0$: nearly all mass sits in a single cell, and the graph is nearly deterministic. Post-elimination, the renormalized external graph is exactly uniform, and $H_{m-1}^{\text{edge}}$ is maximal. Elimination has taken a nearly-zero-entropy object to a maximum-entropy one.

The sign-indefiniteness marks a genuine conceptual distinction: information as measured by entropy is not the same as information as measured by aggregation. Aggregation proper—the node-merge in the mass-preserving, self-loop-permissive setting—destroys information monotonically: rearranging Proposition 11,

$$H_m^{\circlearrowleft}(W) - H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) \;=\; \sum_{j \notin \{a,b\}} p_j^{\text{out}}\, H_2\Big(\tfrac{\pi_{aj}}{p_j^{\text{out}}}, \tfrac{\pi_{bj}}{p_j^{\text{out}}}\Big) + \sum_{i \notin \{a,b\}} p_i^{\text{in}}\, H_2\Big(\tfrac{\pi_{ia}}{p_i^{\text{in}}}, \tfrac{\pi_{ib}}{p_i^{\text{in}}}\Big) + \tau\, H_2\big(\tfrac{\pi_{ab}}{\tau}, \tfrac{\pi_{ba}}{\tau}\big) \;\geq\; 0,$$

a sum of mass-weighted binary entropies, each term the information needed to recover the pre-merge split from the post-merge aggregate. This nonnegative quantity is the formalization of "information destroyed by node-merge," and it requires no renormalization because the mass-preserving merge compares distributions on a common mass base. The elimination step, by contrast, is a change of measurement convention—a reconditioning, not an aggregation—and its entropy effect $\Delta_{\text{cons}}$ carries no definite sign.

**Example 15** (Numerical illustration)**.** Continuing Example 4 with $\tau = 0.4$ and post-merge $H_2^{\text{edge}}(W^{(2,3)}) = 1$ bit:

$$H_2^{\circlearrowleft}(W_{\circlearrowleft}^{(2,3)}) = 0.6 \cdot 1 + H_2(0.4, 0.6) \approx 0.6 + 0.971 = 1.571 \text{ bits.}$$

Consolidation gap: $\Delta_{\text{cons}} \approx 0.971 - 0.4 \cdot 1 = 0.571$ bits. With a richer external graph the gap turns negative: the same $\tau = 0.4$ against a post-merge external graph with $H^{\text{edge}} = 3$ bits gives $\Delta_{\text{cons}} \approx 0.971 - 1.2 = -0.229$ bits. □

### 6.4 Three Entropies for Aggregation Analysis

The framework supports a three-entropy analysis of any node-merge:

$$\begin{aligned} H_m^{\text{edge}}(W) \quad & \text{(pre-merge, full disaggregated information)} \\ H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) \quad & \text{(post-merge, intercompany preserved as self-loop)} \\ H_{m-1}^{\text{edge}}(W^{(a,b)}) \quad & \text{(post-merge, intercompany eliminated)} \end{aligned}$$

The first gap, $H_m^{\text{edge}}(W) - H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)})$, measures pure aggregation loss (the classical Cover–Thomas information destruction); since the pre-merge $W$ has zero diagonal, $H_m^{\text{edge}}(W) = H_m^{\circlearrowleft}(W)$, and Proposition 11 identifies this gap as a sum of mass-weighted binary split entropies—it is always nonnegative. The second gap, $H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) - H_{m-1}^{\text{edge}}(W^{(a,b)}) = \Delta_{\text{cons}}$, measures the elimination *effect*; by Remark 14 it carries no definite sign. The two are separable channels—the first an aggregation, the second a change of measurement convention—and can in principle be measured separately in empirical data if one has both the consolidated and the pre-consolidation financials.

## 7 Random versus Meaningful Merges as a Classification Quality Test

The self-loop-permissive grouping identity (Proposition 11) suggests an empirical application: using aggregation loss under node-merge as a test statistic for how well an accounting classification fits the transaction-flow structure it organizes. Throughout this section the mass-preserving merge of Section 6 is used, so that every merge loss is nonnegative and the sign-indefinite elimination effect of Remark 14 is excluded by construction.

### 7.1 Aggregation Loss and When It Vanishes

For $W \in \mathcal{W}_m^0$ with $S(W) > 0$ and a pair $a \neq b$, define the aggregation loss

$$\begin{aligned} \Delta H(a,b) \;:=\; H_m^{\text{edge}}(W) - H_{m-1}^{\circlearrowleft}(W_{\circlearrowleft}^{(a,b)}) \;=\; & \sum_{j \notin \{a,b\}} p_j^{\text{out}}\, H_2\Big(\tfrac{\pi_{aj}}{p_j^{\text{out}}}, \tfrac{\pi_{bj}}{p_j^{\text{out}}}\Big) \\ & + \sum_{i \notin \{a,b\}} p_i^{\text{in}}\, H_2\Big(\tfrac{\pi_{ia}}{p_i^{\text{in}}}, \tfrac{\pi_{ib}}{p_i^{\text{in}}}\Big) \;+\; \tau\, H_2\big(\tfrac{\pi_{ab}}{\tau}, \tfrac{\pi_{ba}}{\tau}\big), \end{aligned} \tag{15}$$

where the second equality is Proposition 11 (using $H_m^{\text{edge}}(W) = H_m^{\circlearrowleft}(W)$ on the zero-diagonal domain) and terms with zero group mass are read as zero. Each summand is the information required to recover the pre-merge split from the post-merge aggregate, so $\Delta H(a, b) \geq 0$ always.

The structure of (15) determines exactly when aggregation is lossless: $\Delta H(a, b) = 0$ if and only if every split is degenerate—no external account receives flow from both $a$ and $b$, no external account sends flow to both, and the bilateral flow between them runs in one direction only. Lossless is a statement about *non-overlapping flow supports*, and the loss is maximal, relationship by relationship, when $a$ and $b$ split each common relationship evenly.

This inverts a tempting intuition. One might expect that merging two "similar" accounts—same counterparties, comparable magnitudes—destroys little information. Equation (15) says the opposite: identical flow patterns produce 50:50 splits, which maximize every binary entropy in the sum. What makes a merge cheap is *complementarity*, not similarity: the pattern of a sequential process chain, in which each stage has its own external counterparties and internal transfers run one way. A classification whose classes are aggregable in this sense loses little by reporting aggregates—the class total is nearly sufficient for the disaggregated flows. We accordingly read within-class aggregation loss as an (inverse) measure of classification quality *as aggregation*: the linearity law's question of what survives additive aggregation, posed informationally.

### 7.2 Test Statistic

For a bookkeeping graph $W$ on $m$ nodes and a classification $\mathcal{C}$ partitioning the nodes into classes (at least one of size two or more), define

$$\begin{aligned} \overline{\Delta H}_{\text{random}} &:= \mathbb{E}\big[\Delta H(a, b)\big], && (a, b) \text{ uniform over all unordered pairs,} \\ \overline{\Delta H}_{\text{within}} &:= \mathbb{E}\big[\Delta H(a, b)\big], && (a, b) \text{ uniform over unordered pairs sharing a class,} \end{aligned}$$

and, provided $\overline{\Delta H}_{\text{random}} > 0$, the classification quality ratio

$$Q(\mathcal{C}; W) := \frac{\overline{\Delta H}_{\text{random}} - \overline{\Delta H}_{\text{within}}}{\overline{\Delta H}_{\text{random}}}.$$

Since $\Delta H \geq 0$, the statistic satisfies $Q \leq 1$, with $Q = 1$ attained when every within-class merge is exactly lossless. The lower end is not bounded: $Q \geq 0$ holds if and only if within-class merges destroy no more information than random merges, and this is precisely the empirical hypothesis under test, not a property of the construction. A negative value is itself informative: it says the classification groups accounts whose aggregation is *more* destructive than random grouping—parallel near-duplicates rather than complementary process stages.

### 7.3 A Two-Firm Contrast

The conditional nature of $Q$ is visible in a minimal example. Both firms have five accounts—two payables ($\text{AP}_1, \text{AP}_2$), two inventory accounts ($X_1, X_2$), and COGS—and the same classification $\mathcal{C}$: the inventory class $\{X_1, X_2\}$, all other accounts singletons. They differ only in flow topology.

*Chain firm* (sequential production: $X_1$ feeds $X_2$):

$$\pi_{\text{AP}_1, X_1} = \pi_{\text{AP}_2, X_1} = 0.2, \qquad \pi_{X_1, X_2} = 0.3, \qquad \pi_{X_2, \text{COGS}} = 0.3.$$

The within-class merge $\{X_1, X_2\}$ is exactly lossless: the internal transfer is one-directional, $X_1$ alone receives from the payables, and $X_2$ alone feeds COGS, so every split in (15) is degenerate and $\Delta H(X_1, X_2) = 0$. Of the ten unordered pairs, the only lossy merge is $\{\text{AP}_1, \text{AP}_2\}$—the two

payables split the inflow to $X_1$ evenly, costing $0.4 \cdot H_2(1/2, 1/2) = 0.4$ bits—so $\overline{\Delta H}_{\text{random}} = 0.04$, $\overline{\Delta H}_{\text{within}} = 0$, and

$$Q_{\text{chain}} \;=\; 1.$$

*Pool firm* (parallel storage: $X_1, X_2$ are interchangeable pools):

$$\pi_{\text{AP}_1,X_1} = \pi_{\text{AP}_1,X_2} = \pi_{\text{AP}_2,X_1} = \pi_{\text{AP}_2,X_2} = 0.125, \qquad \pi_{X_1,\text{COGS}} = \pi_{X_2,\text{COGS}} = 0.25.$$

Now the within-class merge is the most destructive merge available: $X_1$ and $X_2$ split every relationship evenly, so $\Delta H(X_1, X_2) = 0.5 \cdot 1 + 2 \cdot (0.25 \cdot 1) = 1$ bit. The only other lossy pair is $\{\text{AP}_1, \text{AP}_2\}$ at 0.5 bits, so $\overline{\Delta H}_{\text{random}} = 0.15$, $\overline{\Delta H}_{\text{within}} = 1$, and

$$Q_{\text{pool}} \;=\; \frac{0.15 - 1}{0.15} \;\approx\; -5.67.$$

The same nominal classification—"inventory"—is maximally informative as aggregation for the chain firm and worse than random for the pool firm. $Q$ is not a property of the classification labels; it is a property of the fit between the classification and the firm's flow topology.

### 7.4 Conditions and Caveats for Applied Use

Two considerations govern interpretation in applied settings.

*The hypothesis is conditional on process structure.* Within-class merges are cheap when classes track process chains (one-directional internal transfers, complementary external counterparties) and expensive when classes collect parallel accounts facing common counterparties at comparable scale. Both patterns arise under semantically reasonable classifications, as the two-firm contrast shows: sub-ledgers organized by production stage favor $Q > 0$, while sub-ledgers organized by location or duplicated facility need not. $Q$ should therefore be read as measuring informativeness-as-aggregation relative to the firm's actual flows, not semantic coherence in the abstract.

*Mass confounding.* Each term of (15) is bounded by its group mass, so $\Delta H(a, b)$ is bounded by the total flow incident to the pair; high-flow pairs can dominate both averages. The comparison of within-class to random pairs thus mixes structural fit with the location of mass. A refinement is a mass-matched baseline—random pairs drawn to match the incident-mass profile of the within-class pairs—with the unmatched $Q$ reported alongside.

### 7.5 Applications

With these qualifications, the statistic $Q$ admits several direct applications:

1. *Comparing classification schemes.* Given multiple candidate classifications—e.g., the GAAP balance sheet decomposition versus the Penman-Nissim operating/financing reformulation—compute $Q$ for each on the same underlying graph. The scheme with higher $Q$ aggregates with less information loss on that firm's flows.

2. *Hierarchical quality assessment.* Applied at each level of a nested classification, $Q$ provides a node-by-node quality map of the chart-of-accounts hierarchy.

3. *Industry comparison.* Compute $Q$ across industries for the same classification. Industries with low or negative $Q$ are those whose flow topology the standard classification fits poorly—the pool-firm pattern of the two-firm contrast—potentially identifying where accounting standards need industry-specific modification.

4. *Temporal change detection.* Monitor $Q$ over time for a given firm; sharp drops signal that the firm's underlying transaction structure has evolved away from the classification scheme's assumptions (e.g., a manufacturing firm transitioning to a platform-based business model).

This statistic complements text-similarity-based taxonomy approaches [e.g., Hoberg and Phillips, 2016] by providing an orthogonal, transaction-flow-based assessment of classification fit.

# 8 Positioning: ACE Proposition 1 and Cover–Thomas

## 8.1 Comparison with ACE Proposition 1

ACE Proposition 1 [Li et al., 2025] reproduces the Cover and Thomas [2006] uniqueness theorem on the probability simplex, with accounting fractions $x_i = X_i/K$ substituted for generic probabilities. It continues the tradition of applying information-theoretic measures to financial statement data that runs from Theil [1969] to Fellingham et al. [2019]. The theorem's mathematical content is borrowed; the paper's contribution lies in the interpretive layer—arguing that the Grouping axiom corresponds to the hierarchical classification structure of a financial statement.

Theorem 7 and Proposition 9 together differ in three respects. First, the underlying object is the directed adjacency matrix, not the probability vector; the uniqueness result is therefore not available from Cover–Thomas directly. Second, the axioms and the proof go through the Faddeev functional equation derived on a graph-topology-constrained grouping rule (node-merge rather than atom-merge), which is strictly stronger than the unrestricted atom-merge grouping of Cover–Thomas. Third, Proposition 9 proves the axiom system is minimal: the natural multi-function relaxation of the grouping axiom collapses to the single-function form under the domain symmetry. The ACE setting does not admit an analogous minimality question because its grouping axiom involves only a single binary function by standard convention on the probability simplex.

In the ACE setting, the accounting content enters only in interpretation. In the edge-entropy setting, it enters at two structurally distinct levels: the domain (via node-reordering invariance, encoding chart-of-accounts arbitrariness, and zero-diagonal structure, encoding atomic-level conservation) and the operation (via the zero-diagonal merge convention, encoding intercompany elimination). The axiom level carries no accounting-specific content beyond the generic Faddeev-style requirements, as Proposition 9 makes explicit.

## 8.2 Comparison with Cover–Thomas

Cover and Thomas [2006, Theorem 2.6.4] characterize Shannon entropy on the probability simplex as the unique function satisfying continuity, monotonicity in $m$ for the uniform distribution, and a grouping axiom on binary partitions. Faddeev [1956] gives a tighter version using normalization, continuity, and grouping—with symmetry built into the domain (functions on probability *distributions* rather than tuples) rather than imposed as an axiom. Theorem 7 adopts the Faddeev style: normalization, continuity, and a single grouping axiom, with simultaneous row-column permutation invariance built into the domain $\widetilde{\mathcal{W}}_m^0$.

The substantive differences from the classical case are twofold:

1. *Domain.* The underlying object is $\widetilde{\mathcal{W}}_m^0$, the space of zero-diagonal non-negative matrices modulo simultaneous row-column permutation. Node-reordering invariance is the natural and only invariance built into the domain—weaker than the full $S_{m(m-1)}$ symmetry that would result from treating the off-diagonal cells as an unstructured probability vector. The weaker domain symmetry is essential: it preserves the directed structure of journal-entry flows

and reserves the question of debit-credit equivalence for the domain-level accounting content rather than absorbing it into an over-strong symmetry assumption.

2. *Grouping rule.* Node-merge grouping respects the graph topology and is strictly stronger than arbitrary binary grouping on the edge simplex: it restricts which pairs of cells may be grouped at each step (only pairs corresponding to a node-merge), and the resulting identity carries the $(1-\tau)$ rescaling and bilateral-elimination terms that have no counterpart in the vector case.

Despite the more restrictive grouping rule, the uniqueness argument goes through via Faddeev's functional equation. The key observation is that node-merge at $m = 3$ can be applied in two distinct ways (merging different pairs of nodes), and equating the two expressions yields a Faddeev-type equation on the binary information function. Continuity and normalization then close the argument exactly as in the classical case, with the Sufficiency Lemma (Lemma 5) propagating the determination of $J_2^{\text{bin}}$ to $J_m$ on the full domain.

Proposition 9 establishes an architectural feature absent from the classical case: the grouping axiom's natural multi-function relaxation is redundant under the domain symmetry. This is a statement about the minimality of the axiom system on a quotient domain and has no counterpart in the Cover–Thomas or Faddeev arguments, where the domain symmetry and the axiomatic structure are arranged so that the question does not arise.

# 9 Discussion and Future Directions

## 9.1 What the Theorem Does and Does Not Do

Theorem 7 characterizes Shannon edge entropy as the unique measure on $\widetilde{\mathcal{W}}_m^0$ satisfying Axioms 1–3, and Proposition 9 shows the axiom system is minimal in the precise sense that the natural multi-function relaxation of Axiom 3 collapses under the domain symmetry. It does *not* characterize Shannon edge entropy among all functions of the adjacency matrix—only among functions that depend on the edge distribution alone (a consequence of the domain choice). In particular, edge entropy is order-invariant on the off-diagonal cells: the multiset of edge probabilities is sufficient. For measures that respect the matrix structure beyond the edge distribution—which cell connects to which, the directed cycle structure, the spectral properties of $W$ or its Laplacian—one must turn to spectral functionals on the Laplacian [see Liang, 2026, for the development of bookkeeping graph-entropy on Laplacian eigenvalues]. Uniqueness in the spectral setting is a separate and, we conjecture, considerably harder question; the axioms from Theorem 7 do not lift to the spectral setting in any immediate way.

## 9.2 Extensions

Several extensions suggest themselves:

1. *Relative edge entropy.* A K-L divergence version of Theorem 7, characterizing $H^{\text{edge}}(W\|W') = \sum \pi_{ij} \log(\pi_{ij}/\pi'_{ij})$ axiomatically, would parallel the ACER extension of ACE and would support temporal analysis of bookkeeping graph change.

2. *Conditional edge entropy.* Given a partition of nodes into classes $\mathcal{C}$, the conditional entropy $H^{\text{edge}}(W \mid \mathcal{C})$ would formalize the intuition that node-class labels carry information about edge distributions. The classification quality statistic $Q$ of Section 7 is closely related but not identical.

3. *Axiomatization for joint node-edge entropy.* The Three Laws paper [Liang, 2026] proposes three graph invariants (node-weights, edge-weights, Laplacian eigenvalues) and corresponding entropy measures. An axiomatization that jointly characterizes node entropy and edge entropy via a combined grouping operation is a natural next step.

4. *Asymmetric directed case.* The simultaneous row-column permutation built into the domain $\widetilde{\mathcal{W}}_m^0$ is restrictive; weakening this—for instance, by allowing direction-specific transformations or distinguishing inflow-cells from outflow-cells—would enlarge the family of admissible entropies and might capture flow-direction asymmetries (e.g., inflows versus outflows on a cash account).

### 9.3 Closing Remark

Theorem 7 is Faddeev's uniqueness result localized to a node-merge grouping rule on $\widetilde{\mathcal{W}}_m^0$. Its accounting content is not in additional axioms but in the domain and operation definitions: chart-of-accounts arbitrariness (via node-reordering invariance), atomic-level conservation (via the zero-diagonal structure), and the intercompany-elimination convention (via the merge operation's mass-destroying form). Proposition 9 establishes that these domain and operation choices are sufficient—no separate axiom imposing row-column equality is required, because node-reordering invariance already collapses the natural four-function grouping identity into its single-function form. The theorem should therefore be read as: *Shannon edge entropy is the unique continuous normalized function on $\widetilde{\mathcal{W}}_m^0$ compatible with the node-merge grouping rule under the zero-diagonal merge convention, and this characterization is minimal in the sense that no additional axiomatic content concerning debit-credit duality need be imposed.*

**Acknowledgments.** The debit-credit duality interpretation owes an intellectual debt to Yuji Ijiri and John Fellingham. The critical reading that produced the two-level locus reframing of Section 5 and the Minimality Proposition (Proposition 9) came from a critical reader whose persistence on the question of whether the single-function clause of Axiom 3 carries independent content led directly to the present version of the result.

## A Faddeev's Functional Equation

The proofs of Theorem 7 (Uniqueness of edge entropy) and Proposition 9 (Axiomatic Minimality) invoke the classical fact that Faddeev's functional equation, under continuity, symmetry, and normalization, has the binary Shannon entropy as its unique solution. This appendix states the result in the form used in the main text, records the classical references, and develops its reformulation as a grouping identity. The result originates with Faddeev [1956]; canonical textbook treatments are Aczél and Daróczy [1975, Ch. 3] and Kannappan [2009, Ch. 3].

### Statement

Let $T := \{(x, y) \in \mathbb{R}^2 : x > 0,\ y > 0,\ x + y < 1\}$. For $(x, y) \in T$, both $x/(1-y)$ and $y/(1-x)$ lie in $(0, 1)$.

**Theorem 16.** *Let $\phi : (0, 1) \to \mathbb{R}$ be continuous, satisfy $\phi(x) = \phi(1-x)$ for all $x \in (0, 1)$ and $\phi(1/2) = 1$, and satisfy Faddeev's functional equation*

$$\phi(y) + (1-y)\,\phi\Big(\tfrac{x}{1-y}\Big) \;=\; \phi(x) + (1-x)\,\phi\Big(\tfrac{y}{1-x}\Big) \qquad \text{for all } (x, y) \in T. \tag{16}$$

*Then*

$$\phi(x) \;=\; H_2(x) \;:=\; -x\log_2 x - (1-x)\log_2(1-x).$$

*Proof.* This is Faddeev's uniqueness theorem [Faddeev, 1956]; see Aczél and Daróczy [1975, Ch. 3] for the canonical treatment of equation (16) (there called the fundamental equation of information) and Kannappan [2009, Ch. 3] for a modern exposition. We do not reproduce the argument here. □

**Remark 17** (What the symmetry hypothesis does)**.** The symmetry hypothesis is not decorative. Direct substitution shows that the family $\phi(x) = c\,H_2(x) + b\,x$ satisfies (16) for all constants $b, c$: the equation is linear in $\phi$, the function $H_2$ solves it, and under $\phi(x) = x$ both sides evaluate to $x+y$. Symmetry $\phi(x) = \phi(1-x)$ is what eliminates $b \neq 0$, and normalization $\phi(1/2) = 1$ then fixes $c = 1$. In the main text this symmetry is not an assumption on the function but a consequence of the node-reordering quotient $\widetilde{\mathcal{W}}_2^0$—one more instance of the domain, rather than the axioms, carrying the load.

### Relation to the grouping identity

Faddeev's equation (16) is the binary-entropy form of the statement that grouping a three-element distribution in two different orders must agree. For $(p_1, p_2, p_3)$ with $p_i > 0$ and $p_1 + p_2 + p_3 = 1$, substitute $x = p_1$ and $y = p_3$ into (16) (so $1-y = p_1+p_2$, $x/(1-y) = p_1/(p_1+p_2)$, $1-x = p_2+p_3$, $y/(1-x) = p_3/(p_2+p_3)$):

$$\phi(p_3) + (p_1+p_2)\,\phi\!\left(\tfrac{p_1}{p_1+p_2}\right) \;=\; \phi(p_1) + (p_2+p_3)\,\phi\!\left(\tfrac{p_3}{p_2+p_3}\right).$$

Using $\phi(p_3) = \phi(1-p_3) = \phi(p_1+p_2)$ and $\phi(p_3/(p_2+p_3)) = \phi(p_2/(p_2+p_3))$ (both by symmetry), this rearranges to

$$\phi(p_1+p_2) + (p_1+p_2)\,\phi\!\left(\tfrac{p_1}{p_1+p_2}\right) \;=\; \phi(p_1) + (1-p_1)\,\phi\!\left(\tfrac{p_2}{p_2+p_3}\right). \tag{17}$$

Equation (17) is the classical *grouping identity*: the left-hand side is the entropy of the binary distribution $(p_1+p_2,\, p_3)$ plus the within-group entropy of the merged pair, and the right-hand side is the entropy of the binary distribution $(p_1,\, p_2+p_3)$ plus the within-group entropy of the merged pair on the other side. The functional equation (16) is precisely the statement that these two ways of partitioning the three-element distribution agree. The node-merge grouping identity of Theorem 3 is the adjacency-matrix generalization of this consistency requirement, and the proofs of Theorem 7 and Proposition 9 recover (16) from it on suitable 3-node families.

**Remark 18** (Where the theorem is invoked)**.** Theorem 16 enters the main text at two points.

*Step 2 of Theorem 7.* After Step 1 derives (16) on $\phi := J_2^{\mathrm{bin}}$ from a 3-node application of Axiom 3, Theorem 16 gives $\phi = H_2$. The required hypotheses are: continuity from Axiom 2; symmetry from the domain quotient $\widetilde{\mathcal{W}}_2^0$; normalization from Axiom 1; and equation (16) from Step 1.

*Step 2 of Proposition 9.* After the boundary-limit argument forces $\Phi = 2\phi_{\mathrm{bil}}$, substitution back into the two-function equation recovers (16) on $\phi_{\mathrm{bil}}$, at which point Theorem 16 applies to give $\phi_{\mathrm{bil}} = H_2$ and hence $\Phi = 2H_2$.

## References

Aczél, J., Daróczy, Z., 1975. On Measures of Information and Their Characterizations, vol. 115 of *Mathematics in Science and Engineering*. Academic Press, New York.

Arya, A., Fellingham, J. C., Glover, J. C., Schroeder, D. A., Strang, G., 2000. Inferring transactions from financial statements. Contemporary Accounting Research 17, 366–385.

Cover, T. M., Thomas, J. A., 2006. Elements of information theory. Wiley-Interscience.

Faddeev, D. K., 1956. On the concept of entropy of a finite probabilistic scheme. Uspekhi Matematicheskikh Nauk 11, 227–231, in Russian.

Fellingham, J. C., Lin, H., Schroeder, D., 2019. Entropy and accounting. Available at SSRN .

Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. Journal of Political Economy 124, 1423–1465.

Ijiri, Y., 1975. Theory of accounting measurement. Amer Accounting Assn.

Kannappan, P., 2009. Functional Equations and Inequalities with Applications. Springer Monographs in Mathematics, Springer.

Khinchin, A. I., 1957. Mathematical Foundations of Information Theory. Dover Publications.

Li, N., Liang, P. J., Pyo, J. J., Zhang, G., 2025. Accounting classification entropy, working paper. Available at SSRN: https://ssrn.com/abstract=5588192.

Liang, P. J., 2026. Three entropy measures of double-entry bookkeeping graph classification structure, working paper, CMU Accounting AI Research Lab Working Paper Series. Original draft: March 2025.

Liang, P. J., Pyo, J. J., Zhang, G., Zhang, X.-J., 2026. Accounting graph entropy (AGE): Measuring information of financial statement graph structure, working paper, CMU Accounting AI Research Lab Working Paper Series.

Shannon, C. E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27, 379–423.

Theil, H., 1969. On the use of information theory concepts in the analysis of financial statements. Management Science 15, 459–480.