big data and cognitive computing

MDPI

Article

An Information-Theoretic Framework for Characterizing Interaction-Order Diversity in Temporal Hypergraphs

Francesco Cauteruccio

Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, I84084 Fisciano, Italy; fcauteruccio@unisa.it

Abstract

The proliferation of large-scale interaction datasets, from scientific collaboration networks and legislative records to online communication platforms, has made the analysis of group-based, time-varying systems one of the central challenges of modern data analytics. Hypergraphs provide a natural formalism for such systems, where interactions involve arbitrary groups of agents rather than isolated pairs, and temporal hypergraphs extend this to sequential data by capturing how group interactions evolve over time. Yet quantifying how complex, predictable, or volatile this evolution is remains an open problem: existing entropy-based measures either operate on pairwise projections and thus discard multi-way dependencies or are not naturally defined for varying hyperedge sizes. In this paper, we propose an information–theoretic framework for characterizing how the diversity of interaction orders in a temporal hypergraph evolves over time. We introduce the hyperedge-size distribution entropy of a snapshot and, building on the theory of entropy rates for stochastic processes, we define the temporal hypergraph entropy rate as a principled, dataset-agnostic measure of the average diversity of interaction orders exhibited by the snapshot sequence over time. We further equip the framework with a bias-corrected sliding-window estimator and a lightweight change-point detector, assembling a complete pipeline that runs in time linear in the total number of hyperedges and requires no node alignment across datasets or snapshots. We prove that the measure collapses to zero under clique expansion, demonstrating that it captures interaction-order information that is discarded by the standard size-blind pairwise projection. Experiments on six small and large publicly available benchmark datasets show that the entropy rate spans 1.60 bits across domains, detects unsupervised structural change points, and discriminates between structurally distinct interaction cultures even within the same domain. Our framework is computationally lightweight and applicable to any dataset that can be represented as a temporal sequence of hypergraphs, paving the way for practical, scalable, interaction-order-aware analysis of large-scale higher-order temporal data.

Keywords: temporal network mining; entropy rate; higher-order networks; interaction-order diversity; network science; data analytics

Academic Editor: Giuseppe Ciaburro

Received: 7 May 2026
Revised: 29 June 2026
Accepted: 1 July 2026
Published: 3 July 2026

1. Introduction

The scale and variety of interaction data available for analysis have grown dramatically over the past decade. Co-authorship networks now span millions of publications; legislative cosponsorship records cover decades of congressional activity; online platforms generate hundreds of millions of user interactions per day. A shared structural property of all these systems is that their interactions are inherently group-based: papers are authored

by teams, bills are co-sponsored by coalitions, and forum threads attract sets of contributors. Hypergraphs provide a natural mathematical formalism for such systems, where a hyperedge can connect any number of nodes simultaneously, capturing joint dependencies that pairwise edges inevitably decompose and lose [1,2].

When group interactions are observed over time, the resulting structure is a temporal hypergraph: a sequence of hypergraph snapshots indexed by discrete or continuous time. Temporal hypergraphs arise naturally in co-authorship records, legislative data, and knowledge-tagging systems [3] and have been extensively studied in the past [1–5]. As datasets of this kind grow in size and temporal resolution, a fundamental descriptive question becomes increasingly pressing: *how does the diversity of interaction orders in a temporal hypergraph evolve, and how much does it vary over time?* Equivalently, how much does the higher-order interaction structure change from snapshot to snapshot, and how predictable is that change? Answering this question at scale requires a measure that is computationally efficient, requires no domain-specific tuning, and is comparable across heterogeneous datasets without any preprocessing or node alignment. In this sense, the problem belongs to the broader challenge of extracting meaningful, scalable descriptors from (possibly large-scale) temporal data, which is a challenge that grows more acute as the volume, variety, and temporal resolution of available interaction datasets continue to increase.

For pairwise temporal networks, entropy-based complexity measures have been studied extensively [6–8]. For temporal hypergraphs; however, this question remains largely unaddressed [9]. The most common analytical shortcut is to project the hypergraph onto a weighted pairwise graph and apply standard graph entropy measures. This approach is not only informationally lossy: a hyperedge of size three carries qualitatively different structural content from the three pairwise edges that approximate it, and the distinction matters precisely when higher-order effects are relevant [1,10]. It is also computationally wasteful: the clique expansion of a hyperedge of size s indeed produces $\binom{s}{2}$ pairwise edges [1]. A method that operates directly on the hyperedge-size distribution, without materializing the expanded graph, is therefore preferable both informationally and computationally.

To situate our contribution, we briefly survey the lines of work most directly related to ours. Entropy-based measures of structural complexity are well established for static graphs: degree distribution entropy [11] treats the normalized degree sequence as a probability distribution. The von Neumann entropy [12] is defined via the eigenvalue spectrum of the graph Laplacian, while the graph entropy in the sense of Körner [13] is rooted in combinatorial information theory. More recent work continues to develop entropy-aware graph analytics: structural-entropy methods and their applications have been surveyed comprehensively [14], and information–theoretic dissimilarity measures based on network hierarchy entropy have been shown to track evolving topologies in dynamic networks [15], though these operate on node- and path-level distributions of pairwise graphs rather than on the hyperedge-size distributions that our higher-order measure targets. Overall, we refer the interested reader to [13,16] for comprehensive surveys on graph entropy measures. For temporal pairwise networks, entropy rate has been applied to sequences of adjacency matrices and temporal networks, in general, [17,18], to edge appearance and disappearance processes [19], and to node-activity streams, capturing the long-run complexity of network evolution. Beyond the network-science literature, similar marginal entropy measures have been tracked over time in applied settings to characterize the structural diversification of evolving weighted networks, for instance, in the analysis of global value-chain networks [20]. The study of temporal higher-order networks, instead, has gained momentum more recently. In [3], the authors showed that group interactions exhibit non-trivial memory effects beyond pairwise contacts, while temporal ego-hypergraphs have been characterized in terms of local structural stability and evolution [4], providing

local complexity descriptors that motivate the need for a complementary global measure. On the information–theoretic side, the authors in [21] introduced partial information decomposition to characterize synergy and redundancy in higher-order systems, but this framework operates on joint distributions of node-state variables rather than on the structural properties of hyperedges. To the best of our knowledge, no existing work defines a complexity measure or entropy rate for the structural evolution of temporal hypergraphs based on the distribution of hyperedge sizes, which is the gap we strive to address with our contribution.

Particularly, in this paper, we propose an information–theoretic framework for measuring how the diversity of interaction orders in a temporal hypergraph evolves over time. The central idea is to characterize each snapshot of a temporal hypergraph by the probability distribution over its hyperedge sizes and to quantify the diversity of that distribution via Shannon entropy. Averaging the per-snapshot entropy over time gives the temporal hypergraph entropy rate, a scalar measure that captures how structurally complex and variable the higher-order interactions are across the full observation window. We prove that this measure collapses to zero when the hypergraph is projected onto its pairwise skeleton, establishing a precise, quantifiable gap between what pairwise and higher-order analyses can reveal. To make the framework applicable to finite real-world data, we design a sliding-window estimator and a lightweight change-point detector that translates the temporal complexity profile into a discrete set of structural events. Also, we design the framework to deliberately characterize a single, well-defined structural dimension, i.e., the distribution of interaction orders (in our case, hyperedge sizes) and its evolution over time, rather than the full node-level organization of the hypergraph. Indeed, it does not focus on node identities, hyperedge overlap, or membership patterns. Furthermore, we position our work as a practical, descriptive framework, while its building blocks are individually standard, the contribution lies in adapting and combining them into a dataset-agnostic, linear-time pipeline for a structural object that has not previously been analyzed this way and in characterizing what this descriptor reveals across a range of real datasets. Our aim is to provide a measure that is simple, interpretable, comparable across heterogeneous datasets, and scalable. More in detail, our contribution is as follows:

1. We introduce the hyperedge-size distribution entropy of a temporal hypergraph snapshot, representing a principled measure of the diversity of interaction orders present at a given time.
2. We define the temporal hypergraph entropy rate as the time-average entropy of the snapshot sequence, and we prove that it collapses to zero under clique expansion, demonstrating that it captures interaction-order information discarded by the standard projection.
3. We propose a practical sliding-window estimator that approximates the entropy rate on finite datasets and supports unsupervised change-point detection.
4. We validate the framework on six publicly available temporal hypergraph benchmark datasets spanning diverse domains, showing that the entropy rate discriminates across and within domains, detects unsupervised structural change points, and outperforms simpler size-based baselines. Also, we present a scalability analysis in order to empirically confirm the computational complexity of applying our framework.

The overall structure of the paper is organized as follows. In Section 2, we review background on temporal hypergraphs, Shannon entropy, and graph projections. Then, in Section 3, we introduce and formalize our theoretical framework, detailing all of its components. Afterward, in Section 4, we present our series of experiments, while in Section 5 we provide a thorough discussion of the contribution and indicate some limitations. Finally, in Section 6, we draw our conclusion and outline some future work.

2. Background

2.1. Hypergraphs and Their Temporal Extension

A *hypergraph* $\mathcal{H} = (V, \mathcal{E})$ consists of a finite node set $V = \{v_1, \ldots, v_n\}$ and a family of hyperedges $\mathcal{E} = \{e_1, \ldots, e_m\}$, where each hyperedge $e_i \subseteq V$ may contain any number of nodes [22]. When $|e| = 2$ for all $e \in \mathcal{E}$, the hypergraph reduces to an ordinary graph. We call $|e|$ the *order* (or *size*) of hyperedge e and denote by $s_{\min}$ and $s_{\max}$ the minimum and maximum orders present in $\mathcal{H}$. Note that $\mathcal{H}$ is static, i.e., its nodes and hyperedges do not change over time. The counterpart of static hypergraphs are temporal hypergraphs.

Definition 1. *A* temporal hypergraph *over a discrete time index* $\mathcal{T} = \{t_1, t_2, \ldots, t_T\}$ *is a sequence* $\boldsymbol{\mathcal{H}} = \left(\mathcal{H}^{(1)}, \mathcal{H}^{(2)}, \ldots, \mathcal{H}^{(T)}\right)$*, where each* snapshot $\mathcal{H}^{(t)} = (V^{(t)}, \mathcal{E}^{(t)})$ *is a hypergraph defined on a (possibly varying) node set* $V^{(t)} \subseteq V$ *and a set of hyperedges* $\mathcal{E}^{(t)}$*.*

We refer to T as the *length* of the temporal hypergraph. In this paper, without loss of generality, we assume $V^{(t)} \equiv V$, i.e., the node set is fixed across snapshots.

In order to analyze hypergraphs, both static and temporal ones, a common strategy is to first project a hypergraph onto a graph and then apply standard methods. Several projections have been proposed in the literature [1,23,24], and they differ in what structural information they preserve and what they discard.

We briefly recall the projection called clique expansion. It replaces each hyperedge $e \in \mathcal{E}$ with all $\binom{|e|}{2}$ pairwise edges among its members, assigning to each induced edge a weight equal to the number of hyperedges that contain both endpoints [24]. Among the projections listed above, the clique expansion is the only one that (i) preserves the original node set, and (ii) produces a standard weighted graph to which any existing graph-theoretic tool applies directly without modification. It is consequently the most widely used projection in practice and represents the default analytical choice for practitioners who do not have access to a dedicated hypergraph measure. Other projections are often used in the literature. The star expansion introduces one auxiliary node per hyperedge and connects it to all members of that hyperedge, yielding a bipartite graph between original nodes and hyperedge nodes [23]. The bipartite (incidence) representation is structurally equivalent: nodes on one side, hyperedges on the other, with edges recording membership. Both representations preserve hyperedge identity but introduce a heterogeneous node set, mixing original nodes with auxiliary hyperedge-nodes; entropy measures applied to the resulting bipartite graph conflate these two qualitatively different objects and are not directly comparable to measures on the original node set. Moreover, the line graph of a hypergraph places hyperedges as nodes and connects two hyperedge nodes by an edge whenever the corresponding hyperedges share at least one member [22]. The line graph captures co-membership patterns among hyperedges but operates on an entirely different node set, making comparisons with node-level entropy measures semantically incoherent.

In our paper, we adopt the clique expansion as our sole baseline because it represents the strongest possible competitor within the class of pairwise projections. As it will be clear in the following, if our method detects structural complexity that the clique expansion is not able to detect, then the gap is interpretable as the information exclusively carried by the higher-order structure of the data.

2.2. Shannon Entropy and Entropy Rate

For a discrete random variable X taking values in a finite alphabet $\mathcal{X}$ with probability mass function $p(x) = \Pr[X = x]$, the Shannon entropy is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x), \tag{1}$$

with the convention $0 \log_2 0 = 0$ [25]. Here, $H(X)$ quantifies the average uncertainty about the outcome of X, and is maximized when X is uniformly distributed over $\mathcal{X}$.

For a sequence of random variables $\{X_t\}_{t \geq 1}$ taking values in a finite alphabet, the entropy rate is

$$\mathfrak{h} = \lim_{T \to \infty} \frac{1}{T} H(X_1, X_2, \ldots, X_T), \tag{2}$$

provided the limit exists [26]. Intuitively, $\mathfrak{h}$ measures the average amount of new information contributed by each additional observation in the long run. When the sequence is stationary, i.e., when its statistical properties do not change over time, the entropy rate admits the equivalent characterization $\mathfrak{h} = \lim_{T \to \infty} H(X_T \mid X_1, \ldots, X_{T-1})$, i.e., the limiting conditional entropy of the next observation given the entire past [26]. A high entropy rate signals unpredictable, complex evolution; a low value indicates that the sequence is repetitive or highly structured.

3. The Framework

In this section, we develop the three conceptual layers our framework is based on. The first layer introduces the primitive quantity, i.e., the hyperedge-size distribution of a single snapshot, and defines how to measure its complexity via Shannon entropy. The second layer promotes snapshot entropy to a measure of the entire temporal sequence and establishes the key theoretical property, i.e., collapse under clique expansion. The third layer indicates how to practically apply the framework: we show how to correct for possible bias, how to track complexity over time via a sliding-window profile, and how to extract discrete structural change points from that profile. Together, the three layers form the pipeline that takes a temporal hypergraph as input and produces a complexity characterization as output.

3.1. Hyperedge-Size Distribution of a Snapshot

The core object of our framework is the size distribution of a hypergraph snapshot: the probability that a uniformly sampled hyperedge has a given size.

Definition 2 (Hyperedge-size distribution)**.** *Let $\mathcal{H}^{(t)} = (V, \mathcal{E}^{(t)})$ be a hypergraph snapshot with $m_t = |\mathcal{E}^{(t)}| > 0$ hyperedges. For each integer $s \geq 2$, let $m_t^{(s)} = |\{e \in \mathcal{E}^{(t)} : |e| = s\}|$ be the number of hyperedges of size s. The hyperedge-size distribution of $\mathcal{H}^{(t)}$ is the probability mass function*

$$p_s^{(t)}(s) = \frac{m_t^{(s)}}{m_t}, \qquad s \in \{2, 3, \ldots, n\}, \tag{3}$$

supported on the set $\mathcal{S}^{(t)} = \{s : m_t^{(s)} > 0\}$ of sizes actually present in $\mathcal{H}^{(t)}$.

Here, $p_s^{(t)}$ is the probability distribution we obtain by asking, for a randomly drawn hyperedge, how many nodes participate in it. Note that it deliberately ignores which nodes are involved, retaining only the group sizes. This design choice is what makes the measure dataset agnostic: indeed, it requires no node alignment across snapshots or datasets and is the reason the entropy rate can be compared meaningfully between different kinds of networks. We use $p_s^{(t)}$ as the input to every subsequent quantity in this section.

It is worth making explicit why the hyperedge-size distribution is a principled choice of descriptor for the question we pose. Our goal is not to characterize every aspect of temporal hypergraph structure, but specifically to measure how the diversity of interaction orders evolves over time, and for that question, the size distribution is a natural and minimal sufficient statistic. Three considerations motivate it. First, the order (size) of an interaction is the defining feature that separates higher-order structure from pairwise

structure: a hyperedge of order s is, by definition, an interaction among s agents that no collection of pairwise edges can represent without loss, so the distribution of orders is the most direct summary of *how higher-order* a system is at each instant. Second, it is precisely the dimension that pairwise projections destroy. As Proposition 1 shows, clique expansion maps every interaction to order two and collapses our measure to zero; the size distribution therefore isolates exactly the structural information that the most common pairwise analysis cannot recover. Third, the descriptor is invariant to node relabeling and to the specific identities of the participants, which is what makes it comparable across datasets of different sizes, domains, and vocabularies without any alignment and computable in time linear in the number of hyperedges. These properties are not incidental: they follow directly from restricting attention to interaction order, and they are what make the resulting measure deployable at scale and meaningful across heterogeneous data.

Also, we stress that "sufficient" here is meant relative to this well-defined objective, not as a claim that the size distribution captures all of temporal complexity. Two temporal hypergraphs with different node-level organization but identical size distributions receive identical values, a limitation we make explicit in Section 5. What the descriptor offers is a precise, interpretable, and computationally light characterization of one genuine facet of higher-order temporal structure, and the experiments of Section 4 provide empirical support that this facet is informative and not redundant with established alternatives. Indeed, the measure discriminates across and within domains (Sections 4.2.1 and 4.2.2) and behaves differently from a battery of pairwise and higher-order graph descriptors (Section 4.3.1), confirming that it carries structural information that other descriptors do not.

Definition 3 (Snapshot entropy)**.** *The snapshot entropy of $\mathcal{H}^{(t)}$ is the Shannon entropy of its hyperedge-size distribution:*

$$H(\mathcal{H}^{(t)}) = -\sum_{s\in\mathcal{S}^{(t)}} p_s^{(t)}(s)\,\log_2 p_s^{(t)}(s). \tag{4}$$

Here, $H(\mathcal{H}^{(t)}) = 0$ if all hyperedges in $\mathcal{H}^{(t)}$ have the same size, i.e., a $\mathcal{H}^{(t)}$ is a k-uniform hypergraph. The maximum value $\log_2|\mathcal{S}^{(t)}|$ is attained when all present sizes are equally represented. In particular, when $\mathcal{H}^{(t)}$ is a graph, i.e., all edges have size 2, then $\mathcal{S}^{(t)} = \{2\}$ and $H(\mathcal{H}^{(t)}) = 0$. Practically speaking, $H(\mathcal{H}^{(t)})$ is the building block of the entire framework: it is a single real number in $[0, \log_2|\mathcal{S}^{(t)}|]$ that summarises how diverse the interaction orders are at time t. We will use it as the per-snapshot input to the entropy rate (see Definition 4) and to the sliding-window estimator (see Definition 5). The fact that it equals zero for any pairwise graph is by choice, as we will show that projecting a hypergraph onto its pairwise skeleton always destroys this information entirely (see Proposition 1).

3.2. The Temporal Hypergraph Entropy Rate

We now promote the snapshot entropy to a measure of the temporal evolution of the entire sequence. The key idea is to treat the sequence of size distributions $\left(p_s^{(1)}, p_s^{(2)}, \ldots, p_s^{(T)}\right)$ as the output of a stochastic process and characterize its average per-step entropy.

Definition 4 (Temporal hypergraph entropy rate)**.** *Let $\boldsymbol{\mathcal{H}} = (\mathcal{H}^{(1)}, \ldots, \mathcal{H}^{(T)})$ be a temporal hypergraph. The* temporal hypergraph entropy rate *is*

$$\mathfrak{h}(\boldsymbol{\mathcal{H}}) = \frac{1}{T}\sum_{t=1}^{T} H(\mathcal{H}^{(t)}) = \frac{1}{T}\sum_{t=1}^{T}\left[-\sum_{s\in\mathcal{S}^{(t)}} p_s^{(t)}(s)\,\log_2 p_s^{(t)}(s)\right]. \tag{5}$$

Equation (5) defines $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ as the arithmetic mean of the T snapshot entropies. This is motivated by the classical notion of entropy rate for sequences of random variables (Equation (2)), in which the per-symbol entropy converges to a limit as the sequence grows. For a finite observed window of length T, that limit is not accessible, and we replace it with the sample mean over the available snapshots, which is a standard and well-understood approximation. Practically speaking, $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ is the quantity we use to characterize and compare temporal hypergraphs as a whole. A high value indicates that, on average across all snapshots, the interaction orders are broadly and evenly distributed, therefore, the hypergraph is structurally complex in the higher-order sense. A low value indicates that one or a few interaction sizes dominate consistently over time.

Furthermore, we believe it is worth clarifying in which sense Equation (5) relates to the entropy rate of a stochastic process defined in Equation (2). Let X_t denote the random variable describing the hyperedge-size distribution of snapshot $\mathcal{H}^{(t)}$. By the chain rule for entropy [26], the joint entropy of the snapshot sequence satisfies

$$H(X_1, \ldots, X_T) = \sum_{t=1}^{T} H\big(X_t \mid X_1, \ldots, X_{t-1}\big) \;\leq\; \sum_{t=1}^{T} H(X_t), \tag{6}$$

where the inequality follows because conditioning cannot increase entropy. Dividing by T, the true process entropy rate is therefore bounded above by our measure, i.e., $\frac{1}{T} H(X_1, \ldots, X_T) \leq \mathfrak{h}(\boldsymbol{\mathcal{H}})$. The two coincide exactly when the snapshots are mutually independent, in which case $H(X_t \mid X_1, \ldots, X_{t-1}) = H(X_t)$ for all t and Equation (5) is precisely the finite-sample estimator of the process entropy rate. In the general (dependent) case, $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ measures the average per-snapshot diversity of interaction orders rather than the conditional unpredictability of the next snapshot given the past; the difference between the two quantities equals the average mutual information that successive snapshots share, which our marginal measure deliberately does not capture. Note that by depending only on the marginal hyperedge-size distribution of each snapshot, $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ remains dataset-agnostic, requires no node alignment across time, and is computable in time linear in the number of hyperedges (Section 3.6), at the cost of not modeling inter-snapshot dependence. Analogous marginal, time-resolved entropy measures have been used effectively to characterize the structural evolution of temporal networks outside the hypergraph setting [20], where per-snapshot weighted degree entropies are tracked over time to quantify the diversification of global value-chain networks. Moreover, when the independence assumption is not warranted, we can conservatively describe $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ as the mean temporal snapshot entropy.

3.3. Relation to Pairwise Graph Entropy

Generally, a natural check for any hypergraph measure is to verify its behavior when the data are reduced to an ordinary graph. In our case, such a check also has direct experimental consequences: it tells us what we lose when we apply the most common analytical shortcut, i.e., projecting the hypergraph onto its pairwise skeleton. To formalize this aspect as precisely as possible, we introduce the following proposition.

Proposition 1 (Clique-expansion reduction). *Let $\boldsymbol{\mathcal{H}} = (\mathcal{H}^{(1)}, \ldots, \mathcal{H}^{(T)})$ be a temporal hypergraph in which every snapshot $\mathcal{H}^{(t)}$ is an ordinary graph, i.e., $|e| = 2$ for all $e \in \mathcal{E}^{(t)}$ and all t. Then $\mathfrak{h}(\boldsymbol{\mathcal{H}}) = 0$. Moreover, let $\boldsymbol{\mathcal{H}}'$ be the temporal hypergraph obtained by applying the clique expansion to every snapshot of $\boldsymbol{\mathcal{H}}$. Then $\mathfrak{h}(\boldsymbol{\mathcal{H}}') = 0$ regardless of the original hyperedge sizes.*

Proof. If every snapshot is a graph, then $\mathcal{S}^{(t)} = \{2\}$ for all t, so $p_s^{(t)}(2) = 1$ and $H(\mathcal{H}^{(t)}) = -1 \cdot \log_2 1 = 0$ for every t. The entropy rate is then $\mathfrak{h}(\boldsymbol{\mathcal{H}}) = \frac{1}{T}\sum_{t=1}^{T} 0 = 0$. The second claim follows immediately because the clique expansion replaces every hyperedge with edges of size 2, reducing each snapshot to a graph and applying the first claim. □

Proposition 1 shows that $\mathfrak{h}(\boldsymbol{\mathcal{H}}) = 0$ is a necessary consequence of a purely pairwise structure, but not a sufficient condition: a k-uniform hypergraph (with $k > 2$) also achieves $\mathfrak{h} = 0$ because all edges have the same size. A strictly positive entropy rate therefore certifies the presence of genuinely mixed-order interactions; a high value signals that the mixture itself varies across snapshots.

We consider the clique expansion because it is the most natural and most widely used pairwise baseline. In fact, it always produces $\mathfrak{h} = 0$; therefore, any positive value of $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ measures exactly the size-distribution information that is destroyed by the size-blind clique-expansion entropy defined here. It is worth pointing out that this holds for the specific projection we adopt as a baseline, not for pairwise analysis in general. Richer pairwise constructions, such as weighted projections, motif-based descriptors, or measures of temporal edge correlation, may still encode indirect traces of group-size heterogeneity. With Proposition 1, we show that hyperedge-size information is recovered directly and explicitly by our measure, whereas it is discarded entirely by the standard size-blind projection. Next, we formalize the information gap

$$\Delta(\boldsymbol{\mathcal{H}}) = \mathfrak{h}(\boldsymbol{\mathcal{H}}) - \mathfrak{h}(\boldsymbol{\mathcal{H}}_{\text{clique}}) = \mathfrak{h}(\boldsymbol{\mathcal{H}}) \tag{7}$$

and we use it in the experiments to quantify how much interaction-order diversity is lost, specifically under the size-blind clique-expansion baseline.

3.4. Interpreting the Snapshot Entropy

To give a more interpretable reading of what the snapshot entropy measures, we express it in terms of the Kullback–Leibler (KL) divergence from the uniform distribution over the observed sizes [26]. For a snapshot with $|\mathcal{S}^{(t)}| = k$ distinct sizes,

$$H\big(\mathcal{H}^{(t)}\big) = \log_2 k - D_{\mathrm{KL}}\big(p_s^{(t)} \,\|\, \mathcal{U}_k\big), \tag{8}$$

where $\mathcal{U}_k$ is the uniform distribution over k elements. A snapshot with high entropy is therefore one whose size distribution is close to uniform: no single interaction order dominates. A snapshot with low entropy (and more than one size present) is one concentrated around a few dominant orders, i.e., far from $\mathcal{U}_k$ in KL divergence. The entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ measures the average proximity to uniformity across the entire temporal sequence. Consequently, when applying our framework, in case of a temporal hypergraph having a high entropy rate, we can say its hyperedge-size distribution is consistently near-uniform across time.

3.5. Applying the Framework

Note that the definitions above are stated in terms of the true hyperedge-size distribution $p_s^{(t)}$. In practice, however, this distribution is never observed directly: we have access only to the m_t hyperedges present in snapshot $\mathcal{H}^{(t)}$, and m_t may be small for some snapshots. In what follows, we address how to estimate the framework quantities reliably from finite samples, how to track their evolution over time, and how to extract interpretable structural events from the resulting profile.

3.5.1. Finite-Sample Bias Correction

When m_t is small, the empirical size distribution $\hat{p_s}^{(t)}$ obtained directly from the observed hyperedges may be a poor estimate of the true distribution, and we know from the literature that the naive entropy estimator $\hat{H}(\mathcal{H}^{(t)}) = -\sum_s \hat{p}_s^{(t)} \log_2 \hat{p}_s^{(t)}$ is known to be biased downward, i.e., it systematically underestimates the entropy of the true distribution because rare size categories may be absent from the observed sample [27]. In our scenario, we adopt the Miller–Madow correction [28], defined as

$$\hat{H}_{\mathrm{MM}}(\mathcal{H}^{(t)}) = \hat{H}(\mathcal{H}^{(t)}) + \frac{|\mathcal{S}^{(t)}| - 1}{2m_t} \cdot \log_2 e, \tag{9}$$

which adds an upward correction proportional to the number of observed sizes and inversely proportional to the snapshot size. For large m_t, the correction term is negligible; it is most consequential for sparse snapshots where few hyperedges are present. In other words, this correction is what allows us to compare entropy values across snapshots of very different sizes without systematic bias. Without it, a snapshot with $m_t = 10$ hyperedges would appear structurally simpler than a snapshot with $m_t = 1000$ hyperedges even if both have the same true size distribution, simply because the smaller sample is less likely to have observed all present sizes. We apply $\hat{H}_{\mathrm{MM}}$ as the per-snapshot entropy estimate throughout all subsequent calculations.

3.5.2. Sliding-Window Entropy Rate

Rather than summarizing the entire temporal sequence as a single entropy rate, we propose a classical sliding-window estimator that reveals how structural complexity evolves over time. Let w be a window size and let τ range over the valid window centers $\lceil w/2 \rceil, \ldots, T - \lfloor w/2 \rfloor$.

Definition 5 (Sliding-window entropy rate)**.** *The sliding-window entropy rate at time τ with window size w is*

$$\mathfrak{h}_w(\tau) = \frac{1}{w} \sum_{t=\tau-\lfloor w/2 \rfloor}^{\tau+\lfloor w/2 \rfloor} \hat{H}_{\mathrm{MM}}\big(\mathcal{H}^{(t)}\big). \tag{10}$$

Here, the sequence $\big(\mathfrak{h}_w(\tau)\big)_\tau$ constitutes a temporal complexity profile of the hypergraph. Setting $w = T$ recovers the global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ defined in (5). The global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ is useful for between-dataset comparison, but it merges all T snapshots into a single number and therefore cannot reveal whether complexity is rising, falling, or abruptly shifting over time. Therefore, one wants to analyze the sliding-window profile $\big(\mathfrak{h}_w(\tau)\big)_\tau$. Moreover, w controls the trade-off between temporal resolution and statistical stability of each local estimate, which can be domain-dependent.

3.5.3. Change-Point Detection

Having the possibility of calculating a temporal complexity profile, one of its natural applications is the detection of moments at which the structural regime of the hypergraph changes abruptly. We flag a time τ^* as a candidate change point if the profile exhibits a sharp jump at that location. Formally, we write

$$\tau^* \in \mathcal{C} \iff \big|\mathfrak{h}_w(\tau^*) - \mathfrak{h}_w(\tau^* - 1)\big| > \mu + \alpha\sigma, \tag{11}$$

where μ and σ are the mean and standard deviation of the sequence of consecutive absolute differences $\big(|\mathfrak{h}_w(\tau) - \mathfrak{h}_w(\tau - 1)|\big)_\tau$, and $\alpha \geq 1$ is a user-defined sensitivity parameter. This is a lightweight, non-parametric heuristic; obviously, more principled change-point

procedures [29] could be substituted in place of Equation (11) without modifying any other part of the framework. Note that the change-point set $\mathcal{C}$ converts the continuous profile into a discrete list of structural events. From an analytical point of view, this allows one to align detected transitions with domain events, as we will discuss in Section 4.

3.6. The Pipeline

For the ease of the reader in understanding the application of our framework, in Algorithm 1 we summarize the complete computation, from raw temporal hypergraph to complexity profile and change-point set. It takes in input the temporal hypergraph $\boldsymbol{\mathcal{H}} = (\mathcal{H}^{(1)}, \ldots, \mathcal{H}^{(T)})$, the window size w, and the sensitivity parameter α. Steps 1 to 4 instantiate the theoretical layer, i.e., snapshot distributions and entropies, while steps 5 to 7 instantiate the estimation layer, i.e., the sliding-window profile and change-point detection. Finally, note that the overall time complexity is $\mathcal{O}(M)$, where M is the total number of hyperedges in all snapshots.

Algorithm 1 Temporal Hypergraph Entropy Rate Pipeline

Require: Temporal hypergraph $\boldsymbol{\mathcal{H}} = (\mathcal{H}^{(1)}, \ldots, \mathcal{H}^{(T)})$, window size w, sensitivity α
Ensure: Complexity profile $(\mathfrak{h}_w(\tau))_\tau$, change-point set $\mathcal{C}$
1: **for** $t = 1$ **to** T **do**
2: Compute $p_s^{(t)}$ from $\mathcal{E}^{(t)}$ via (3)
3: Compute $\hat{H}_{\mathrm{MM}}(\mathcal{H}^{(t)})$ via (9)
4: **end for**
5: **for** $\tau = \lceil w/2 \rceil$ **to** $T - \lfloor w/2 \rfloor$ **do**
6: Compute $\mathfrak{h}_w(\tau)$ via (10)
7: **end for**
8: Compute μ, σ of consecutive differences
9: $\mathcal{C} \leftarrow \emptyset$
10: **for** each valid τ **do**
11: **if** $|\mathfrak{h}_w(\tau) - \mathfrak{h}_w(\tau - 1)| > \mu + \alpha\sigma$ **then**
12: $\mathcal{C} \leftarrow \mathcal{C} \cup \{\tau\}$
13: **end if**
14: **end for**
return $(\mathfrak{h}_w(\tau))_\tau, \mathcal{C}$

4. Experiments

In this section, we present the experimental campaign designed to assess the usefulness of our measure. After introducing the datasets and setup (Section 4.1), we organize the experiments in four parts. We first report the core results (Section 4.2): the global entropy rate and the significance of its cross-dataset differences, the temporal complexity profiles and their change points, and the clique-expansion baseline. We then study how the measure relates to alternative descriptors (Section 4.3), comparing it against pairwise and higher-order graph descriptors, simpler size statistics, a permutation null model, and established change-point methods. Next, we examine the robustness of the framework to its parameters and preprocessing choices (Section 4.4). Finally, we assess the computational scalability of the pipeline (Section 4.5), verifying empirically that it scales linearly with the total number of hyperedges.

4.1. Datasets and Experimental Setup

We evaluate the framework on six publicly available temporal hypergraph benchmark datasets spanning different domains [30]:

- coauth-DBLP: Each hyperedge is the author set of a publication indexed by DBLP; timestamps are publication years. We restrict to 1970 onward, where annual coverage is sufficiently dense, thus having $T = 49$ annual snapshots.
- coauth-MAG-History: Each hyperedge is the author set of a history publication in the Microsoft Academic Graph [31]; timestamps are publication years. We apply the same 1970 cutoff as for coauth-DBLP, giving $T = 49$ annual snapshots. We include these dataset specifically to contrast with coauth-DBLP: the two share interaction type, format, and temporal resolution, but represent structurally different collaboration cultures.
- email-Enron: Each hyperedge comprises the sender and all recipients of an email among a core set of Enron employees. We aggregate into weekly snapshots ($T = 170$) covering December 1998 through early 2002.
- congress-bills: Each hyperedge is the set of sponsors and co-sponsors of a bill introduced in the US Congress [32,33]; timestamps are days from the opening of the 93rd Congress (3 January 1973). We aggregate into 30-day snapshots ($T = 372$), covering 1973 through 2016.
- tags-math-sx: Each hyperedge is the set of tags applied to a question on Mathematics Stack Exchange (https://math.stackexchange.com/); (accessed on 30 June 2026) timestamps are question-posting times. We aggregate into weekly snapshots ($T = 372$).
- threads-ask-ubuntu: Each hyperedge is the set of users who contributed to the same thread on Ask Ubuntu (https://askubuntu.com/); (accessed on 30 June 2026) timestamps are thread-creation times. We aggregate into weekly snapshots ($T = 371$), covering August 2010 through August 2017.

As for preprocessing, each hyperedge is retained with its observed size, including hyperedges of size $s = 1$, for instance, sole-authored papers or single-participant threads. We retain singletons because they could carry genuine information about the distribution of interaction orders: the presence of a large fraction of size-1 records is itself a structural property of a dataset, and removing them would discard that signal and change the very distribution whose diversity we set out to measure. Nonetheless, we will also study the impact of excluding size-1 hyperedges subsequently. Furthermore, snapshots containing fewer than two hyperedges after binning are discarded, as the size distribution is undefined for a single hyperedge. In Table 1, we report summary statistics of the datasets after preprocessing.

Table 1. Summary statistics of the datasets. T is the number of snapshots; $\bar{m}$ is the mean number of hyperedges per snapshot; $\bar{s}$ is the mean hyperedge size including size-1 hyperedges.

Dataset	Domain	T	$\bar{m}$	$\bar{s}$
coauth-DBLP [30]	Co-authorship (computer science)	49	75,256.5	2.79
coauth-MAG-History [30,31]	Co-authorship (humanities)	49	32,656.4	1.31
email-Enron [30]	Email communication	170	64.0	2.47
congress-bills [30,32,33]	Legislative cosponsorship	372	701.2	3.66
tags-math-sx [30]	Q&A tag co-occurrence	372	2209.8	2.19
threads-ask-ubuntu [30]	Forum thread membership	371	520.1	1.80

All experiments are implemented in Python 3.12. To foster reproducibility and transparency, the source code used for the experiments is available at the following repository: https://github.com/finalfire/entropy-in-temporal-hypergraphs (accessed on 30 June 2026).

As far as the parameters are concerned, for the sliding-window estimator we use window size $w = 10$. Change-point detection uses the threshold in (11) with a dataset-specific sensitivity parameter α. We use the default $\alpha = 2.0$ for email-Enron, congress-bills, and threads-ask-ubuntu. For tags-math-sx, we raise α to 3.5 due to its very low intrinsic volatility, which will be clear in the following. For the two co-authorship datasets, we use $\alpha = 1.5$. Both exhibit smooth long-term trends in which the standard deviation of consecutive differences is small, and the lower threshold confirms that no abrupt breaks exist even under elevated sensitivity. For each dataset, we additionally compute the entropy rate of its clique-expanded temporal graph as a baseline; per Proposition 1, this is identically zero for all datasets.

4.2. Core Results

We begin with the core empirical findings. We first report the global entropy rate of each dataset and test the significance of the differences between datasets; we then examine how complexity evolves over time through the sliding-window profiles and their change points; and we finally confirm that the clique-expansion baseline is identically zero, isolating the information our measure attributes to the hyperedge-size dimension.

4.2.1. Global Entropy Rate Comparison

With this first experiment, we ask whether the entropy rate assigns meaningfully different values to datasets drawn from different domains and, more stringently, to datasets drawn from the same domain but representing structurally distinct interaction cultures. To this end, we apply Algorithm 1 to each of the six datasets with a window size of $w = 10$. For each dataset, the algorithm first computes $\hat{H}_{\mathrm{MM}}(\mathcal{H}^{(t)})$ for every snapshot $t = 1, \ldots, T$ via Equations (3)–(9). The global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ is then obtained as the arithmetic mean of these T values, as in Equation (5). The sliding-window profile $(\mathfrak{h}_w(\tau))_\tau$ is computed via Equation (10); we report its mean $\bar{\mathfrak{h}}$ and standard deviation $\sigma_{\mathfrak{h}}$ as summary statistics of how complexity evolves over time. Change points are detected from the profile via Equation (11) with the per-dataset sensitivity parameter α discussed in Section 4.1. Finally, we re-run the pipeline on the clique-expanded version of each temporal hypergraph to obtain the pairwise baseline $\mathfrak{h}(\boldsymbol{\mathcal{H}}_{\mathrm{clique}})$. In Table 2, we report all of the obtained quantities.

Table 2. Global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$, sliding-window mean $\bar{\mathfrak{h}}$ and standard deviation $\sigma_{\mathfrak{h}}$ (window $w = 10$), change-point sensitivity α, number of detected change points $|\mathcal{C}|$, and clique-expansion baseline $\mathfrak{h}(\boldsymbol{\mathcal{H}}_{\mathrm{clique}})$. Higher $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ indicates more structurally complex temporal evolution. $\mathfrak{h}(\boldsymbol{\mathcal{H}}_{\mathrm{clique}}) = 0$ for all datasets, confirming Proposition 1 empirically. Datasets are sorted by decreasing $\mathfrak{h}(\boldsymbol{\mathcal{H}})$.

Dataset	$\mathfrak{h}(\boldsymbol{\mathcal{H}})$	$\bar{\mathfrak{h}}$	$\sigma_{\mathfrak{h}}$	α	$\lvert\mathcal{C}\rvert$	$\mathfrak{h}(\boldsymbol{\mathcal{H}}_{\mathrm{clique}})$
congress-bills	2.5500	2.5637	0.2765	2.0	17	0
coauth-DBLP	2.0116	2.0234	0.3445	1.5	0	0
tags-math-sx	1.9586	1.9561	0.1190	3.5	5	0
threads-ask-ubuntu	1.5562	1.5555	0.1682	2.0	16	0
email-Enron	1.1692	1.2129	0.3040	2.0	7	0
coauth-MAG-History	0.9549	0.9945	0.0688	1.5	1	0

The entropy rate discriminates across all six datasets, spanning 1.60 bits from 0.95 (coauth-MAG-History) to 2.55 (congress-bills), with all six values distinct. The measure also discriminates within domains. In fact, the two co-authorship datasets differ by 1.06 bits despite sharing interaction type, format, and temporal resolution, a gap that reflects a structural difference in collaboration culture between computer science and history rather than any artefact of dataset size or format.

Table 2 reveals several insights. First, congress-bills attains the highest entropy rate ($\mathfrak{h} = 2.55$ bits), reflecting the high diversity of cosponsorship group sizes across the US legislative corpus: bills range from solo-sponsored measures to large coalitions, and the balance among these sizes shifts substantially across congressional eras. The standard deviation $\sigma_{\mathfrak{h}} = 0.277$ indicates moderate temporal volatility, consistent with a legislative body whose compositional norms evolve with each new Congress and administration. coauth-DBLP ranks second ($\mathfrak{h} = 2.01$ bits), with the highest standard deviation of the six datasets ($\sigma_{\mathfrak{h}} = 0.345$), which reflects a long-term monotone rise rather than random fluctuation: collaboration group sizes in CS have diversified continuously over five decades, driving a sustained increase in entropy. tags-math-sx ranks third ($\mathfrak{h} = 1.96$ bits) but has the lowest standard deviation ($\sigma_{\mathfrak{h}} = 0.119$). The combination of high entropy and low volatility indicates a stable, near-uniform distribution of tag-set sizes: the community sustains a consistent mix of one-, two-, and three-tag questions week after week. threads-ask-ubuntu is intermediate ($\mathfrak{h} = 1.56$ bits, $\sigma_{\mathfrak{h}} = 0.168$), with a moderately diverse but declining profile reflecting a maturing online community. email-Enron attains the second-lowest entropy rate ($\mathfrak{h} = 1.17$ bits), consistent with the structural constraint of email: most messages go to a small number of recipients, concentrating the size distribution near $s = 2$–3. Yet its standard deviation ($\sigma_{\mathfrak{h}} = 0.304$) is the second largest, indicating that while average complexity is low, it is far from constant in the sense that the Enron corpus undergoes substantial structural evolution over the observation window. Finally, coauth-MAG-History attains the lowest entropy rate ($\mathfrak{h} = 0.95$ bits) and the lowest standard deviation ($\sigma_{\mathfrak{h}} = 0.069$), reflecting the conservative authorship norms of historical scholarship, in which sole- and dual-authored publications predominate and this dominance is stable across the full observation window.

4.2.2. Statistical Significance

The comparison in Section 4.2.1 is based on point estimates of the global entropy rate. To assess whether the observed differences are statistically meaningful, we attach uncertainty intervals to each $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ and test the pairwise differences for significance. Because the global entropy rate is the mean of a temporally autocorrelated sequence of per-snapshot entropies, an ordinary bootstrap that resamples snapshots independently would understate the variance. We therefore use a moving-block bootstrap [34], which resamples contiguous blocks of snapshots and so preserves local temporal dependence. We use block length $L = \lceil T^{1/3} \rceil$ and $B = 10{,}000$ resamples and report 95% percentile intervals.

Table 3 reports the per-dataset intervals. As expected, the two co-authorship datasets, which have only $T = 49$ annual snapshots, carry the widest intervals (half-widths of 0.22 and 0.07 bits), while the datasets with several hundred snapshots are estimated much more tightly (half-widths of 0.03–0.14 bits). We then test all $\binom{6}{2} = 15$ pairwise differences using the difference of the two independent bootstrap distributions; a difference is declared significant at the 5% level when its 95% confidence interval excludes zero.

Table 3. Moving-block bootstrap 95% confidence intervals for the global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ ($B = 10{,}000$ resamples, block length $L = \lceil T^{1/3} \rceil$).

Dataset	T	$\mathfrak{h}(\boldsymbol{\mathcal{H}})$	95% CI
congress-bills	372	2.5500	[2.458, 2.639]
coauth-DBLP	49	2.0116	[1.792, 2.227]
tags-math-sx	372	1.9586	[1.926, 1.990]
threads-ask-ubuntu	371	1.5562	[1.508, 1.603]
email-Enron	170	1.1692	[1.030, 1.300]
coauth-MAG-History	49	0.9549	[0.880, 1.015]

The result is that 14 of the 15 pairwise differences are statistically significant ($p < 0.01$ in every case). In particular, the central within-domain contrast between coauth-DBLP and coauth-MAG-History is highly significant: the difference is 1.057 bits with a 95% confidence interval of $[0.818,\ 1.305]$ and $p < 10^{-4}$, even under this conservative block bootstrap. The single exception is the pair coauth-DBLP ($\mathfrak{h} = 2.012$) and tags-math-sx ($\mathfrak{h} = 1.959$), which differ by only 0.053 bits with a confidence interval of $[-0.170,\ 0.272]$ ($p = 0.63$) and are therefore not statistically distinguishable. We report this openly: while the six datasets are distinct as point estimates and span a range of 1.60 bits, the entropy rate resolves them into clearly separated levels with one tied pair, rather than six strictly ordered values. This does not affect the paper's main claims, which concern the overall spread across domains and the within-domain co-authorship contrast, both of which are statistically robust; it simply delimits the resolution of the measure on the present data. The complete set of all pairwise differences, each with its confidence interval and p-value, is reported in Appendix A.

4.2.3. Temporal Complexity Profiles and Change Points

In this second experiment, we ask whether the entropy rate captures the temporal dynamics of structural complexity, rather than merely summarizing it as a single scalar. Concretely, we examine whether the sliding-window profile $(\mathfrak{h}_w(\tau))_\tau$ reveals meaningful patterns of change over time, such as trends, phase transitions, and abrupt structural breaks, and whether the change points detected by Algorithm 1 align with possible events within the datasets. Figure 1 shows the profiles for all six datasets with $w = 10$, with detected change points marked as vertical dashed lines. For each plot, the x-axis shows the snapshot index, while the y-axis reports the entropy rate (in bits). In what follows, we present an exploratory interpretation rather than a formal validation. The temporal datasets that we consider here do not come with an objective, independently specified list of structural events against which detected change points could be tested, and several of the candidate events (for instance, regularly spaced legislative sessions or software release cycles) are near-periodic, which makes any proximity-based test of limited statistical power. We therefore describe the alignments below as plausible, qualitatively consistent readings of the detected transitions, and we do not claim that they constitute confirmatory evidence. We further stress that, since the framework operates solely on the distribution of hyperedge sizes and has no access to the semantic, political, or behavioral content of the interactions, any correspondence between a detected transition and an external event is necessarily correlational: the entropy rate can indicate when the size distribution changes, but not why, and the explanations we suggest are offered only as candidate readings to be tested by domain-specific analysis beyond the scope of this measure.

Figure 1 reveals several notable patterns. For coauth-DBLP, the profile rises smoothly and near-monotonically from approximately 1.3 bits in the early 1970s to approximately 2.7 bits by 2018. No change points are detected at $\alpha = 1.5$. In this case, the rise is too uniform for any single annual transition to exceed the detection threshold. This is somewhat meaningful in itself: the structural evolution of computer science co-authorship complexity might have no abrupt breaks, only continuous drift. Instead, for coauth-MAG-History, the profile is low and nearly flat throughout, oscillating around $\mathfrak{h} \approx 0.95$ bits. A single change point is detected at snapshot index 43 (year 2013). This is temporally close to the period in which digital-humanities collaboration is often reported to have grown, which may correspond to a modest increase in multi-author history publications [30]; we note this as a tentative association rather than an established cause. Taken together with coauth-DBLP, this pair of profiles illustrates that the entropy rate distinguishes collaboration

cultures within a shared domain: the same measure, under identical conditions, yields qualitatively different shapes and ranges.

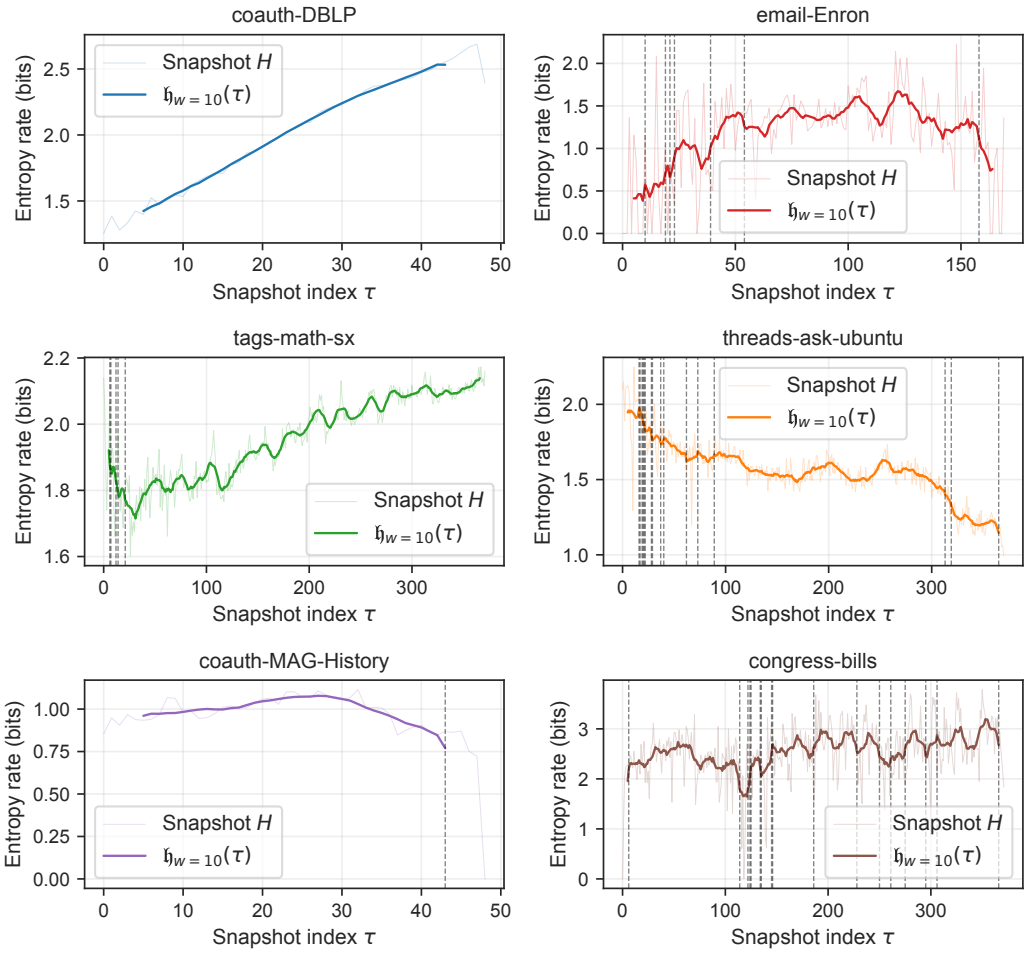

Figure 1. Sliding-window entropy rate profiles for all six datasets ($w = 10$, arranged in a 2×3 grid). The faint trace shows the raw per-snapshot entropy; the solid line is $\mathfrak{h}_w(\tau)$. Vertical dashed lines mark detected change points $\tau^* \in \mathcal{C}$.

The email-Enron dataset instead tells a different story. The profile rises from near zero in early 1999 to approximately 1.3 bits by mid-2000, then remains broadly stable until early 2002. Seven change points are detected. Six (indices 10, 19, 21, 23, 39, 54; March 1999 through January 2000) fall within the first year of the window, when the monitored employee set was still expanding and patterns had not yet stabilized. The seventh change point (index 158; week of 5 January 2002) is qualitatively distinct: it falls approximately five weeks after Enron filed for bankruptcy on 2 December 2001, while we do not establish a causal link, the proximity is suggestive of a lag between the legal event and an observable change in communication patterns. In congress-bills, the profile rises from about 1.8 bits at the opening of the 93rd Congress (1973) to above 3.0 bits in recent sessions, reflecting a long-term increase in cosponsorship diversity, and seventeen change points are detected. Most clusters in the early 1980s (the high-activity 97th–98th Congresses) and several others fall near recognizable legislative milestones such as the 1994 midterm realignment and the 104th Congress [32,33] (https://www.congress.gov/bill/97th-congress/house-bill/4961) (accessed on 30 June 2026). Because congressional sessions recur on a near-regular schedule, we treat these alignments as illustrative temporal coincidences rather than confirmatory evidence. The profile of tags-math-sx is the most stable of the six datasets ($\sigma_{\mathfrak{h}} = 0.119$). Five change points are detected (indices 6, 7, 12,

14, 21), all within the first 21 weeks of the observation window. This early clustering is consistent with the platform's formation phase following the public beta of Mathematics Stack Exchange in October 2010. After approximately five months, the profile stabilizes, and no further change points are detected for the remaining seven years, suggesting that community tagging norms, once established, are remarkably persistent. Finally, the profile of threads-ask-ubuntu declines monotonically from approximately 1.97 bits in August 2010 to approximately 1.15 bits by August 2017. Sixteen change points fall into two groups. The first (indices 16–89; December 2010 through May 2012) coincides with the rapid growth phase of the community in its first two years. The second group (indices 313, 319, 365; August–September 2016 and August 2017) is temporally close to the Ubuntu 16.04 LTS and 17.04 release periods. We mention this as a possible association only because Ubuntu releases recur on a fixed semi-annual schedule; such proximity can also arise by chance. The long-term decline somewhat reflects community maturation: thread participation converges towards smaller, more focused groups of expert contributors, reducing entropy over time.

4.2.4. Comparison with Clique-Expansion Baseline

In this experiment, we ask how much structural complexity is exclusively carried by the higher-order dimension of the data, i.e., how much information would be lost if we resort to projecting the hypergraph onto its pairwise skeleton before applying any complexity measure.

As established by Proposition 1, the entropy rate of the clique-expanded temporal graph is exactly zero for any dataset. Table 2 confirms this empirically for all considered datasets. Indeed, the information gap defined in Equation (7) ranges from 0.95 bits (coauth-MAG-History) to 2.55 bits (congress-bills): the entire signal captured by our measure is attributable to the hyperedge-size dimension of the data, which the size-blind clique expansion discards by construction. Note that this does not imply the information is unrecoverable by every possible pairwise method, but rather that it is lost under the standard projection most commonly used in practice. The 1.06 bit gap between the two co-authorship datasets further demonstrates that the information gap measures genuine structural properties of the interaction data rather than incidental features of dataset scale or format.

4.3. Relation to Alternative Measures

Having established the core results, we now ask whether the entropy rate is genuinely distinct from existing descriptors, or whether a simpler or established alternative would capture the same signal. We compare our profile against a battery of pairwise and higher-order graph descriptors, against a mean-size statistic and a permutation null model, and we validate the change-point detector against several principled change-point methods.

4.3.1. Comparison with Non-Trivial Pairwise and Higher-Order Descriptors

Proposition 1 shows that the size-based entropy rate vanishes on any clique expansion. As this is a definitional consequence of projecting onto size-2 edges, it does not, by itself, establish that the temporal signal captured by our measure is inaccessible to other descriptors, whether pairwise or higher-order. To test this directly, we compute five established descriptors on each snapshot and turn each into a temporal profile using the same sliding window ($w = 10$). Four are pairwise descriptors of the clique-expanded weighted graph: the entropy of the weighted-degree (node-strength) distribution, the entropy of the edge-weight distribution, the von Neumann (spectral) entropy of the Laplacian density matrix [12], and the temporal edge turnover (the Jaccard distance between the edge sets of consecutive snapshots). The fifth is a genuinely higher-order descriptor that operates

on the hypergraph directly. Building on the notion of simplicial closure [30], we measure the fraction of hyperedges of order at least three whose proper subfaces are all present in the same snapshot, a quantity we refer to as the downward-closure fraction. It quantifies downward closure (inclusion) rather than size diversity. We then measure the Pearson correlation between our entropy-rate profile and each descriptor profile. Table 4 reports the results. For the largest snapshots, the von Neumann entropy is computed via the standard degree-based quadratic approximation rather than exact eigen-decomposition, and the downward-closure fraction is evaluated exactly for hyperedges up to order six; neither approximation affects the qualitative conclusions, since the size distributions concentrate on small orders.

Table 4. Pearson correlation between our entropy-rate profile and five descriptor profiles ($w = 10$): weighted-degree entropy (r_{deg}), edge-weight entropy (r_{ew}), von Neumann entropy (r_{vN}), downward-closure fraction (r_{DC}), and edge turnover (r_{turn}). Values near ± 1 indicate the descriptor tracks our profile; values near 0 indicate our profile carries information the descriptor does not. "–" denotes an undefined correlation (constant descriptor profile).

Dataset	r_{deg}	r_{ew}	r_{vN}	r_{DC}	r_{turn}
coauth-DBLP	0.997	0.998	−0.057	−0.807	−0.830
email-Enron	0.851	0.877	0.863	0.471	−0.144
tags-math-sx	0.893	0.901	0.889	0.629	−0.805
threads-ask-ubuntu	−0.592	−0.384	−0.557	–	−0.266
coauth-MAG-History	−0.140	−0.140	0.532	−0.792	0.160
congress-bills	0.456	0.286	0.434	−0.008	−0.293

On three datasets, namely, coauth-DBLP, email-Enron, and tags-math-sx, our entropy-rate profile is strongly correlated with the weighted-degree and edge-weight entropies (up to $r = 0.998$ for coauth-DBLP). On these datasets, the hyperedge-size distribution and the pairwise degree structure co-evolve: as collaborations or interactions grow, both the group sizes and the node degrees increase together, so a pairwise degree descriptor tracks our measure closely. We do not claim, therefore, that the entropy rate is orthogonal to pairwise descriptors, in general,; on datasets where size and connectivity move together, a standard pairwise entropy can be an effective proxy for it.

Two findings nonetheless show that the measure is not reducible to these descriptors, in general,. First, on the remaining three datasets the profiles decouple: for threads-ask-ubuntu the correlations with degree, edge-weight, and von Neumann entropy are weak or negative (-0.38 to -0.59), for coauth-MAG-History the degree and edge-weight correlations are essentially zero (-0.14), and for congress-bills they are only moderate (0.29 to 0.46). On these datasets the size mix and the pairwise connectivity evolve differently, and the pairwise descriptors do not reproduce our profile. Of the eighteen comparisons against the three entropy descriptors, seven fall below $|r| = 0.5$. Second, and more uniformly, edge turnover is weakly or negatively correlated with our profile on every dataset (r between -0.83 and $+0.16$). The information our measure captures is thus never well explained by how quickly the pairwise edge set churns.

The same pattern holds against the higher-order baseline. The downward-closure fraction measures downward closure rather than size diversity, and it does not reproduce our profile either: the two are essentially independent on congress-bills ($r = -0.008$), moderately correlated on email-Enron and tags-math-sx ($r = 0.47$ and 0.63), and strongly inversely correlated on the two co-authorship datasets ($r \approx -0.80$). The inverse relationship is interpretable rather than incidental: as collaboration sizes diversify and the entropy rate rises, large author teams appear whose smaller sub-collaborations are not separately recorded, so downward closure falls. The two measures thus move in opposite directions

because they track different structural facts. On threads-ask-ubuntu the downward-closure fraction is constant across snapshots (the data contain almost no closed higher-order simplices), so the correlation is undefined. We conclude that the established higher-order temporal descriptor does not subsume our measure: it is independent of it on one dataset, inversely related on two, moderate on two, and uninformative on one.

To conclude this experiment, we note that the clique-expansion result of Proposition 1 is a definitional property, not an empirical demonstration of superiority, and on datasets where interaction size and node degree co-evolve, our measure can indeed be approximated by a pairwise degree entropy. Where the framework adds value is on datasets whose higher-order size structure evolves independently of pairwise connectivity and of downward closure: here, the entropy rate behaves differently from every descriptor we tested, pairwise and higher-order alike. The measure should therefore be understood as a compact, dataset-agnostic, and linear-time descriptor of interaction-order diversity that is complementary to existing pairwise and higher-order descriptors, coinciding with them when the underlying structural dimensions are entangled and diverging from them when they are not, rather than as a uniformly alternative. A fuller comparison against other higher-order descriptors, motif-based profiles, higher-order random-walk entropy, and spectral hypergraph complexity would further sharpen this picture; each targets a distinct structural axis (local topology, diffusion dynamics, and spectral structure, respectively), and we leave their systematic integration to future work.

4.3.2. Comparison with Additional Baselines

Finally, we evaluate the framework against two additional baselines designed to address two distinct questions, namely: (i) does the entropy rate convey information beyond what a simpler size statistic could provide?; (ii) does the temporal ordering of snapshots carry genuine structure, or are the observed profiles consistent with a random arrangement of the same data? To this end, we design two baselines, namely, (i) mean hyperedge size and (ii) permutation null model, which we introduce in the following.

The former baseline is the simplest non-trivial alternative to the entropy rate. We formally define it as $\bar{s}^{(t)} = \sum_s s \cdot p_s^{(t)}(s)$, which is also a marginal statistic derived solely from the size sequence but requires no information–theoretic machinery. We compute a sliding-window profile of $\bar{s}^{(t)}$ using the same window size $w = 10$ and apply the same change-point detector with the same per-dataset α values. The resulting change-point counts are reported as $|\mathcal{C}_{\bar{s}}|$ in Table 5.

Table 5. Results for baseline comparison. $|\mathcal{C}|$ is the observed change-point count (entropy rate). $|\mathcal{C}_{\bar{s}}|$ is the change-point count for the mean hyperedge size baseline. z_σ is the z-score of observed profile standard deviation $\sigma_\mathfrak{h}$ under the permutation null (200 permutations, fixed seed). $z_{|\mathcal{C}|}$ is the z-score of observed change-point count under the permutation null.

Dataset	$\lvert\mathcal{C}\rvert$	$\lvert\mathcal{C}_{\bar{s}}\rvert$	z_σ	$z_{\lvert\mathcal{C}\rvert}$
congress-bills	17	14	6.99	0.12
coauth-DBLP	0	2	7.23	−3.25
tags-math-sx	5	5	24.36	14.48
threads-ask-ubuntu	16	12	19.60	0.08
email-Enron	7	9	7.40	−0.59
coauth-MAG-History	1	1	2.65	−1.36

The latter baseline is useful to assess whether the temporal ordering of snapshots carries genuine structure beyond what a random arrangement of the same data would produce. For each dataset, we generate 200 permutations of the snapshot order, recompute the sliding-window profile for each, and record two order-sensitive quantities: the profile

standard deviation $\sigma_{\mathfrak{h}}$, which measures the overall volatility of the complexity trajectory, and the change-point count $|\mathcal{C}|$. We report z-scores measuring how many standard deviations the observed values lie above the permutation null means [35]. Note that the global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ itself is order-invariant by construction, due to it being a time-average of per-snapshot entropies, and is therefore excluded from this test. Indeed, the null comparison is meaningful only for quantities that depend on the temporal arrangement of snapshots. Also in this case, the results are reported in Table 5.

Several insights emerge from Table 5. First, the two statistics are not equivalent. On two datasets (tags-math-sx and coauth-MAG-History), they produce identical change-point counts, indicating that the signal is strong enough for a simpler summary to capture it. On the remaining four, they diverge: for coauth-DBLP, the mean-size profile flags two spurious change points that the entropy rate correctly suppresses, consistent with the fact that the entropy profile is a genuinely smooth monotone ramp; for email-Enron, the mean-size profile produces two additional detections beyond the entropy rate's seven, adding noise rather than resolution; and for threads-ask-ubuntu and congress-bills, the entropy rate detects four and three more change points, respectively, than the mean-size profile, suggesting that the entropy captures higher-order distributional changes that the mean misses. Taken together, the entropy rate is the more consistent and less noisy of the two statistics: it avoids spurious detections when the profile is smooth and provides additional sensitivity when the distributional structure is richer than a single moment can capture. Moreover, the z_σ results are consistent across all six datasets: every observed profile standard deviation is significantly larger than the permutation null (z-scores ranging from 2.65 for coauth-MAG-History to 24.36 for tags-math-sx), confirming that the temporal ordering of snapshots produces volatility patterns that are highly unlikely under a random arrangement. The entropy rate framework is therefore measuring genuine temporal structure in every dataset, not merely the marginal distribution of per-snapshot entropy values.

Interestingly, the $z_{|\mathcal{C}|}$ results are more informative precisely because they differ across datasets, revealing qualitatively distinct structural regimes. The strongest case is tags-math-sx ($z_{|\mathcal{C}|} = 14.48$): the five detected change points are far more concentrated than any random permutation would produce (permutation null 95th percentile: 1). This confirms that the changes are not noise but reflect a genuine structural phase transition at platform launch, after which the tagging regime stabilizes. coauth-DBLP ($z_{|\mathcal{C}|} = -3.25$) is equally informative in the opposite direction: the zero observed change points are significantly fewer than the permutation null would produce (null mean: 3.6; null 95th percentile: 5). A random ordering of the same 49 annual snapshots would generate change points on average; the real temporal sequence does not. This confirms that the absence of change points is a genuine property of the smooth monotone trajectory, not a failure of the detector. email-Enron, threads-ask-ubuntu, and congress-bills all have $z_{|\mathcal{C}|}$ near zero, meaning their observed change-point counts are consistent with a random ordering of the same snapshots. For these datasets, the value of the framework lies not in the raw count of change points but in the specific temporal locations of the detected transitions, which are interpretable against external ground truth that the null cannot access. The high z_σ values for these three datasets (6.99, 19.60, and 7.40, respectively) confirm that their profiles are genuinely structured; the change points are real events embedded in a non-random temporal trajectory.

4.3.3. Validation of the Change-Point Detector

Our change-point detector (Equation (11)) is deliberately lightweight, and it is therefore important to verify that the transitions it reports reflect genuine structure rather than artifacts of the thresholding heuristic. We bring three pieces of evidence to bear, two of which are already established in the previous subsections. First, the permutation-null analysis of

Section 4.3.2 shows that the detected counts and the profile volatility are highly unlikely under a random ordering of the same snapshots, so the structure the detector responds to is genuinely temporal. Second, the threshold-sensitivity analysis of Section 4.4.2 (reported later, among the robustness analyses) shows that the detected count varies smoothly and monotonically with α, with no value at which it changes abruptly, and that the strongest transitions persist across the whole range. Third, and to address the concern directly, we now compare our detector against a principled, established change-point method.

We use four established change-point methods spanning distinct algorithmic families: the Pruned Exact Linear Time (PELT) algorithm [29] and binary segmentation [29], both penalty-based segmentation methods; the classical CUSUM (cumulative sum) sequential control chart [36]; and Bayesian Blocks [37], a Bayesian segmentation method. All are applied to the same sliding-window profiles with standard penalties or priors and no per-dataset tuning. For each dataset, we match our detected change points to those of each method, counting a match when a change point lies within three snapshots of the other method's nearest boundary. We report the results of this comparison in Table 6.

Table 6. Comparison of our threshold-based change-point detector with four established methods on the same profiles: PELT and binary segmentation (Binseg) [29], CUSUM [36], and Bayesian Blocks (BB) [37]. For each method, we report its detected count and the number of our change points confirmed by it (within three snapshots).

Dataset	$\lvert\mathcal{C}\rvert_{\text{ours}}$	$\lvert\mathcal{C}\rvert_{\text{PELT}}$	m	$\lvert\mathcal{C}\rvert_{\text{CUSUM}}$	m	$\lvert\mathcal{C}\rvert_{\text{Binseg}}$	m	$\lvert\mathcal{C}\rvert_{\text{BB}}$
congress-bills	17	43	15	9	3	42	14	8
coauth-DBLP	0	6	0	1	0	6	0	1
tags-math-sx	5	37	4	2	0	33	4	3
threads-ask-ubuntu	16	32	12	2	1	29	11	3
email-Enron	7	14	7	3	2	15	7	3
coauth-MAG-History	1	7	1	1	1	7	1	1

The four methods fall into two groups of differing sensitivity, and our detector sits between them. The two segmentation methods, PELT and binary segmentation, report more change points than we do (typically two to eight times as many), because they partition the whole profile into homogeneous pieces and so subdivide even smooth trends; relative to them, our change points are a confirmed subset, with 39 of our 46 points (85%) matched by PELT and 37 (80%) by binary segmentation. The CUSUM chart and Bayesian Blocks, by contrast, are more conservative than our detector, reporting fewer points (one to nine); here the relationship inverts, and the points they find are essentially a subset of ours. This is most striking for Bayesian Blocks: on the three datasets where it detects a non-trivial number of transitions, namely, congress-bills, threads-ask-ubuntu, and email-Enron, all change points (8/8, 3/3, and 3/3, respectively) coincide with one of ours. In other words, no principled method, from the most liberal to the most conservative and including a Bayesian one, identifies strong structure where our detector reports none, and the conservative methods confirm precisely the most prominent transitions we flag.

The clearest illustration is coauth-DBLP, whose profile is a smooth monotone rise: our detector reports no change points, the conservative methods (CUSUM and Bayesian Blocks) report at most one, and only the segmentation methods impose several boundaries on the gradual drift. Our abstention there is the desired behavior, and it is independently corroborated by the permutation null of Section 4.3.2, where coauth-DBLP has $z_{|\mathcal{C}|} = -3.25$, i.e., significantly fewer change points than a random ordering would produce. In summary, the detector's positive detections agree with multiple principled methods spanning penalty-based, sequential, and Bayesian families, and its conservatism is a deliberate and validated design choice rather than a source of spurious detections.

4.4. Robustness

We now assess how the framework behaves under variation of its design choices: the window size w, the change-point threshold α, the treatment of singleton hyperedges, and the entropy estimator.

4.4.1. Sensitivity to the Window Size

The sliding-window estimator uses a window size w, set to $w = 10$ in the experiments above. We now examine how the results change as w ranges over $\{5, 10, 20\}$. Before reporting the table, we note an important structural fact: the global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ of Definition 4 does not depend on w at all, since it is the time average of the per-snapshot entropies and the window enters only the sliding-window profile of Equation (10). The window affects only two downstream quantities: the volatility of the profile, $\sigma_{\mathfrak{h}}$, and the change-point count $|\mathcal{C}|$. In Table 7, we report the profile mean $\bar{\mathfrak{h}}$, the volatility $\sigma_{\mathfrak{h}}$, and $|\mathcal{C}|$ for each window size.

Table 7. Sensitivity to the window size $w \in \{5, 10, 20\}$. For each w we report the profile mean $\bar{\mathfrak{h}}$, the profile volatility $\sigma_{\mathfrak{h}}$, and the change-point count $|\mathcal{C}|$ (per-dataset α as in Table 2).

Dataset	$\bar{\mathfrak{h}}$			$\sigma_{\mathfrak{h}}$			$\lvert\mathcal{C}\rvert$		
	$w = 5$	$w = 10$	$w = 20$	$w = 5$	$w = 10$	$w = 20$	$w = 5$	$w = 10$	$w = 20$
congress-bills	2.561	2.564	2.563	0.354	0.277	0.238	18	17	18
coauth-DBLP	2.019	2.023	2.030	0.393	0.344	0.258	1	0	0
tags-math-sx	1.957	1.956	1.955	0.121	0.119	0.117	5	5	4
threads-ask-ubuntu	1.556	1.556	1.555	0.180	0.168	0.157	12	16	14
email-Enron	1.188	1.213	1.244	0.360	0.304	0.251	8	7	10
coauth-MAG-History	0.980	0.995	1.011	0.090	0.069	0.036	1	1	1

Three observations follow. First, the profile mean $\bar{\mathfrak{h}}$ is essentially unchanged across window sizes: the largest variation over the whole range is 0.056 bits (email-Enron), and the ordering of the six datasets by entropy rate is identical at every w. The window size therefore does not affect the cross-dataset conclusions. Second, the profile volatility $\sigma_{\mathfrak{h}}$ decreases monotonically with w for every dataset, exactly as expected: a larger window averages over more snapshots and smooths the profile. This is a predictable, well-behaved dependence rather than a source of instability, and the relative ordering of the datasets by volatility is preserved. Third, the change-point count is stable: the two co-authorship datasets yield zero or one change point at every w, and the active datasets vary by only a few detections (for instance, congress-bills stays at 17–18 and tags-math-sx at 4–5) while the strongest transitions persist. The choice $w = 10$ sits in the middle of this range and balances temporal resolution against the stability of each local estimate; the results are not sensitive to that choice.

4.4.2. Sensitivity to the Change-Point Threshold

The change-point detector of Equation (11) depends on a single sensitivity parameter α, which we set per dataset in the experiments above. We now report how the number of detected change points $|\mathcal{C}|$ varies as α ranges over the values $\{1.5, 2.0, 2.5, 3.0, 3.5\}$, holding the window size fixed at $w = 10$. Note that the sliding-window profile itself does not depend on α; only the detection step is repeated. We report the results in Table 8.

Several observations follow from Table 8. First, the detected count is a smooth, monotonically non-increasing function of α for every dataset: raising the threshold removes the weakest detections first, and there is no value of α at which the count changes abruptly. The detector therefore behaves predictably, and small perturbations of α do not qualitatively alter the results. Second, the two co-authorship datasets, which carry the central

within-domain contrast of our study, are entirely insensitive to α: coauth-DBLP yields zero change points and coauth-MAG-History yields exactly one across the entire set of values. The conclusion that computer-science co-authorship evolves through smooth drift while historical co-authorship is essentially static thus holds independently of the threshold, as does the global entropy rate contrast. Finally, the four higher-activity datasets (congress-bills, threads-ask-ubuntu, email-Enron, and tags-math-sx) exhibit a dependence on α, as expected for any threshold-based detector: a more conservative threshold reports fewer transitions. We emphasize that our interpretation of these datasets in Section 4.2.3 rests on the locations of the most prominent transitions rather than on their exact number, and the strongest change points (for instance, the post-bankruptcy transition in email-Enron and the platform-formation cluster in tags-math-sx) persist across the grid. The per-dataset values $\alpha^\star$ were chosen to reflect each profile's intrinsic volatility $\sigma_\mathfrak{h}$: the least volatile profile (tags-math-sx, $\sigma_\mathfrak{h} = 0.119$) requires the most conservative threshold ($\alpha^\star = 3.5$) to avoid flagging routine fluctuation as structural change, whereas the most volatile profile (coauth-DBLP, $\sigma_\mathfrak{h} = 0.345$) tolerates the least conservative threshold ($\alpha^\star = 1.5$). The relationship is not strict, yet it shows that $\alpha^\star$ tracks a measurable property of each profile rather than being tuned to produce a specific count.

Table 8. Sensitivity of the detected change-point count $|\mathcal{C}|$ to the threshold parameter α (window $w = 10$). $\sigma_\mathfrak{h}$ is the profile volatility (standard deviation of the sliding-window profile) and $\alpha^\star$ is the per-dataset value used in the main experiments. The column matching $\alpha^\star$ is shown in bold. Counts are non-increasing in α for every dataset.

Dataset	$\sigma_\mathfrak{h}$	$\alpha^\star$	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 2.5$	$\alpha = 3.0$	$\alpha = 3.5$
congress-bills	0.2765	2.0	25	**17**	11	7	4
coauth-DBLP	0.3445	1.5	**0**	0	0	0	0
tags-math-sx	0.1190	3.5	21	12	11	7	**5**
threads-ask-ubuntu	0.1682	2.0	22	**16**	9	7	5
email-Enron	0.3040	2.0	15	**7**	5	1	1
coauth-MAG-History	0.0688	1.5	**1**	1	1	1	1

4.4.3. Sensitivity to Singleton Hyperedges

We now study the measure when singletons, i.e., size-1 hyperedges, are excluded. Starting from the considered hypergraphs in Table 2, we remove size-1 hyperedges and compute the global entropy rate $\mathfrak{h}(\boldsymbol{\mathcal{H}})$, the profile volatility $\sigma_\mathfrak{h}$, and the change-point count $|\mathcal{C}|$. We report the results in Table 9. Here, the first column f_1 is the fraction of size-1 hyperedges in the original data. As far as the measures are concerned, the subscript "incl" (resp., "excl") indicates that the measure is computed on the hypergraphs including (excluding, respectively) singletons.

The effect of the preprocessing rule depends systematically on the singleton fraction f_1, and its direction is somewhat informative. When singletons dominate a dataset ($f_1 > 0.5$), excluding them removes the modal category and exposes the diversity among the remaining group sizes, so the entropy rate rises: for coauth-MAG-History ($f_1 = 0.83$), it increases from 0.95 to 1.60 bits, and for congress-bills ($f_1 = 0.59$), from 2.55 to 3.64 bits. When singletons are a minority well mixed with larger sizes, excluding them instead removes a category that was contributing diversity, so the entropy rate falls relatively; for example, coauth-DBLP drops from 2.01 to 1.74 bits and email-Enron from 1.17 to 1.01 bits. This confirms that singletons are not a neutral artifact to be filtered away; for datasets such as coauth-MAG-History, where sole-authored papers constitute the large majority of records, the prevalence of singletons is the dominant structural fact, and a measure of interaction-order diversity should reflect this, indicating that they should be retained.

Table 9. Effect of including ("incl") versus excluding ("excl") size-1 hyperedges; window size is $w = 10$, α is set per-dataset as in Table 2; results are sorted by decreasing f_1 (fraction of singletons in the original data).

Dataset	f_1	$\mathfrak{h}_{\text{incl}}$	$\mathfrak{h}_{\text{excl}}$	σ_{incl}	σ_{excl}	$\lvert\mathcal{C}\rvert_{\text{incl}}$	$\lvert\mathcal{C}\rvert_{\text{excl}}$
coauth-MAG-History	0.831	0.9549	1.5960	0.0688	0.1024	1	1
congress-bills	0.594	2.5500	3.6418	0.2765	0.1651	17	10
threads-ask-ubuntu	0.390	1.5562	1.0464	0.1682	0.1709	16	13
tags-math-sx	0.321	1.9586	1.5734	0.1190	0.1324	5	7
coauth-DBLP	0.200	2.0116	1.7425	0.3445	0.3139	0	0
email-Enron	0.040	1.1692	1.0119	0.3040	0.2725	7	8

Also, two robustness observations follow. First, the change-point counts of the two co-authorship datasets are unchanged under both rules (coauth-DBLP: 0; coauth-MAG-History: 1), and the relative ordering of the datasets by entropy rate is largely preserved (congress-bills remains the most complex under both rules). Second, we note that the magnitude of the within-domain co-authorship gap does depend on this choice: under inclusion the gap between coauth-DBLP and coauth-MAG-History is 1.06 bits, whereas under exclusion it narrows to roughly 0.15 bits, because exclusion disproportionately raises the entropy rate of the singleton-dominated coauth-MAG-History. The qualitative conclusion holds under both rules, but its quantitative size is specific to the treatment of size-1 hyperedges. We regard this as the precise, transparent characterization of the preprocessing effect that Table 9 is intended to provide.

4.4.4. Validation of the Entropy Estimator

As introduced in Section 3.5.1, we use the Miller–Madow correction (Equation (9)) to reduce the finite-sample bias of the entropy estimator in Equation (4), which we denote as "plugin" here. To motivate this choice, we conduct an experiment simulating with known ground-truth size distributions and compare the Miller–Madow estimator against the naive plugin estimator and three classical alternatives: a Bayesian Dirichlet smoother (Krichevsky–Trofimov, add-$\frac{1}{2}$ [38]), the coverage-adjusted Chao–Shen estimator [39], and a nonparametric bootstrap bias correction [40]. The first replaces the empirical counts with a Dirichlet-smoothed distribution; the second corrects the plugin estimator using an estimate of the sample coverage; finally, the last resamples the observed hyperedge sizes with replacement to estimate the plugin estimator's bias directly. We build four ground-truth distributions chosen to resemble the empirical hyperedge-size distributions in our datasets: a geometric-decay distribution (most interactions small), a near-uniform distribution over a few sizes (as in tags-math-sx), a power-law distribution with exponent $\alpha = 2$ (broad range of sizes, as in congress-bills), and a dominant-mode distribution (one size carrying most of the mass). For each ground-truth distribution, we test how accurately each estimator recovers its known entropy as the sample size grows. We use sample sizes m spanning the per-snapshot hyperedge counts in our datasets, from $m = 1$ to $m = 65{,}536$ (recall that $\bar{m}$ ranges from 64 for email-Enron to over 75,000 for coauth-DBLP). For each distribution and each m, we draw m sizes at random, apply each estimator, and record the error (estimate minus true entropy), averaging over 2000 repetitions to obtain the typical bias.

We show in Figure 2 the bias of each estimator as a function of m, and in Table 10 a summary of the bias at $m = 64$. Table 10 and Figure 2 together support several observations. First, the naive plugin estimator is substantially biased downward at small sample sizes, confirming that a correction is necessary. Second, the Miller–Madow correction removes most of this bias: at $m = 64$, its absolute bias is at most 0.04 bits across all four distributions, a three- to five-fold reduction relative to the plugin estimator. Third, the more elaborate

estimators are competitive with Miller–Madow but do not consistently outperform it: the Dirichlet and bootstrap corrections are comparable, and while the Chao–Shen estimator is the most accurate at the very smallest samples ($m \leq 32$) on the most skewed distributions, it is no better, and sometimes worse, in the near-uniform and dominant-mode cases. No single estimator dominates. Furthermore, by $m \geq 256$ all estimators, including the naive plugin, have negligible bias (below 0.02 bits). Five of our six datasets have mean snapshot sizes well above this threshold ($\bar{m}$ between 520 and 75,256), so for them the choice of estimator is immaterial. Only email-Enron ($\bar{m} = 64$) lies in the regime where the correction has any practical effect, and there the Miller–Madow estimator already reduces the bias to under 0.03 bits. These results motivate our adoption of the Miller–Madow correction. It is accurate where accuracy is needed and negligible elsewhere, yet far cheaper than the bootstrap, which would require resampling at every snapshot, and simpler than the coverage-adjusted and Bayesian alternatives, whose marginal accuracy gains do not materialize in the sample-size regime our datasets occupy.

Table 10. Estimator bias (estimated entropy minus true entropy, in bits) at sample size $m = 64$, the smallest mean snapshot size among our datasets (email-Enron). Values are averaged over 2000 repetitions.

Estimator	Geometric	Near-Unif.	Power-Law	Dom.-Mode
plugin (naive)	−0.086	−0.034	−0.137	−0.063
Miller-Madow (ours)	−0.023	−0.000	−0.038	−0.011
Dirichlet (KT)	−0.016	−0.032	−0.023	+0.052
Chao–Shen	−0.003	−0.034	+0.024	+0.043
bootstrap	−0.018	+0.000	−0.029	−0.005

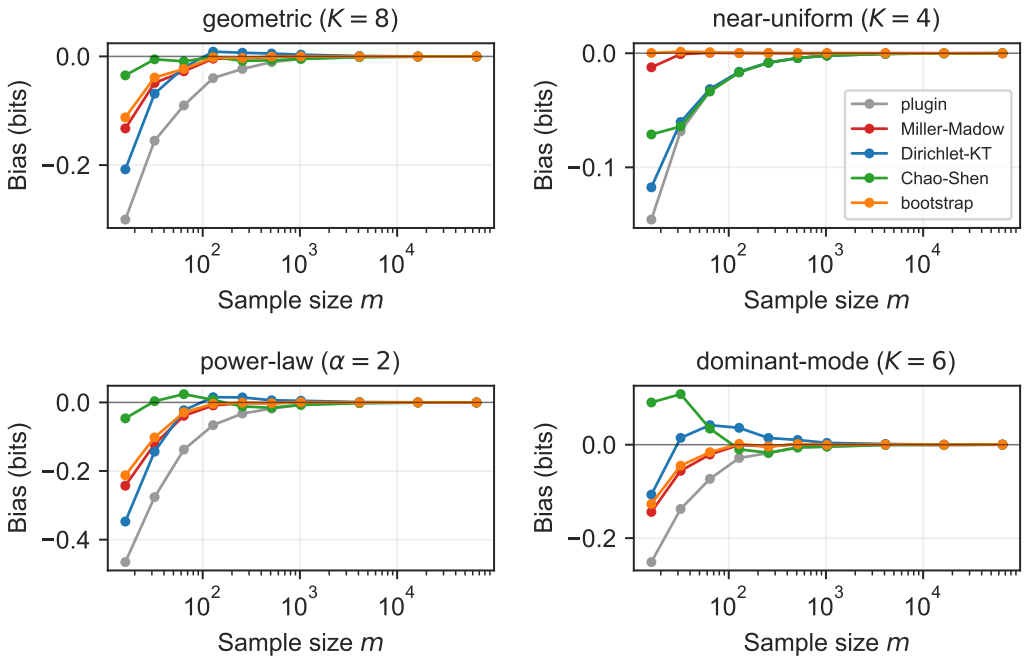

Figure 2. Bias (estimated minus true entropy, in bits) of five entropy estimators as a function of sample size m, for four ground-truth size distributions resembling those in our datasets.

4.5. Scalability Analysis

The fifth and final experiment asks whether the pipeline remains computationally tractable as the dataset size grows, both in terms of the number of snapshots T and in terms of the number of hyperedges per snapshot m_t. Note that in Section 3.6 we establish an

$\mathcal{O}(M)$ time bound, where $M = \sum_{t=1}^{T} m_t$ is the total hyperedge count; in what follows, we verify this bound empirically and quantify the practical constant.

As the experiment setting, we measure wall-clock time for the core pipeline steps (per-snapshot size distribution, Miller–Madow entropy, and sliding-window profile, as reported in Algorithm 1, lines 2–7) using two controlled experiments. In the first, we vary T by taking contiguous prefixes of each dataset at 25%, 50%, 75%, and 100% of the full sequence, holding $\bar{m}$ approximately constant. In the second, we vary m_t by uniformly subsampling hyperedges within each snapshot at fractions $\{0.10, 0.25, 0.50, 0.75, 1.00\}$ of the full count, holding $T = 49$ fixed; we use the two largest datasets for this experiment, namely, coauth-DBLP, with $\bar{m} \approx 75{,}000$, and coauth-MAG-History, with $\bar{m} \approx 32{,}000$. Each configuration is timed over ten repetitions, and we report the mean.

The results of these experiments are depicted in Figure 3, and they show that the runtime grows linearly with M. We fit a linear model $t = a \cdot M$ (forced through the origin), which gives $R^2 \geq 0.999$ for five of the six datasets in the varying-T experiment and $R^2 \geq 0.999$ for both datasets in the varying-m experiment. The single exception is email-Enron ($R^2 = 0.961$), whose absolute runtimes lie in the sub-millisecond range (0.1–0.8 ms), which we believe is due to the implementation rather than any algorithmic effect. The fitted slope is consistent across datasets, ranging from 0.027 to 0.037 µs per hyperedge for the five datasets with reliable timing, confirming that the constant in the $\mathcal{O}(M)$ bound is both small and stable across domains. At full scale, the largest dataset (coauth-DBLP, $M = 3{,}687{,}570$ total hyperedges) completes in 101 ms; the smallest (email-Enron, $M = 10{,}876$) completes in under 1 ms. The complete pipeline for all six datasets runs in well under one second of total CPU time, making the framework practical even on modest hardware and at dataset scales substantially larger than those considered here.

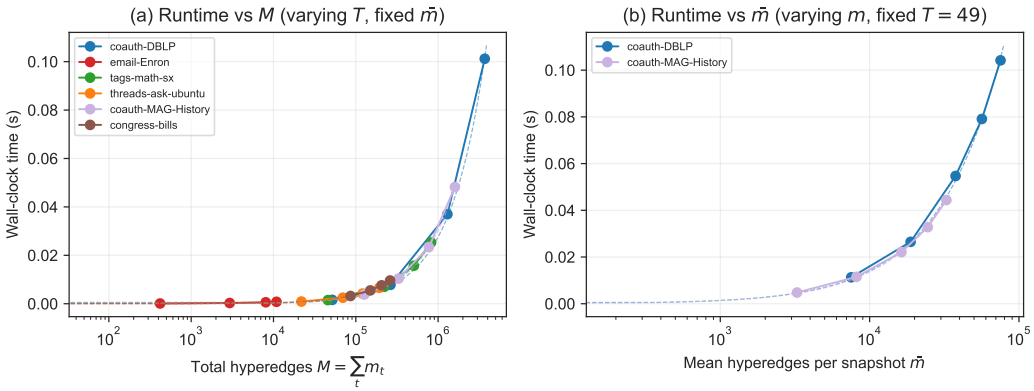

Figure 3. Results on the scalability of Algorithm 1. (**a**) wall-clock time vs. total hyperedge count M as T varies, for all datasets. (**b**) wall-clock time vs. mean hyperedges per snapshot $\bar{m}$ as hyperedge density varies (fixed $T = 49$), for the two largest datasets.

5. Discussion and Limitations

The proposed series of experiments confirms that temporal hypergraph complexity is a meaningful and measurable property that goes beyond what pairwise analysis can reveal. Across six datasets spanning five domains, the entropy rate spans 1.60 bits (from 0.95 to 2.55) with all values distinct, and the clique-expansion baseline is identically zero in every case. Particularly, the within-domain contrast between coauth-DBLP ($\mathfrak{h} = 2.01$ bits) and coauth-MAG-History ($\mathfrak{h} = 0.95$ bits) is instructive: the two datasets share interaction type, format, and temporal resolution, yet the entropy rate reveals a 1.06-bit gap that reflects a genuine difference in collaboration culture between computer science and historical scholarship. This result could have been overlooked by pairwise projection, since both datasets would collapse to an entropy rate of zero under clique expansion.

Beyond the experimental validation, the framework has concrete practical utility for analysts working especially with large-scale temporal interaction data. As a first application, the entropy rate pipeline can serve as a rapid structural characterization tool at dataset intake: given a new temporal hypergraph, Algorithm 1 completes in time linear in the total hyperedge count and immediately produces a complexity profile that reveals whether the interaction size distribution is stable, volatile, monotonically drifting, or punctuated by abrupt transitions. Such information is directly actionable: it could guide the choice of a downstream modeling approach before any expensive computation is performed, and it could flag datasets where a simple pairwise model would lose a quantifiable amount of structural information. Another application is the use of the unsupervised change-point detector, which provides a lightweight monitoring tool for large communication and collaboration platforms: structural regime changes are detected automatically, without labeled examples or domain-specific tuning beyond the sensitivity parameter α. As a third application, the dataset-agnostic design of the measure enables principled comparative benchmarking: since the entropy rate requires no node alignment and makes no assumptions about network size or domain, it can be used to rank or cluster a heterogeneous collection of temporal hypergraph datasets by structural complexity, providing a global descriptor that complements local measures such as degree distributions or motif counts.

Furthermore, it is useful to situate our work within the rapidly growing recent literature on higher-order and temporal hypergraph analysis. One prominent line of work pursues expressive, learning-based models of higher-order temporal structure. Heterogeneous temporal hypergraph neural networks, for instance, combine hierarchical attention with temporal message-passing to capture higher-order group interactions for downstream prediction tasks [41], and higher-order structure has been shown to improve link prediction on temporal graphs more broadly; the representation-learning perspective on higher-order networks is surveyed in [42]. A second, complementary line focuses on mining recurrent higher-order patterns: a recent survey lists the patterns, tools, and generators for hypergraph mining [43], and new structural primitives, such as hypermotifs, have been proposed as higher-order fingerprints of real-world hypergraphs [44]. Our work is positioned differently from both lines. Rather than learning a parameterized model or enumerating local higher-order patterns, we provide a single, closed-form, parameter-free descriptor of one global property that is computable in linear time and requires no training or node alignment. The learning-based and mining-based approaches are substantially more expressive and target prediction or pattern discovery, whereas our measure targets rapid, interpretable, and comparable characterization at dataset scale. We therefore see these directions as complementary to ours: the descriptor could, for example, serve as a lightweight pre-analysis or feature that informs when the richer machinery of higher-order temporal learning is warranted.

Finally, we believe it is worth pointing out some limitations. First, the hyperedge-size distribution is a deliberately marginal statistic: it retains only the distribution of interaction orders at each time step and therefore discards several distinct kinds of structural information: the identities of the nodes participating in each hyperedge, the community structure among them, the overlap topology and recurrence of hyperedges across time, higher-order dependency patterns among interactions, and the broader connectivity organization of the system. As a direct consequence of this design, the measure is invariant to any structural change that leaves the hyperedge-size distribution unchanged: two temporal hypergraphs with entirely different node-level organization receive identical entropy values whenever their size distributions coincide. The framework should thus be read as characterizing one well-defined dimension of higher-order structure rather than the full structural complexity of a temporal hypergraph. We make this representational scope explicit as a design choice:

it is precisely this marginalization that makes the measure dataset-agnostic, free of node alignment, and computable in linear time. We also note that the relationship between our measure and the discarded connectivity dimension is not merely asserted but examined empirically in Section 4.3.1, where we compare our profile against pairwise-graph descriptors and find that it coincides with them on datasets where size and connectivity co-evolve and diverges from them where they do not. Richer encodings, such as hypergraph motif types or node-membership patterns [43], could capture more structural detail at the cost of higher computational complexity and reduced interpretability. Moreover, the change-point sensitivity parameter α is set per dataset rather than globally, reflecting differences in intrinsic volatility across domains, while each choice is motivated by the statistical properties of the corresponding profile, a principled data-driven method for selecting α could be useful in various scenarios. Furthermore, the entropy rate as defined in Definition 4 is a time average, which implicitly assumes that the underlying process is at least approximately stationary over the observation window. This could conflate heterogeneous regimes of the global entropy rate into a single summary statistic. However, the sliding-window profile addresses this by providing local estimates, making the non-stationarity visible rather than hiding it. Finally, note that $\mathfrak{h}(\boldsymbol{\mathcal{H}})$ is built from the marginal hyperedge-size distribution of each snapshot; therefore, it quantifies the average diversity of interaction orders rather than the conditional unpredictability of one snapshot given its predecessors. As we pointed out in Section 3.2, the two coincide under snapshot independence and otherwise differ by the average mutual information between successive snapshots. Extending the framework to explicitly model inter-snapshot dependence, thereby capturing the process entropy rate in the strict sense, is a natural and worthwhile direction for future work.

6. Conclusions

In this paper, we have introduced an information–theoretic framework for measuring how the diversity of interaction orders in a temporal hypergraph evolves over time. The framework is built around the temporal hypergraph entropy rate, defined as the time average of per-snapshot Shannon entropies computed from the hyperedge-size distribution. We proved that the measure collapses to zero under clique expansion, establishing that it captures interaction-order information that the standard size-blind pairwise projection discards entirely. We provided a bias-corrected sliding-window estimator that makes the framework applicable to finite real-world datasets, equipped it with a lightweight change-point detector, and validated the whole pipeline on six publicly available benchmark datasets spanning different domains. Indeed, our framework is built from standard components deliberately combined for simplicity, scalability, and cross-dataset compatibility. The experiments show that the entropy rate spans 1.60 bits across the six datasets with all values distinct, discriminates between structurally different collaboration cultures within the same domain, and detects unsupervised structural transitions whose locations are qualitatively consistent with known external events. The permutation null model analysis confirms that the observed temporal profiles are highly unlikely under a random arrangement of the same snapshots, providing evidence that the framework measures genuine temporal structure rather than merely marginal properties of the per-snapshot entropy distribution.

Several directions naturally extend the present work. The most immediate one concerns the richness of the primitive statistic: replacing the hyperedge-size distribution with a joint distribution over size and node-degree pairs, or with a distribution over hypergraph motif types [43], would capture structural detail that a marginal size summary cannot express, at the cost of higher computational complexity and reduced dataset agnosticism. On the applied side, extending the framework to temporal hypergraphs with dynamic node sets requires normalizing the size distribution by the active node set at each step, introduc-

ing non-trivial definitional choices that we leave for future work. Finally, two connections to existing frameworks appear particularly promising. Correlating the global entropy rate with the local structural descriptors derived from temporal ego-hypergraphs [4] could reveal how macro-level complexity relates to the micro-level evolution of individual neighborhoods. Connecting snapshot entropy to partial information decomposition [21] could clarify the roles of synergy and redundancy in driving the entropy rate, potentially paving the way for a more principled decomposition of higher-order structural complexity.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and source code used in this study are available at https://github.com/finalfire/entropy-in-temporal-hypergraphs (accessed on 30 June 2026).

Acknowledgments: This work was supported by EU NextGenerationEU programme under the funding schemes PNRR-PE-AI FAIR (Future Artificial Intelligence Research). Furthermore, the author thanks Parus major for insightful discussions on temporal evolution of higher-order complex systems.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Statistical Significance

For completeness, Table A1 gives all fifteen pairwise differences between the datasets' global entropy rates, each with its moving-block bootstrap 95% confidence interval and two-sided p-value. A difference is statistically significant at the 5% level precisely when its confidence interval excludes zero; fourteen of the fifteen pairs meet this criterion, with the only exception being coauth-DBLP versus tags-math-sx.

Table A1. All 15 pairwise differences in global entropy rate between datasets, with moving-block bootstrap 95% confidence intervals and two-sided bootstrap p-values ($B = 10{,}000$). Diff is $\mathfrak{h}_A - \mathfrak{h}_B$.

Dataset A	Dataset B	Diff	95% CI	p
coauth-DBLP	email-Enron	0.842	[0.584, 1.103]	<0.001
coauth-DBLP	tags-math-sx	0.053	[−0.170, 0.272]	0.63
coauth-DBLP	threads-ask-ubuntu	0.455	[0.223, 0.677]	<0.001
coauth-DBLP	coauth-MAG-History	1.057	[0.818, 1.305]	<0.001
coauth-DBLP	congress-bills	−0.538	[−0.777, −0.299]	<0.001
email-Enron	tags-math-sx	−0.789	[−0.933, −0.654]	<0.001
email-Enron	threads-ask-ubuntu	−0.387	[−0.534, −0.248]	<0.001
email-Enron	coauth-MAG-History	0.214	[0.057, 0.366]	0.006
email-Enron	congress-bills	−1.381	[−1.547, −1.222]	<0.001
tags-math-sx	threads-ask-ubuntu	0.402	[0.344, 0.460]	<0.001
tags-math-sx	coauth-MAG-History	1.004	[0.935, 1.085]	<0.001
tags-math-sx	congress-bills	−0.592	[−0.669, −0.510]	<0.001
threads-ask-ubuntu	coauth-MAG-History	0.601	[0.524, 0.690]	<0.001
threads-ask-ubuntu	congress-bills	−0.994	[−1.094, −0.891]	<0.001
coauth-MAG-History	congress-bills	−1.595	[−1.708, −1.483]	<0.001

References

1. Battiston, F.; Cencetti, G.; Iacopini, I.; Latora, V.; Lucas, M.; Patania, A.; Young, J.G.; Petri, G. Networks beyond pairwise interactions: Structure and dynamics. *Phys. Rep.* **2020**, *874*, 1–92. [CrossRef]
2. Benson, A.R.; Gleich, D.F.; Leskovec, J. Higher-order organization of complex networks. *Science* **2016**, *353*, 163–166. [CrossRef] [PubMed]
3. Cencetti, G.; Battiston, F.; Lepri, B.; Karsai, M. Temporal properties of higher-order interactions in social networks. *Sci. Rep.* **2021**, *11*, 7028. [CrossRef] [PubMed]

4. Cauteruccio, F.; Citraro, S.; Failla, A.; Rossetti, G. Generalizing Hypergraph Ego-Networks and Their Temporal Stability. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 53–67.
5. Failla, A.; Citraro, S.; Rossetti, G.; Cauteruccio, F. Characterizing User Archetypes and Discussions on Social Hypernetworks. *Big Data Cogn. Comput.* **2025**, *9*, 236. [CrossRef]
6. Zhao, K.; Karsai, M.; Bianconi, G. Entropy of dynamical social networks. *PLoS ONE* **2011**, *6*, e28116. [CrossRef] [PubMed]
7. Pfitzner, R.; Scholtes, I.; Garas, A.; Tessone, C.J.; Schweitzer, F. Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *arXiv* **2012**, arXiv:1208.0588.
8. Das, K.; Samanta, S.; Pal, M. Study on centrality measures in social networks: A survey. *Soc. Netw. Anal. Min.* **2018**, *8*, 13. [CrossRef]
9. Chun, J.; Bu, F.; Kim, Y.; Miyauchi, A.; Bonchi, F.; Shin, K. A Survey on Centrality and Importance Measures in Hypergraphs: Categorization and Empirical Insights. *arXiv* **2025**, arXiv:2512.00107.
10. Bick, C.; Gross, E.; Harrington, H.A.; Schaub, M.T. What are higher-order networks? *SIAM Rev.* **2023**, *65*, 686–731. [CrossRef]
11. Solé, R.V.; Valverde, S. Information theory of complex networks: On evolution and architectural constraints. In *Complex Networks*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 189–207.
12. Braunstein, S.L.; Ghosh, S.; Severini, S. The Laplacian of a graph as a density matrix: A basic combinatorial approach to separability of mixed states. *Ann. Comb.* **2006**, *10*, 291–317. [CrossRef]
13. Simonyi, G. Graph entropy: A survey. *Comb. Optim.* **1995**, *20*, 399–441. [CrossRef]
14. Su, D.; Peng, H.; Pan, Y.; Li, A. A survey of structural entropy: Theory, methods, and applications. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2025), Montreal, QC, Canada, 16–22 August 2025; pp. 10660–10668.
15. Mou, J.; Wang, L.; Zhang, C.; Luo, W.; Tan, S.; Zhou, B.; Lu, X. Network hierarchy entropy for quantifying graph dissimilarity. *Commun. Phys.* **2026**, *9*, 83. [CrossRef]
16. Dehmer, M.; Mowshowitz, A. A history of graph entropy measures. *Inf. Sci.* **2011**, *181*, 57–78. [CrossRef]
17. Tang, D.; Du, W.; Shekhtman, L.; Wang, Y.; Havlin, S.; Cao, X.; Yan, G. Predictability of real temporal networks. *Natl. Sci. Rev.* **2020**, *7*, 929–937. [CrossRef] [PubMed]
18. Weng, T.; Zhang, J.; Small, M.; Zheng, R.; Hui, P. Memory and betweenness preference in temporal networks induced from time series. *Sci. Rep.* **2017**, *7*, 41951. [CrossRef] [PubMed]
19. Tang, J.; Musolesi, M.; Mascolo, C.; Latora, V.; Nicosia, V. Analysing information flows and key mediators through temporal centrality metrics. In Proceedings of the 3rd Workshop on Social Network Systems, Paris, France, 13 April 2010; pp. 1–6.
20. Angelidis, G.; Ioannidis, E.; Makris, G.; Antoniou, I.; Varsakelis, N. Competitive conditions in global value chain networks: An assessment using entropy and network analysis. *Entropy* **2020**, *22*, 1068. [CrossRef] [PubMed]
21. Rosas, F.E.; Mediano, P.A.; Gastpar, M.; Jensen, H.J. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys. Rev. E* **2019**, *100*, 032305. [CrossRef] [PubMed]
22. Bretto, A. *Hypergraph Theory: An Introduction*; Springer: Cham, Switzerland, 2013.
23. Agarwal, S.; Branson, K.; Belongie, S. Higher order learning with graphs. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 17–24.
24. Zhou, D.; Huang, J.; Schölkopf, B. Learning with hypergraphs: Clustering, classification, and embedding. *Adv. Neural Inf. Process. Syst.* **2006**, *19*, 1–8.
25. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
26. Cover, T.M. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
27. Paninski, L. Estimation of entropy and mutual information. *Neural Comput.* **2003**, *15*, 1191–1253. [CrossRef]
28. Miller, G. Note on the bias of information estimates. *Inf. Theory Psychol. Probl. Methods* **1955**.
29. Killick, R.; Fearnhead, P.; Eckley, I.A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **2012**, *107*, 1590–1598. [CrossRef]
30. Benson, A.R.; Abebe, R.; Schaub, M.T.; Jadbabaie, A.; Kleinberg, J. Simplicial closure and higher-order link prediction. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E11221–E11230. [CrossRef] [PubMed]
31. Sinha, A.; Shen, Z.; Song, Y.; Ma, H.; Eide, D.; Hsu, B.J.P.; Wang, K. An Overview of Microsoft Academic Service (MAS) and Applications. In *Proceedings of the 24th International Conference on World Wide Web*; ACM Press: New York, NY, USA, 2015.
32. Fowler, J.H. Connecting the Congress: A Study of Cosponsorship Networks. *Political Anal.* **2006**, *14*, 456–487. [CrossRef]
33. Fowler, J.H. Legislative cosponsorship networks in the US House and Senate. *Soc. Netw.* **2006**, *28*, 454–465. [CrossRef]
34. Lahiri, S.N. *Resampling Methods for Dependent Data*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
35. Hastie, T. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.
36. Page, E.S. Continuous Inspection Schemes. *Biometrika* **1954**, *41*, 100–115. [CrossRef]

37. Scargle, J.D.; Norris, J.P.; Jackson, B.; Chiang, J. Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations. *Astrophys. J.* **2013**, *764*, 167. [CrossRef]
38. Krichevsky, R.; Trofimov, V. The performance of universal encoding. *IEEE Trans. Inf. Theory* **1981**, *27*, 199–207. [CrossRef]
39. Chao, A.; Shen, T.J. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ. Ecol. Stat.* **2003**, *10*, 429–443. [CrossRef]
40. Tibshirani, R.J.; Efron, B. *An Introduction to the Bootstrap*; Monographs on Statistics and Applied Probability 57; CRC: Boca Raton, FL, USA, 1993; pp. 1–436.
41. Liu, H.; Jiao, P.; Gao, M.; Chen, C.; Jin, D. Heterogeneous Temporal Hypergraph Neural Network. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2025), Montreal, QC, Canada, 16–22 August 2025.
42. Tian, H.; Zafarani, R. Higher-Order Networks Representation and Learning: A Survey. *ACM SIGKDD Explor. Newsl.* **2024**, *26*, 1–18. [CrossRef]
43. Lee, G.; Bu, F.; Eliassi-Rad, T.; Shin, K. A survey on hypergraph mining: Patterns, tools, and generators. *ACM Comput. Surv.* **2025**, *57*, 1–36. [CrossRef]
44. Meng, X.; Zhai, X.; Fei, G.; Wen, S.; Hu, G. Census and Analysis of Higher-Order Interactions in Real-World Hypergraphs. *Big Data Min. Anal.* **2025**, *8*, 383–406. [CrossRef]