Relationships Between Node Degrees and Hyperedge Sizes in Empirical Hypergraphs

Bogumił Kamiński[1]*, Paweł Prałat[2]†, Aleksander Wojnarowicz[1]‡, Mateusz Zawisza[1]§

[1]Decision Analysis and Support Unit, SGH Warsaw School of Economics, Warsaw, Poland
[2]Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada

Abstract

We investigate networks represented as hypergraphs and propose a a novel measure that captures the relationship between their node degrees and hyperedge sizes. We test the presence of such an association in 36 empirical hypergraphs from diverse domains, with a focus on social networks. Using nested model comparisons, we classify each such relationship as linear, monotonic, non-monotonic, or absent. Results reveal that true absence of this relationship is rare, while nearly half exhibit non-monotonic patterns. We evaluate three correlation measures of this association and find that Pearson correlation best aligns with relationship direction. We also consider three ways to capture this relationship (called: bipartite, node-centric or edge-centric) and show that the bipartite one yields most consistent results. We discuss the implications of existence of relationship between node degrees and hyperedge sizes for dynamic processes on social systems.

Keywords: empirical hypergraphs, node degree, hyperedge size, bipartite representation, Pearson correlation, non-monotonic relationship

1 Introduction

In recent years, hypergraphs have emerged as a powerful generalization of traditional pairwise graphs [3, 38, 95], particularly suited for modelling complex systems involving higher-order relationships [19, 88, 94, 123]. Unlike standard dyadic networks where edges connect pairs of nodes, hypergraphs allow hyperedges to connect any number of nodes, enabling a more expressive modelling framework [17, 20, 26]. This makes hypergraphs ideal for capturing group interactions found in diverse social networks, such as co-authorship networks [105, 89], affiliation or membership networks [36, 86, 128], social media [85, 7], social tagging [18], team sports [52, 103], but also ecological systems [55] and joint protein interactions in biological networks [91].

Due to their growing importance, many structural and statistical properties of empirical hypergraphs have been the focus of recent research [35, 78, 15]. Some of these measures, like node degree distribution [30], modularity [65, 64, 66, 31], clustering coefficients [40, 2] or the Bonacich eigenvector centrality [24] have analogues in traditional graphs [3, 92, 129, 22, 23, 41, 99, 98]. Others, such as hyperedge size distribution, node-hyperedge size correlation or the simplicial closure [105, 18] are unique to hypergraphs and open new questions for exploratory data analysis [1, 48]. Understanding these properties is essential for both descriptive purposes and for informing the design of algorithms and models tailored to higher-order data.

Descriptive analysis of hypergraph properties is not only of theoretical interest but also of practical significance. Structural characteristics—such as node degree distributions and hyperedge sizes—play a crucial role in shaping the dynamics of processes occurring on these networks [21]. For example, the spread of information, opinions, or infectious diseases can behave qualitatively differently depending on whether the system is modelled as a traditional graph or a hypergraph [77, 120, 139, 75]. Consequently, understanding and quantifying key

*Email: bkamins@sgh.waw.pl
†Email: pralat@torontomu.ca
‡Email: alwojnarowicz@gmail.com
§Corresponding author: Mateusz Zawisza, SGH Warsaw School of Economics, Institute of Econometrics, ul. Madalińskiego 6/8, 02-513 Warsaw, Poland.
Tel.: +48 502 190 746.
Email: mzawisz@sgh.waw.pl

structural features of hypergraphs is essential for developing accurate and predictive models of complex social systems.

In this paper, we propose and investigate a novel relationship in hypergraphs: between node degrees and their hyperedge sizes. While this relationship is non-existing in standard graphs, where all edges connect exactly two nodes, hypergraphs allow for nontrivial hyperedge size variability. This enables the study of correlations between how many hyperedges a node participates in (its degree) and how large those hyperedges tend to be. This important relationship has received limited attention in the literature up to our knowledge. Several generative and random models for hypergraphs explicitly analyze the distributions of node degree (number of hyperedges a node participates in) and hyperedge size (number of nodes in a hyperedge), often deriving these distributions in terms of model parameters. These models show that the mechanisms governing hyperedge formation, such as preferential attachment or copying, directly influence both node degree and hyperedge size distributions, and their interplay can lead to power-law or other heavy-tailed behaviours in real-world hypergraphs [16, 57, 109].

Beyond the mere existence of correlation between node degree and hyperedge size in empirical hypergraphs, we aim to explore whether this relationship is consistently non-zero and varies across semantic categories of hypergraphs. Such variation could help distinguish between types of real-world systems. For example, in co-authorship networks where hyperedges represent papers and nodes are authors, a positive correlation is expected: prolific researchers tend to publish more papers and often do so in larger teams [105]. This pattern contrasts with other domains, such as user tagging systems, where different interaction dynamics may apply.

Detecting systematic patterns in this relationship could inform the development of generative models of hypergraphs [30, 31]. Notably, many current generative frameworks such as h–ABCD [67, 69] implicitly assume zero correlation between node degree and hyperedge size. Incorporating flexible control over this correlation could improve the realism and utility of synthetic hypergraph models in both simulation and inference settings.

A core motivation for this study stems from the potential influence that the correlation between node degree and hyperedge size can exert on social dynamics unfolding on hypergraphs. In classical network science, it is well established that structural features such as degree distribution and assortativity shape fundamental dynamical processes, including epidemic spreading [104, 97], diffusion [71, 29], and the evolution of cooperation [107, 117]. By analogy, we hypothesize that in higher-order systems represented as hypergraphs, the relationship between how many groups individuals belong to (node degree) and the sizes of those groups (hyperedge size) may play a similarly pivotal role. As reviewed in Section 4, such correlations can modulate contagion speed [21], alter the size of the seed set needed for global influence in social diffusion models [9, 10] and affect cooperation levels in multiplayer public goods games [51, 4]. Identifying and characterizing these correlations in empirical hypergraphs is therefore not only a matter of structural interest, but a critical step toward understanding and ultimately modelling complex social dynamics in real-world higher-order systems.

To measure the relationship between node degree and hyperedge size, we propose a set of alternative measures to quantify the relationship between node degrees and hyperedge sizes. These include classical correlation coefficients (Pearson, Spearman, Kendall) applied to three types of hypergraph representations (node-centric, edge-centric, bipartite). Using a curated collection of 36 empirical hypergraphs from diverse domains, we assess the consistency and informativeness of these measures and identify general trends.

The remainder of the article is organized as follows. In Section 2, we introduce the foundational concepts, including hypergraph notation and key properties. Section 2.1 presents three data preprocessing strategies designed to enable meaningful comparisons between node degree and hyperedge size. Section 2.2 introduces the Eta-squared (η^2) statistic for evaluating alignment with semantic groupings. In Section 2.3, we outline a nested statistical testing procedure to classify the relationship between structural quantities. The results are presented in three parts. Section 3.1 evaluates and justifies the choice of the optimal data preprocessing strategy. Section 3.2 investigates which correlation metric best captures global structural trends. Section 3.3 applies model-based tests to classify empirical hypergraphs by the type of relationship between node degree and hyperedge size. In Section 4, we explore the potential impact of the identified relationship between node degree and hyperedge size on social dynamics, focusing on three selected problems: social contagion and spreading, social influence maximization, and cooperation in public goods games. Finally, conclusions and directions for future research are offered in Section 5. The Appendix provides detailed background on the empirical hypergraph datasets used in this study, including semantic segments, node and hyperedge interpretations, and descriptive statistics (Subsection A.1). It also documents implementation details and computational complexity (Subsection A.2), defines the correlation measures employed (Subsection A.3), and includes supplementary analyses referenced in the main text (Subsection A.4).

2 Notation, Methods & Data

In this section, we introduce the fundamental definitions, methods, and datasets that underpin our analysis of structural patterns in empirical hypergraphs. We begin by formalizing hypergraph notation and its incidence graph representation, which serves as the mathematical foundation for our computations (Section 2.1), together with the data preprocessing strategies designed to enable meaningful comparisons between node degree and hyperedge size. Subsequently, we present a principled method for evaluating the alignment between correlation values and semantic segments using the η^2 statistic (Section 2.2). We also describe a statistical procedure to classify the functional relationship between node degree and hyperedge size (Section 2.3). Finally, we give the overview of the empirical datasets used in this study and computational handling (Section 2.4), while the more detailed discussion of these topics is provided in Appendix (Sections A.1 and A.2) together with the review of correlation measures suitable for quantifying the statistical association (Section A.3).

2.1 Hypergraph Notation and Data Preprocessing Steps

Understanding the relationship between node degree and hyperedge size in a hypergraph requires precise definitions and careful data transformation. In this section, we formalize the notation used throughout the paper and describe three alternative preprocessing strategies that enable meaningful correlation and relationship analysis between these two structural quantities.

Hypergraph and Its Bipartite Representation as an Incidence Graph A hypergraph is a generalization of a graph in which edges, called hyperedges, can connect any number of vertices. Formally, a hypergraph is defined as $H = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ is the set of vertices, and $E = \{e_1, \ldots, e_m\}$ is the set of hyperedges, with each $e_j \subseteq V$ [68].

To facilitate analysis, we use the bipartite representation of a hypergraph, also known as its incidence graph. This representation is information-preserving and equivalent to the original hypergraph structure. Specifically, we define the incidence graph $IG(H) = (V', E')$, where the vertex set $V' = V \cup E$ includes both the original vertices and hyperedges of H, and edges E' connect $v \in V$ to $e \in E$ if and only if $v \in e$. This structure naturally yields an incidence matrix $B = (b_{ij}) \in \{0, 1\}^{n \times m}$, where each entry $b_{ij} = 1$ if vertex v_i belongs to hyperedge e_j, and $b_{ij} = 0$ otherwise. Both H and $IG(H)$ are fully recoverable from this matrix.

Hypergraph Data Preprocessing Steps Let the degree of a vertex $v_i \in V$ be defined as the number of hyperedges that include it, i.e.,

$$\text{degree}_i = \deg(v_i) = |\{e \in E : v_i \in e\}| .$$

Similarly, the size of a hyperedge $e_j \in E$ is given by

$$\text{hEdge}_j = |e_j| = |\{v \in V : v \in e_j\}| .$$

To quantify the relationship between node degrees and hyperedge sizes, these measures must be defined in the same domain. However, node degree is inherently a vertex-level property, while hyperedge size is defined at the hyperedge level. Thus, we construct data pre-processing transformations to bring both into a common domain for meaningful comparison.

To define a node-centric counterpart to hyperedge size, we compute for each node v_i the average size of hyperedges in which it participates:

$$\text{avgHEdgeSize}_i = \frac{\sum_{\{j : v_i \in e_j\}} |e_j|}{\deg(v_i)} = \frac{\sum_j b_{ij} \sum_k b_{kj}}{\sum_j b_{ij}}.$$

Algebraically, this corresponds to summing over the columns of the incidence matrix B for those hyperedges e_j that contain node v_i, and dividing by v_i's degree. This defines the node-centric preprocessing step, which produces a dataset of the form $\{(\text{degree}_i, \text{avgHEdgeSize}_i)\}_{i=1}^n$. One may then compute, for example, the Pearson correlation between node degree and average hyperedge size. This correlation reflects the expected hyperedge size for a randomly chosen node (uniformly at random) that has above-average degree.

Analogously, we define a hyperedge-centric counterpart to node degree by computing, for each hyperedge e_j, the average degree of its participating nodes:

$$\text{avgDegree}_j = \frac{\sum_{\{i:v_i \in e_j\}} \deg(v_i)}{|e_j|} = \frac{\sum_i b_{ij} \sum_k b_{ik}}{\sum_i b_{ij}}.$$

This amounts to summing over the rows of the incidence matrix B corresponding to nodes v_i in hyperedge e_j, and dividing by the hyperedge size. This defines the edge-centric[1] preprocessing step resulting in a dataset $\{(\text{avgDegree}_j, \text{hEdge}_j)\}_{j=1}^m$. The Pearson correlation between these quantities captures the expected degree of nodes involved in a hyperedge with size above the average.

A third approach is the bipartite representation preprocessing step, based on the incidence graph $IG(H)$ of the hypergraph H, where nodes and hyperedges form two disjoint sets of vertices. This representation leads to a dataset $\{(\text{degree}_i, \text{hEdge}_j) : b_{ij} = 1\}$ defined over the edges of the bipartite graph. Here, one can compute the assortativity coefficient as the Pearson correlation of the degrees of incident vertex pairs in the bipartite graph [96, 68]. This statistic reflects the expected hyperedge size for a randomly sampled node–hyperedge pair, conditional on the node having above-average degree.

2.2 Assessing the Alignment between Correlation and Hypergraph Segment

To evaluate how well different correlation designs align with semantic distinctions in hypergraph structure, we employ the Eta-squared statistic (η^2). This metric quantifies the proportion of variance in a continuous variable that is explained by a categorical predictor, and is commonly interpreted as a measure of effect size in analysis of variance (ANOVA) [112]. In our case, η^2 captures how well a given correlation coefficient (e.g., between hyperedge size and node degree) can be predicted based on the hypergraph segment to which it belongs.

Formally, let Y_i denote the correlation value computed for the i-th hypergraph, and let $\mathcal{S}_i$ be its segment label (e.g., *email*, *tag-question*, etc.). Then, η^2 is defined as:

$$\eta^2 = \frac{\text{SS}_{\text{between}}}{\text{SS}_{\text{total}}} = \frac{\sum_{g=1}^{G} n_g(\bar{Y}_g - \bar{Y})^2}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}, \tag{1}$$

where G is the number of groups (segments), n_g is the number of observations in group g, $\bar{Y}_g$ is the mean correlation in group g, and $\bar{Y}$ is the overall mean. The numerator is the between-group sum of squares ($\text{SS}_{\text{between}}$), and the denominator is the total sum of squares (SS_{total}).

We consider nine correlation configurations, defined by the combination of three data preprocessing strategies (edge-centric, node-centric, and bipartite representation) with three correlation measures (Pearson, Spearman, and Kendall). For each configuration, a single correlation value is computed for each of the 12 hypergraph segments under study, which include: *user-answer*, *physical contact*, *part-whole*, *diseases and genes*, *email*, *person-place*, *political*, *participant-conference*, *user-review*, *drugs*, *tag-question*, and *user-thread*.

To compute η^2, we fit a linear model where the response variable is the correlation value and the predictor is a one-hot encoded vector representing the hypergraph segment. The resulting η^2 is equivalent to the R^2 of this model and indicates the proportion of variance in the correlation values that can be attributed to segment identity. All η^2 values reported in this paper were calculated in `R` using the `etaSquared()` function from the `lsr` package [90].

A higher η^2 signifies a stronger alignment between correlations and hypergraph segments. Thus, we use η^2 as a criterion for selecting the optimal design of data preprocessing method and correlation measure that yields the most segment-sensitive correlation values.

2.3 Statistical Identification of Relationship Type

This subsection outlines the statistical procedure used to classify the relationship between hyperedge size and node degree into one of four categories: non-monotonic, monotonic (but not linear), linear, or no relationship. This classification is later employed in the results Subsection 3.3. The identification procedure is model-based and involves fitting three types of models to each dataset: (i) an unrestricted Generalised Additive Model

[1] A more precise name would be *hyperedge-centric preprocessing step*, but for conciseness, we refer to it simply as *edge-centric* throughout the text.

(GAM) [56, 132], (ii) a monotonic GAM, and (iii) a simple linear regression model estimated using ordinary least squares (OLS) [42].

In this work, we define a monotonic GAM as a special case of a shape-constrained additive model (SCAM) in which the smooth term is restricted to be either monotonically increasing or decreasing. To implement this, we fit two SCAMs with a monotonic increasing constraint and another with a decreasing constraint. Then our procedure selects the model with the lower residual sum of squares as the representative monotonic GAM. This approach follows the methodology introduced by [110] and is implemented using the `mgcv` and `scam` packages in R [132]. For general background on generalized additive models, we refer readers to [56].

The classification is conducted via a sequence of statistical comparisons using ANOVA tests for nested models [44]. As illustrated in Figure 1, the procedure begins by comparing the unrestricted GAM to the best-fitting monotonic GAM using an F-test. The null hypothesis (H_0) states that the monotonic GAM sufficiently explains the data; rejecting it (at $\alpha < 10^{-5}$) implies the presence of significant non-monotonicity, and the relationship is classified as non-monotonic. The choice of such a stringent significance threshold is motivated by the large size of most empirical hypergraphs and the need to limit false discovery in favour of simpler models unless the data strongly supports added complexity.

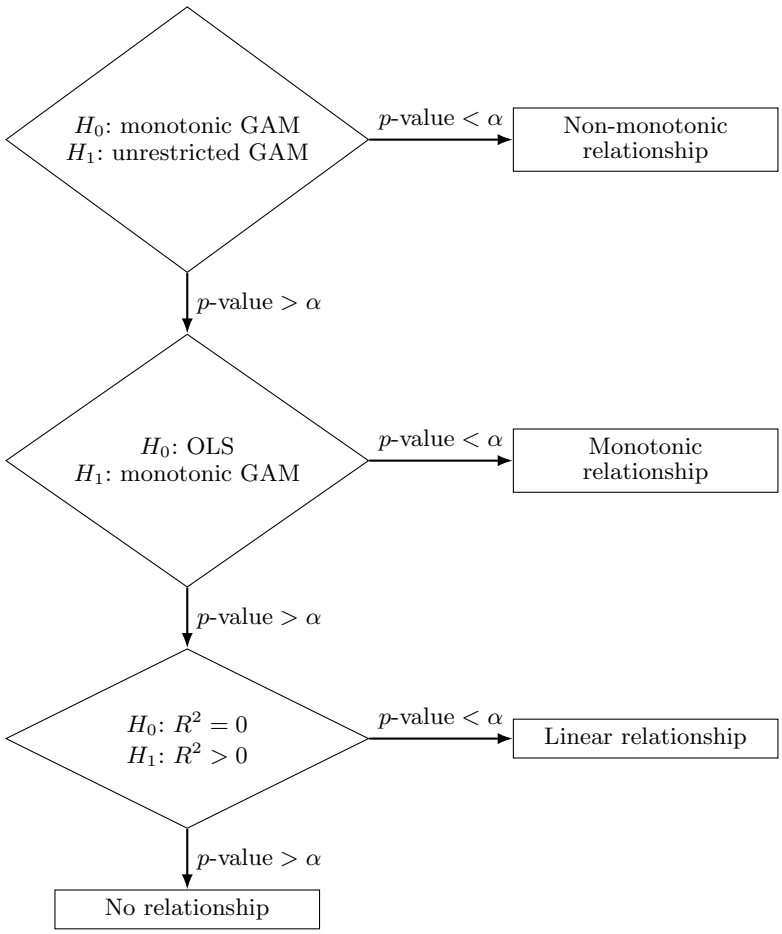


Figure 1: Diagram illustrating the decision procedure for classifying the relationship between two variables (hyperedge size and node degree) into one of four categories: non-monotonic, monotonic (but not linear), linear, or no relationship. Decisions are based on successive ANOVA comparisons of nested models: unrestricted GAM, monotonic GAM, and OLS.

If the null is not rejected in the first test, the relationship is assumed monotonic, and a second ANOVA test compares the monotonic GAM to a linear OLS model. Rejection of linearity indicates a monotonic (but non-linear) pattern. If the test fails to reject the linear model, a final F-test is performed to check whether the linear model explains any significant variation in the data (i.e., whether $R^2 > 0$). If this is not the case, the relationship is classified as "no relationship." Thus, the decision tree ensures a structured, conservative, and

data-driven assignment of relationship type.

2.4 Data and Computation Overview

Our analysis draws upon a curated collection of 36 empirical hypergraphs spanning diverse domains—including physical contact networks, online user interactions, political affiliations, and biomedical associations. Each dataset was selected for its interpretability and structural diversity, with nodes and hyperedges carrying clear semantic meanings relevant to their source context (e.g., people, genes, products). This breadth ensures that observed statistical relationships are robust across domains and not artifacts of a single data type. Detailed descriptions of all datasets, their assigned segments, and statistical summaries of node degrees and hyperedge sizes are presented in Appendix A.1.

To process and analyze these datasets, we developed a flexible computational pipeline tailored to varied file formats and scales. Hypergraphs were represented using sparse matrix structures to ensure memory efficiency and speed. Different sparse formats were tested and selected based on input type and construction cost, with final conversion to CSR (Compressed Sparse Row) format to enable fast indexing and vectorized computations. Our implementation makes use of Python libraries such as `scipy`, `numpy`, and, where applicable, `numba` for optimization. All statistical analyses, including correlation computations, model fitting, and figure generation, were performed using the R programming language [111]. The full technical details of data ingestion, matrix construction, and performance considerations are documented in Appendix A.2.

All reproducible code used in this study, including both Python and R scripts for data processing, statistical analysis, and figure generation, is available in the public repository: `https://github.com/AleksanderWWW/hypergraph-properties`.

3 Results

This section is divided into four main parts. In the first subsection, Subsection 3.1, we justify the optimal choice of hypergraph data preprocessing techniques introduced in Subsection 2.1, namely, the edge-centric, node-centric, and bipartite representations. In the second subsection, Subsection 3.2, we delve into the selection of one of three correlation measures: Pearson, Spearman, or Kendall. These two parts provide recommendations for choosing a specific design when a single correlation coefficient is needed to summarize the relationship between hyperedge size and node degree.

However, a single coefficient rarely captures the full complexity of these relationships. Therefore, in the third subsection, Subsection 3.3, we move beyond global coefficients to characterize the nature of the observed relationships in more qualitative manner. We classify them into one of four categories: non-monotonic, monotonic (but non-linear), linear, or no apparent relationship. We then assess the distribution of these categories and discuss the kinds of patterns typically observed. Additionally, we perform robustness checks to ensure the reliability of our findings.

3.1 Optimal Choice of Hypergraph Preprocessing Step

In this subsection, we assess the effectiveness of various hypergraph data preprocessing strategies for quantifying the relationship between hyperedge size and node degree. Specifically, we compare three hypergraph representation methods: edge-centric, node-centric, and bipartite representation. Our aim is to identify the preprocessing method that yields a single numerical descriptor most strongly aligned with the segmentation of the hypergraph. This alignment is quantified using the Eta-squared statistic (η^2), which captures the proportion of variance explained. The findings presented here offer practical guidance for selecting an appropriate setup when a single numerical summary of the hyperedge size–degree relationship is required.

The goal of this subsection is to determine which of the three hypergraph data preprocessing strategies introduced in Subsection 2.1 is most suitable for capturing the relationship between hyperedge size and node degree. The criterion for this comparison is the degree of alignment between the computed correlation values and the categorical segmentation of hypergraphs. This alignment is quantified using the Eta-squared statistic (η^2), as described in subsection 2.2.

	Pearson	Spearman	Kendall
node-centric	0.5380	0.6062	0.4360
edge-centric	0.6096	0.4354	0.4360
bipartite representation	0.6656	0.6782	0.6605

Table 1: Eta-squared (η^2) values measuring the proportion of variance in each continuous variable explained by the nominal `Category` variable. Rows correspond to type of hypergraph data pre-processing (`node-centric`, `edge-centric`, `bipartite representation`), and columns indicate the correlation type used (Pearson, Spearman, Kendall).

3.1.1 Comparing Preprocessing Strategies via η^2

To compute the correlation between hyperedge size and node degree, two design choices must be made: the hypergraph preprocessing method and the type of correlation coefficient. The latter includes Pearson, Spearman, and Kendall coefficients, each described in Subsection A.3. These choices yield a total of nine (3×3) possible combinations. For each such combination, a correlation value is computed for every hypergraph in the dataset. These correlation values are then evaluated for their segment-level alignment using η^2. The resulting η^2 scores, summarized in Table 1, serve as the basis for identifying the most informative preprocessing approach.

Table 1 presents the η^2 values for each combination of hypergraph data preprocessing method (rows) and correlation measure (columns). Recall that $0 \leqslant \eta^2 \leqslant 1$ quantifies the proportion of variance in the correlation coefficients that can be explained by the categorical variable representing hypergraph segment identity. Higher values of η^2 indicate stronger alignment between the correlation values and the semantic distinctions between hypergraph types. While the absolute differences in η^2 are moderate, these differences are consistent and informative enough to guide design choices.

The most striking observation is that the choice of data preprocessing method exerts the greatest influence on η^2. Across all three correlation measures, the bipartite representation consistently yields the highest η^2 scores, clearly outperforming both the node-centric and edge-centric approaches. Specifically, bipartite representation achieves $\eta^2 = 0.6656$ with Pearson, 0.6782 with Spearman, and 0.6605 with Kendall, all substantially higher than their respective scores under alternative preprocessing methods. This robustness suggests that the bipartite structure more faithfully preserves segment-level variability relevant to the correlation between hyperedge size and node degree.

In contrast, the choice of correlation measure appears to matter less, especially within the bipartite setting, where all three correlations perform comparably. This implies that once an appropriate data structure is chosen, the precise choice of correlation coefficient has limited impact on segment-level discriminatory power. For the detailed comparative analyses of preprocessing strategies, we therefore rely on Pearson correlation as a representative metric. Pearson not only performs nearly as well as Spearman in the bipartite case (0.6656 vs. 0.6782), but is the best-performing measure for the edge-centric view and second-best in the node-centric case. This makes it a stable and informative benchmark for deeper investigation of structural differences among preprocessing strategies.

Summary This analysis demonstrates that the choice of preprocessing strategy plays a more decisive role than the choice of correlation coefficient in capturing segment-level differences between hypergraphs. Among all tested combinations, the bipartite representation consistently yields the highest η^2 values across Pearson, Spearman, and Kendall correlations, indicating that it best preserves meaningful structural variability across semantic categories. Consequently, we recommend bipartite preprocessing as the default strategy.

3.1.2 Explaining η^2 Values via Within-Segment Variability in Correlation Estimates

To better understand the η^2 values reported in Table 1, we visualize the distribution of correlation coefficients within 12 hypergraph segments in Figure 6. The figure displays the variability of correlation values for six selected combinations of data preprocessing method and correlation coefficient, i.e., all pairings of Pearson and Spearman correlations with the three preprocessing strategies: node-centric, edge-centric, and bipartite representation. We omit Kendall's τ for clarity, focusing on the two more commonly used and better-performing measures.

The figure illustrates how well each combination discriminates among the 12 hypergraph segments. Lower

variability in correlation values across hypergraphs within the same segment implies stronger between-group effects, resulting in higher η^2. While this variability is visualised using the estimated interquartile range (IQR), the formal η^2 is computed using within-group sums of squares. A pattern emerges: the bipartite representation consistently shows low within-segment variability across all 12 categories, regardless of whether Pearson or Spearman is used. This uniformity explains its dominant performance in Table 1, where both correlations achieve the highest η^2 values (0.6656 and 0.6782, respectively).

In contrast, node-centric and edge-centric approaches show more fluctuation across segments. For instance, within the `user-thread` segment, the Pearson correlation under the node-centric view shows very low variability. However, the same configuration yields high variability for other segments, such as `physical contact`, resulting in a lower overall η^2 of 0.5380. This value is not only lower than the Spearman correlation for the same preprocessing ($\eta^2 = 0.6062$), but also substantially below the Pearson result for the bipartite representation ($\eta^2 = 0.6656$). These observations support the conclusion that bipartite preprocessing is both the most consistent and most informative representation for capturing segment-level differences in the relationship between hyperedge size and node degree.

A closer comparison of correlation variability across the 12 hypergraph segments highlights consistent advantages of bipartite representation. For segments such as `diseases and genes`, `drugs`, `email`, `part-whole`, and `person-place`, both Pearson and Spearman correlations under the bipartite representation show minimal within-segment variability, making them clearly superior to other combinations. In contrast, Spearman correlations under both node-centric and edge-centric preprocessing tend to exhibit the highest variability in these segments, making them the least effective. For the `participant-conference` segment, all six combinations perform well, showing tight clustering of correlation values and indicating that this segment is structurally well captured regardless of design choice. A similar trend holds for `user-answer` and `tag-question`, though in the latter, Spearman under node-centric preprocessing shows noticeably more dispersion. In the `physical contact` segment, both bipartite and edge-centric preprocessing yield relatively consistent correlation values, whereas node-centric shows a much wider spread. Lastly, in the `political` segment, the bipartite representation again performs best, with both Pearson and Spearman yielding compact distributions; node-centric performs moderately well, while edge-centric displays considerably higher internal variability. These patterns reinforce the advantage of bipartite preprocessing for producing stable, segment-discriminative correlation estimates across a diverse range of hypergraph types.

Another insight from Figure 6 concerns the variation of correlation values across different correlation measures, namely Pearson and Spearman and preprocessing steps, examined within each of the 12 hypergraph segments. It is rare to find a segment where all six combinations (2 correlation types $\times$ 3 preprocessing methods) yield similar values. The closest example is the `user-answer` segment, which exhibits consistently mild negative correlations around -0.11 across four of the six configurations. Exceptions include the edge-centric Spearman correlation ($r = -0.010$) and the node-centric Spearman correlation ($r = 0.062$), which deviate from this pattern.

Another example of such a consistency is found in the `participant-conference` segment, where five of the six measures fall within a narrow interval of $(-0.0251,\ 0.0382)$, indicating near-zero correlation. The outlier is again node-centric Spearman, which produces a noticeably higher value of $r = 0.306$. Although node-centric Spearman differs substantially in average correlation from the other five measures for several segments (e.g., `Drugs`, `email`, `participant-conference`, `person-place`, `physical contact`), it still achieves relatively low within-segment variance, resulting in a high η^2 value of 0.6062. This is one of the two highest η^2 (another is edge-centric Pearson) scores outside the bipartite representation.

For most other segments, the variation across measures is even greater, with discrepancies driven more by differences in preprocessing strategy than by choice of correlation method. Notably, in the bipartite representation, both Pearson and Spearman produce remarkably similar correlation values across all segments. This consistency is reflected in the nearly identical η^2 scores for these two measures reported in Table 1, and will be further explored in detail in Subsection 3.2.

Summary This subsection explains the high η^2 values associated with the bipartite preprocessing strategy by examining within-segment variability in correlation coefficients. Visualizations across 12 semantic hypergraph segments reveal that bipartite representation yields consistently low variance in both Pearson and Spearman correlations, leading to clearer segment-level differentiation. In contrast, node-centric and edge-centric strategies exhibit higher and more inconsistent variability, especially across structurally diverse segments. These findings confirm that the superior η^2 scores of bipartite preprocessing stem from its ability to produce stable, segment-

informative correlation estimates across multiple hypergraph types.

3.1.3 Pairwise Comparison of Preprocessing Strategies

A deeper analysis of the interrelation between the three data preprocessing strategies is provided in Figure 2. Like Figure 7, it reports only Pearson correlation values, but this time in pairwise comparisons between preprocessing types. Subfigure (a) displays a scatterplot comparing Pearson correlations under the bipartite representation versus those under edge-centric processing. The reported R^2 of 0.60 indicates a moderate positive association between these two measurements. This is consistent with previous results in Figures 7 and 6, which demonstrated general alignment between these two correlation estimates, although with notable exceptions.

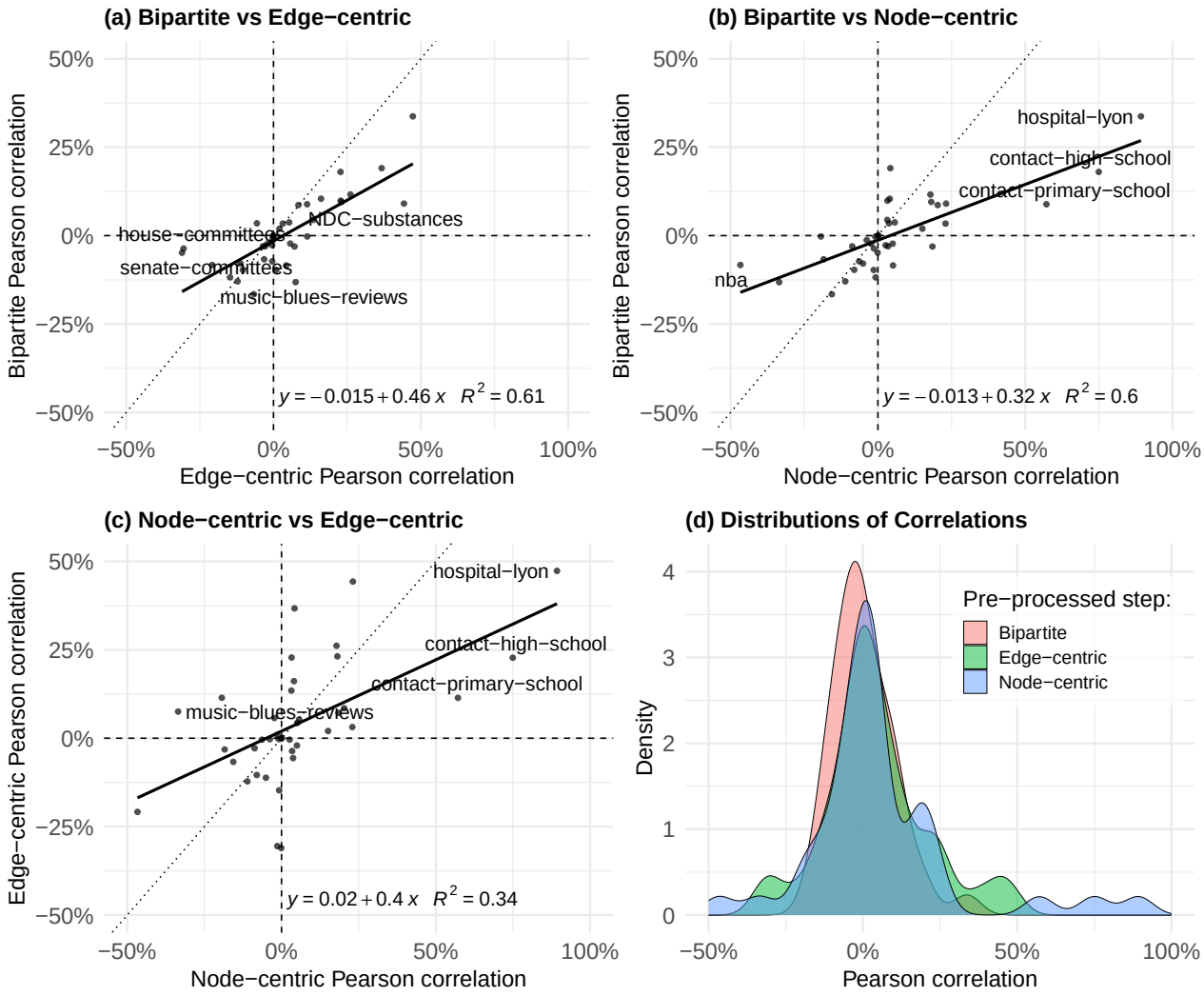


Figure 2: (a–c) Scatterplots of Pearson correlations between pairs of hypergraph preprocessing strategies, with fitted regression line (solid) and identity line (dashed); (d) distribution of Pearson correlations across all hypergraphs. Outlier hypergraphs with largest discrepancies are labelled.

Four hypergraphs with the largest discrepancies between bipartite and edge-centric Pearson correlations are labelled in the figure. Among them, `music-blues-reviews` stands out, with a positive edge-centric Pearson correlation of 0.0753, while its bipartite Pearson is substantially negative at −0.132. Interestingly, its node-centric Pearson is even more negative at −0.335, aligning more closely with the bipartite estimate. Similar discrepancies are observed in `house-committees` ($r_{\text{edge}} = -0.305$, $r_{\text{bipartite}} = -0.037$) and `senate-committees` ($r_{\text{edge}} = -0.310$, $r_{\text{bipartite}} = -0.048$), both from the `political` segment. Again, their node-centric correlations

(-0.014 and -0.001, respectively) are closer to bipartite values. The regression line fitted between edge-centric and bipartite Pearson correlations is $r_{\text{bipartite}} = 0.015 + 0.46 \times r_{\text{edge}}$, indicating no systematic bias (as the intercept is small and statistically non-significant).

Figure 2(b) shows the scatterplot of bipartite versus node-centric Pearson correlations. The relationship between these two is similar, with $R^2 = 0.60$ and an estimated regression line of $r_{\text{bipartite}} = 0.013 + 0.48 \times r_{\text{node}}$, again indicating no significant bias. Nonetheless, some hypergraphs exhibit large discrepancies. For example, `contact-high-school` has a node-centric Pearson of 0.750 but a bipartite value of only 0.180, a difference of 0.570. Similar discrepancies are found in `hospital-lyon` ($r_{\text{node}} = 0.893$, $r_{\text{bipartite}} = 0.337$), `contact-primary-school` (0.572 vs. 0.089), and `nba` (-0.467 vs. -0.083). As previously, the third preprocessing measurements, i.e., edge-centric Pearson correlations for these hypergraphs are closer to bipartite correlations.

Finally, Figure 2(c) compares Pearson correlations between node-centric and edge-centric representations. This pair exhibits the weakest relationship, with $R^2 = 0.34$. The estimated regression line is $r_{\text{node}} = 0.020 + 0.52 \times r_{\text{edge}}$, again with a non-significant intercept. The most prominent outliers are again hypergraphs from the `physical contact` segment: `contact-high-school` ($r_{\text{node}} = 0.750$, $r_{\text{edge}} = 0.228$), `contact-primary-school` (0.572 vs. 0.114), `hospital-lyon` (0.893 vs. 0.473), and `music-blues-reviews` (-0.335 vs. 0.075). These illustrate that node- and edge-centric preprocessing steps can yield substantially different correlation estimates even when applied to structurally similar hypergraphs.

Figure 2(d) displays overlaid distributions of Pearson correlations for the three hypergraph preprocessing strategies: edge-centric, node-centric, and bipartite representation. Consistent with the regression results discussed earlier, the distributions share similar central tendencies, indicating no systematic location bias across preprocessing types. However, the distributions differ notably in their spread. The node-centric correlations exhibit the widest dispersion, reflecting the presence of several hypergraphs with exceptionally especially high, but also low correlation values. In contrast, the bipartite-based Pearson correlations are the most concentrated around its mean, suggesting greater stability and less variability across datasets. Edge-centric correlations fall in between, showing moderate variability. These observations further reinforce the finding that the bipartite representation yields more stable and interpretable correlation estimates across diverse hypergraph structures.

Summary This subsection compares the three preprocessing strategies by examining pairwise relationships between their Pearson correlation estimates. The bipartite representation shows moderate but consistent agreement with both node- and edge-centric approaches ($R^2 \approx 0.60$), while node- and edge-centric correlations are less aligned ($R^2 = 0.34$), reflecting structural differences in how each method aggregates information. Several outlier hypergraphs exhibit large discrepancies, often with bipartite values closer to node-centric than edge-centric correlations. Distributional analysis confirms these trends: bipartite correlations are the most stable and tightly clustered, while node-centric correlations display the greatest spread. Overall, these results underscore the superior consistency of the bipartite strategy for capturing correlation patterns in empirical hypergraphs.

3.1.4 Interpretation Corner: Pearson Correlations by Segment

To further understand the behaviour and consistency of correlation estimates across different, we examine Pearson correlation values between hyperedge size and node degree for all hypergraphs, computed under three data preprocessing methods: node-centric, edge-centric, and bipartite representation in Figure 7. The hypergraphs are sorted by decreasing Pearson correlation under bipartite representation. This ordering allows us to visually identify clusters of hypergraphs that exhibit similar correlation structure, both in magnitude and in sign.

Several coherent segment-level patterns emerge. For instance, the `physical contact` segment, comprising `hospital-lyon` ($r = 0.337$), `contact-high-school` ($r = 0.180$), `contact-primary-school` ($r = 0.089$), `InVS13` ($r = -0.030$), `InVS15` ($r = 0.020$), `Malawi-village` ($r = 0.034$), and `Science-Gallery` ($r = 0.086$), appears mostly in the upper half of the ranking, exhibiting generally positive Pearson correlations under bipartite representation. Similarly, the `Drugs` segment: `NDC-classes` ($r = 0.191$) and `NDC-substances` ($r = 0.091$), and the `user-thread` group: `threads-ask-ubuntu` ($r = 0.104$), `threads-math-sx` ($r = 0.099$), `twitter` ($r = 0.035$), also cluster together with consistently positive, though more moderate, correlation values. Likewise, hypergraphs from the `tag-question` segment:`tags-math-sx` ($r = 0.116$), `tags-ask-ubuntu` ($r = 0.095$), exhibit moderate and comparable correlations.

Conversely, several hypergraphs appear at the lower end of the ranking with negative correlation values. These include `disgenenet` ($r = -0.166$) and `diseasome` ($r = -0.067$) from the `Diseases and genes` segment; `geometry` ($r = -0.129$) and `algebra` ($r = -0.0968$) from `user-answer`; and 3 hypergraphs from the

`user-review` segment, including `music-blues-reviews` ($r = -0.132$), `restaurant-reviews` ($r = -0.079$), and `vegas-bars-reviews` ($r = 0.037$), which show weak to moderately negative or near-zero correlations. These trends illustrate how semantic categories often align with shared correlation patterns, hinting at underlying structural regularities that will be explored further in the following analyses.

Summary This subsection highlights the interpretability of Pearson correlations between hyperedge size and node degree when grouped by semantic hypergraph segments. Using bipartite preprocessing as a reference, the results reveal clear and coherent segment-level trends. Segments such as `physical contact`, `user-thread`, `tag-question`, and `Drugs` exhibit consistently positive correlations, suggesting a shared structural tendency across these domains. Conversely, segments like `Diseases and genes`, `user-answer`, and `user-review` tend to show negative or near-zero correlations. These patterns underscore that the relationship between node degree and hyperedge size is not only statistically detectable but also semantically meaningful, aligning with domain-specific mechanisms such as contact dynamics, specialization, or institutional constraints. The consistency of signs across preprocessing methods further reinforces the robustness of these patterns.

3.1.5 Summary and Recommendations

This analysis demonstrates that the bipartite representation is the most effective preprocessing strategy for capturing the relationship between hyperedge size and node degree in hypergraph data. It consistently yields the highest Eta-squared (η^2) values across various correlation measures, indicating strong alignment with the semantic segmentation of hypergraphs and low within-segment variability. While the choice of correlation coefficient (Pearson, Spearman, or Kendall) has some influence, it is secondary to the choice of representation: in the bipartite framework, all three yield similar and reliable results. By contrast, node-centric correlations are highly variable and tend to overestimate strength, while edge-centric ones are somewhat more stable but less consistent than bipartite. These findings emphasize that selecting an appropriate preprocessing method, i.e., the bipartite representation, is more crucial than the specific correlation measure when summarizing the hyperedge size–degree relationship. However, a detailed comparative analysis of Pearson and Spearman correlations, accompanied by additional evaluation criteria beyond η^2, follows in the next subsection and is expected to shed new light on the importance of choosing an appropriate correlation metric.

3.2 Optimal Choice of Correlation Coefficient

The previous subsection demonstrated that the bipartite representation is the preferred preprocessing method for computing correlations between hyperedge size and node degree, outperforming both node-centric and edge-centric approaches. Once this design choice is fixed, the next natural question arises: which correlation coefficient, e.g., Pearson, Spearman, or Kendall, should be used as a single, interpretable summary of this relationship? This subsection aims to provide practical guidelines for selecting the most appropriate correlation measure. A qualitative discussion of the types of relationships underlying these correlations will follow in subsection 3.3.

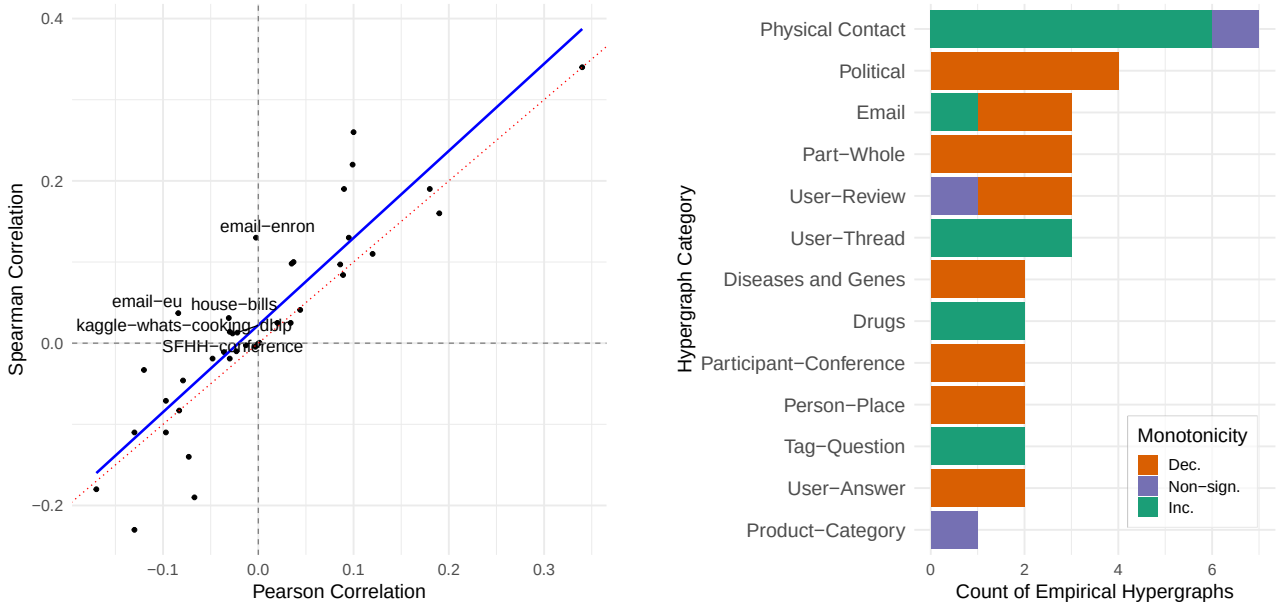
As shown earlier in Table 1, the η^2 scores, measuring how well each correlation coefficient aligns with known hypergraph categories, do not provide a decisive basis for choosing among Pearson, Spearman, and Kendall. In the bipartite setting, all three perform comparably: η^2 equals 0.6656 for Pearson, 0.6782 for Spearman, and 0.6605 for Kendall. These marginal differences are too small to support the selection of one coefficient over the others based solely on η^2.

To resolve this ambiguity, we proceed with a more detailed investigation focused on comparing Pearson and Spearman coefficients. (Since Spearman and Kendall are almost perfectly correlated in our dataset, with over 99% agreement, we focus only on Spearman as a representative of the nonparametric class.) Our analysis follows two complementary approaches: first, a quantitative comparison of the magnitude of Pearson and Spearman correlations' differences across empirical hypergraphs; and second, a qualitative comparison based on the alignment of their signs with the global trends identified by monotonic GAM models.

3.2.1 Quantitative Comparison of Pearson and Spearman correlations

A quantitative approach to compare Pearson and Spearman coefficients is to examine their alignment and investigate hypergraphs where the two diverge significantly. Noting that Spearman and Kendall are nearly identical with the Pearson correlation of 99% between them, therefore, we focus exclusively on Spearman for comparison against Pearson.

Figure 3a presents a scatterplot of Pearson (X-axis) vs. Spearman (Y-axis) correlation coefficients for 36 empirical hypergraphs. The dashed red line indicates the identity line (45°), while the solid blue line is the fitted linear regression. Six hypergraphs for which the two coefficients differ in sign are labelled: `email-enron`, `email-eu`, `kaggle-whats-cooking`, `dblp`, `SFHH-conference`, and `house-bills`. Visual inspection reveals a general alignment along the identity line, with minor fluctuations. The fitted regression line supports this, showing a slope close to one and an intercept estimate of 2.43% (SE = 0.988%), which is statistically significant at $\alpha = 0.05$ ($p = 0.0192$). This indicates a small but systematic upward bias in Spearman relative to Pearson.



(a) Scatterplot of Pearson vs. Spearman correlation coefficients for 36 hypergraphs. The dashed red line indicates the identity line (45°), and the solid blue line shows the fitted linear regression. The six hypergraphs for which the two correlation coefficients differ in sign are labeled.

(b) Frequency of monotonic trend classes by hypergraph category. Each horizontal bar shows the count of hypergraphs in that category with an increasing (Inc.), decreasing (Dec.), or non-significant (Non-sign.) monotonic GAM fit.

Figure 3: Comparison of correlation measures (left) and their monotonicity alignment with hypergraph categories (right).

The goodness-of-fit of this regression, $R^2 = 79.0\%$, suggests a medium-to-high alignment between the two metrics. Half of the hypergraphs show an absolute difference between the two metrics of less than 2.95pp, and 75% have a difference below 6.4pp. Nevertheless, a few outliers show discrepancies exceeding 12pp in absolute terms: `email-eu` (Spearman vs. Perason difference: 12.1pp), `threads-math-sx` (12.1pp), `diseasome` (-12.3pp), `email-enron` (13.2pp), and `threads-ask-ubuntu` (16pp). Notably, four of these belong to either the `email` or `threads` categories. In each of these four cases, Spearman is systematically higher than Pearson by at least 12pp; in two instances (`email-enron`, `email-eu`) this leads to opposite signs. The common factor appears to be an initial upward trend in dense regions of the data, which biases rank-based Spearman estimates, as discussed for `email-eu` in Figure 4a. While Pearson reflects the global trend, Spearman is heavily influenced by early dense data ranges.

The case of `diseasome` presents a reverse scenario. Here, Spearman is 12.3pp lower than Pearson. Figure 8 from Appendix reveals a complex structure: while the global trend is slightly negative (Pearson = -0.067), the end of the range features an upward fluctuation. Spearman emphasizes early data, which starts with a mild upward slope but transitions quickly to a descending pattern, better captured by the negative Spearman (-0.19). Importantly, Spearman is statistically significant at $\alpha = 0.00001$, whereas Pearson is only marginally significant at $\alpha = 0.05$ ($p = 0.026$). This is the one case in which Spearman is better aligned with the monotonic GAM classification than Pearson, especially considering the small dataset size ($N = 1109$).

Summary In summary, Pearson and Spearman show strong alignment overall, with $R^2 = 79\%$ and a modest 2.43% upward bias in Spearman. For 75% of hypergraphs, the absolute difference does not exceed 4.85pp. However, a closer inspection of the five most divergent hypergraphs shows that in four of them, Pearson is more robust and consistent with visual trends and fitted GAM models. Beyond quantitative differences, qualitative discrepancies in sign between Pearson and Spearman are especially consequential and motivate the subsequent part of paper.

3.2.2 Assessing Alignment Between GAM Monotonicity and Correlation Coefficients' Signs

In this part, we propose an additional criterion that is not captured by η^2: namely, the degree to which the sign of a correlation coefficient reflects the global trend in the data. This trend is approximated using monotonic Generalized Additive Models (GAMs), fitted separately under increasing and decreasing shape constraints. Although Pearson measures linear dependence and Spearman/Kendall assess monotonic relationships, we observe in following subsection 2.3 that approximately half of the empirical hypergraphs exhibit complex, non-monotonic relationships, often with a clear dominant trend. For such cases, it is desirable that the sign of the selected correlation coefficient be aligned with the direction of this dominant trend.

To operationalize this comparison, we classify each empirical hypergraph in two ways. First, based on the correlation coefficient (Pearson, Spearman, or Kendall), we assign it to one of three categories: (i) significantly negative, (ii) non-significant, or (iii) significantly positive. Second, based on the monotonic GAM fit, we classify it into: (i) decreasing trend, (ii) no significant trend, or (iii) increasing trend. The alignment between these two classifications is evaluated using contingency tables in Table 2. The top table concerns Pearson, while the bottom aggregates Spearman and Kendall, which yield identical classifications.

Sign of Pearson	**GAM direction** Dec.	Non-sign.	Inc.	**Sum**
Negative	16	0	0	16
Non-sign.	3	3	1	7
Positive	0	0	13	13
Sum	19	3	14	36

(a) Pearson correlation.

Sign of Spearman	**GAM direction** Dec.	Non-sign.	Inc.	**Sum**
Negative	8	0	0	8
Non-sign.	8	2	0	10
Positive	3	1	14	18
Sum	19	3	14	36

(b) Spearman or Kendall correlations.

Table 2: Contingency tables comparing the sign of correlations with the direction of monotonicity inferred from monotonic GAMs.

In both the top and bottom panels of Table 2, all 36 empirical hypergraphs are classified according to the monotonicity direction inferred from monotonic GAM fits (columns). According to this classification, the majority, 19 out of 36 hypergraphs (approximately 53%), exhibit a statistically significant *decreasing* relationship between hyperedge size and node degree in bipartite representation. A slightly smaller group, 14 hypergraphs (39%), shows a statistically significant *increasing* relationship. Only 3 hypergraphs (about 8%) are found to have no statistically significant monotonic trend at a stringent threshold of $\alpha = 0.00001$. How do these GAM-derived trend directions align with the sign of standard correlation coefficients such as Pearson, Spearman, and Kendall? We begin by evaluating this alignment for Pearson coefficients, presented in the top table.

Alignment Between GAM Monotonicity and Pearson Signs The top panel of Table 2 compares the GAM monotonicity classification with the sign and statistical significance of Pearson correlations. According to the Pearson-based classification, the most common group consists of hypergraphs with significantly negative correlations, 16 out of 36 cases (approximately 44%), all of them belong to 19 hypergraphs that monotonic GAM classifies as decreasing. The second most frequent group comprises hypergraphs with significantly positive Pearson correlations—13 out of 36 (36%), again, all of them belong to 14 increasing cases identified by the GAM. The least represented group consists of non-significant Pearson correlations, occurring in 7 hypergraphs, while GAM identifies only 3 hypergraphs with no significant monotonicity-all of these 3 hypergraphs belong to 7 hypergraphs indicated by Pearson correlation. In total, the alignment between Pearson sign and monotonic GAM direction is 32 out of 36 cases (about 89%). The 4 misaligned cases are all classified as non-significant by Pearson (second row), while GAM assigns 3 of them (`house-committees`, `senate-committees`, `diseasome` as

visible in Table 8) to the decreasing category and 1 (email-enron) to the increasing category. The Pearson coefficients for these four hypergraphs are all close to zero: −0.0669 (diseasome), −0.0483 (senate-committees), −0.0365 (house-committees), and −0.0024 (email-enron), see Table 8. While these values are not significant at $\alpha = 0.00001$, three out of four are significant at a more conventional threshold of $\alpha = 0.05$, with p-values of 0.0260, 0.8720, 0.0000713, and 0.00038, respectively. Importantly, the first three have negative Pearson coefficients, consistent with the decreasing classification by the GAM model. If we relax the significance threshold to $\alpha = 0.05$, the total alignment between Pearson and monotonic GAM increases to 35 out of 36 hypergraphs (approximately 97%). Moreover, this change would also reclassify three other hypergraphs, amazon, InVS13, and vegas-bars-reviews, from non-significant to either increasing or decreasing, again in agreement with the GAM classification. Thus, under a more conventional $\alpha = 0.05$ threshold, the alignment remains consistently high at around 97%.

The only hypergraph whose sign remained misaligned even under more conventional significance levels of 1% or 5% is the email-enron hypergraph. To inspect this case in detail, Figure 8 in the Appendix presents the scatter plot of data points from its bipartite representation, upon which all correlation coefficients and GAM models were fit. The plot includes three fitted models: an unrestricted GAM (blue solid line), a monotonic GAM (green dashed line), and a linear regression line (red dotted line). The unrestricted GAM reveals a highly complex, non-monotonic relationship. Specifically, the expected degree initially increases from ca. 40 to ca. 55 with hyperedge size from 1 to around 5, then stabilizes at the predicted degree of ca. 55 with statistically insignificant fluctuations between hyperedge sizes of approximately 5 to 15. After that, the predicted degree drops from 55 to around 20 for hyperedge sizes near 30, though this drop is accompanied by wide confidence intervals, suggesting weak statistical support. Notably, the expected degree rises again to approximately 40 for hyperedge sizes above 35. This final level has tighter confidence intervals and an upper bound that lies below the initial degree level of 50, indicating an overall upside-down U-shape.

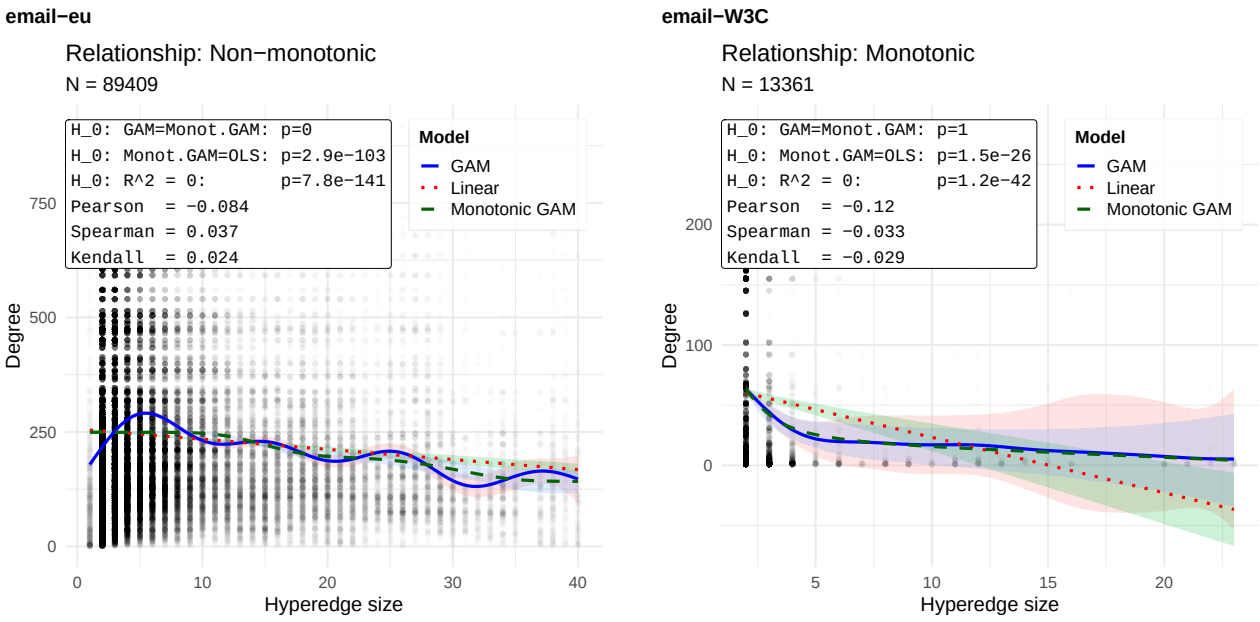
Such a unimodal and balanced pattern, where expected degrees at both tails are similar, results in a flat linear regression fit and a Pearson correlation near zero. In these cases, the sign of the Pearson coefficient may be either positive or negative depending on slight asymmetries in the curve, and can still be statistically significant under conventional α thresholds such as 1% or 5% due to the relatively large sample size (here, $N = 4623$). Meanwhile, the monotonic GAM is constrained to fit either an increasing or decreasing trend. Similar to the Pearson correlation, it is forced to choose one "arm" of the relationship. In this example, the monotonic increasing model results in slightly lower residual sum of squares (SSE), leading to its selection as the better monotonic fit. Consequently, the monotonic GAM indicates an increasing trend, creating a formal misalignment with the negative Pearson coefficient.

Alignment Between GAM Monotonicity and Spearman and Kendall Signs Turning to the bottom panel of Table 2, Spearman and Kendall coefficients exhibit notably lower alignment with the monotonic GAM direction, with only 24 out of 36 hypergraphs (67%) classified in agreement. While these non-parametric coefficients correctly identify all 14 cases with an increasing trend, they misclassify 11 out of 19 hypergraphs with a decreasing trend—labeling 8 as non-significant and, in 3 cases, incorrectly assigning them a statistically significant positive correlation.

Notably, three hypergraphs—kaggle-whats-cooking, house-bills, and email-eu—are assigned an opposite classification. While monotonic GAM identifies these relationships as decreasing, Spearman and Kendall both indicate statistically significant positive correlations. The corresponding Spearman values are 0.014, 0.031, and 0.037, and the Kendall values are 0.009, 0.020, and 0.024, respectively. Although these correlation values are small and only marginally above zero, they are statistically significant at a stringent significance level of $\alpha = 0.00001$, owing to the large sample sizes of their bipartite representations (428,249; 1,248,666; and 89,409 data points, respectively). While their numerical values suggest only a weak positive trend, their qualitative misalignment with the GAM-based classification is noteworthy—particularly when inspecting the underlying scatterplots.

Among the three hypergraphs where the sign of the Spearman and Kendall coefficients misaligns with the direction indicated by the monotonic GAM, kaggle-whats-cooking, house-bills, and email-eu, we focus here on the email-eu hypergraph. However, the qualitative characteristics observed in this case are also present in the remaining two examples that a reader can examine on his own in Figures 10 in Appendix. Figure 4a displays the scatterplot for email-eu, which consists of 89,409 data points in its bipartite representation. The unrestricted GAM (blue solid line) reveals a complex, multimodal structure, characterized by several statistically

significant fluctuations, as evidenced by the narrow confidence intervals[2] estimated at a stringent significance level of $\alpha = 0.00001$.



(a) Example of a non-monotonic relationship in the *email-eu* hypergraph between the number of recipients of a given e-mail (hyperedge size) and the total number of emails sent and received by a given sender/recipient (node degree).

(b) Example of a monotonic relationship in the *email-W3C* hypergraph between the number of recipients of a given e-mail (hyperedge size) and the total number of emails sent and received by a given sender/recipient (node degree).

Figure 4: Examples of non-monotonic (left) and monotonic (right) relationships between hyperedge size and node degree in two e-mail-based hypergraphs.

Visual inspection of Figure 4a, along with the monotonic GAM fit (green dashed line), suggests a globally decreasing relationship, consistent with the negative linear regression line and Pearson coefficient of -0.084. Nevertheless, both the Spearman and Kendall coefficients are positive (0.037 and 0.024, respectively), which likely results from the initial upward trend for small hyperedge sizes (approximately 1 to 5). This is also the most common range of hyperedge sizes in the data, with a mean of 3.56 and standard deviation of 3.40, as summarized in Table 7. Because non-parametric rank-based correlations are sensitive to data density, this dense early region of increasing trend may bias the coefficients toward a positive value, even when the global relationship is decreasing.

Interestingly, the qualitative shape of the relationship in `email-eu` closely resembles that observed in `email-enron` (Figure 8). Both hypergraphs exhibit an initial increase in node degree for small hyperedge sizes, followed by a global decline. In both cases, the resulting relationship forms an upside-down U-shape. The key difference lies in the symmetry: for `email-eu`, the right-side "arm" of the U is longer and more pronounced, which leads to a stronger preference for a decreasing trend in both monotonic GAM and Pearson regression. By contrast, the earlier-discussed `email-enron` case showed a slight preference for the increasing arm.

These complex fluctuations observed in both `email-enron` and `email-eu` may stem from dependencies among observations, violating the assumption of independence. In bipartite representations, such dependencies naturally arise: two edges may share the same node (implying identical or correlated node degrees), or belong to the same hyperedge (implying shared hyperedge sizes). Beyond these structural sources, domain-specific behaviours, such as regular organizational mailing lists, can introduce further correlations. For example, recurring weekly emails sent to the same group of 36 recipients, each with node degree around 30, could generate clusters of duplicate observations. These correlated clusters may skew fitted models, introducing fluctuations in unrestricted GAM fits or biasing regression lines.

[2]Confidence intervals are computed under the assumption that observations are independent. This assumption is revisited later in the text, where we discuss domain-specific and structural reasons why it may be violated, and outline potential approaches for addressing such dependencies.

Addressing these dependencies could improve fitted model robustness but also its indications of monotonicity direction or correlation coefficients. For bipartite-structure-induced correlations, one could model shared nodes or hyperedges explicitly by assuming the correlation among them. However, correlations arising from repeated organizational behaviour (e.g., mailing lists) are harder to detect and may require heuristic de-duplication, such as removing repeated instances of specific hyperedge-size/degree pairs (e.g., many observations with size 36 and degree 30). Developing principled methods to address both sources of correlation remains an interesting direction for future work, which we leave outside the scope of the present study.

So far, we have discussed three hypergraphs that exhibit a clear misalignment between the direction indicated by monotonic GAM and the sign of the nonparametric Spearman and Kendall correlations. These examples: `kaggle-whats-cooking`, `house-bills`, and `email-eu`, are particularly striking, as they not only indicate opposite directional trends but do so with very low p-values, which may lead to overconfident and misleading conclusions. Visually, the relationships in these cases appear strongly non-monotonic or globally trending in the opposite direction, further reinforcing the argument that nonparametric coefficients may fail as reliable indicators of global trend direction.

However, beyond these qualitative discrepancies, a broader issue emerges from the contingency analysis in Table 2, which shows that the largest source of misalignment with GAM monotonicity comes from eight hypergraphs classified as decreasing by GAM, but non-significant according to Spearman and Kendall. Among these, six hypergraphs do have negative correlation coefficients, aligning in sign with the GAM classification. Of these, half: `Hypertext-conference`, `restaurant-reviews`, and `email-W3C`, become statistically significant at the more conventional $\alpha = 0.05$ level, see Table 8. The remaining three: `house-committees`, `senate-committees`, and `diseasome`, remain non-significant even at that threshold.

The remaining two hypergraphs in this group: `dblp` and `SFHH-conference`, are misaligned not only in terms of statistical significance but also in the direction of the correlation sign. Both are classified as *decreasing* by monotonic GAM yet have *positive* Spearman and Kendall coefficients. Although not significant at the stringent $\alpha = 0.00001$ level, these correlations become significant at $\alpha = 0.01$, see Table 8. A closer look at their scatterplots, available in the Appendix, reveals familiar patterns: fluctuating relationships that start with a short upward trend heavily supported by dense data, followed by a broader, downward trend.

The case of `SFHH-conference`, illustrated in Figure 12, is particularly noteworthy. The unrestricted GAM reveals a smooth, complex, and non-monotonic pattern, starting with a slight but data-dense upward trend and evolving into a pronounced decline. The monotonic GAM and linear regression both capture this global downward trend, the latter supported by a negative Pearson coefficient of -0.027. In contrast, Spearman and Kendall yield small positive values of 0.012 and 0.0097, respectively, both statistically significant at $\alpha = 0.01$. This stark divergence exemplifies once more how nonparametric rank-based correlations, while useful for monotonic relationships, can fail to reflect global trends in complex empirical data.

These two misaligned hypergraphs, when combined with the three earlier cases of strong misclassification, yield a total of five hypergraphs where Spearman and Kendall correlations exhibit a statistically significant sign opposite to the monotonic trend identified by GAM at the $\alpha = 0.01$ level. This underscores a substantial limitation in using nonparametric correlation coefficients as standalone indicators of relationship direction in empirical hypergraph data.

Summary This analysis demonstrates that Pearson correlation coefficients are remarkably well-aligned with the direction of global trends inferred from monotonic GAM models, with an agreement rate of 89% under a stringent $\alpha = 0.00001$ and rising to 97% at a conventional threshold of $\alpha = 0.05$. In contrast, non-parametric Spearman and Kendall correlations exhibit considerably lower alignment, agreeing with monotonic GAM in only 67% of cases. Notably, five hypergraphs show statistically significant signs in Spearman or Kendall that are opposite to the monotonic direction indicated by the GAM. This discrepancy is most apparent in complex, non-monotonic settings where local data density can bias rank-based measures. The results caution against relying solely on Spearman or Kendall for detecting global trends in empirical hypergraphs, especially under large sample sizes and non-linear structures. Instead, monotonic GAMs, paired with Pearson coefficients and visual inspection, offer a more robust framework for trend identification in complex bipartite hypergraph data.

3.2.3 Interpretation Corner: Segment-Level Monotonicity Patterns

So far, we have focused on accurately measuring the relationship between hyperedge size and node degree. However, given the empirical nature of the hypergraphs under analysis and their semantic interpretations, we

now aim to interpret the observed relationship signs through the lens of what hyperedges and nodes represent in each case. This interpretation goes beyond pure measurement and offers rationale grounded in the semantics of each dataset.

Figure 3b presents the frequency of monotonicity directions (increasing, decreasing, or non-significant) detected by monotonic GAMs, grouped by hypergraph segment. As described in subsection 2.3, these monotonicity classes are assigned based on statistical testing using a stringent significance level $\alpha = 0.00001$. A monotonic trend is deemed "non-significant" if neither an increasing nor a decreasing GAM model fits significantly better than a flat baseline.

The figure reveals strong homogeneity of GAM monotonicity direction within hypergraph segments: 10 out of 13 categories exhibit perfect consistency in trend direction across all included hypergraphs. Of the remaining three, two segments (`Physical Contact` and `User-review`) contain a single outlier hypergraph (`InVS13` and `vegas-bars-reviews`, respectively), both classified as non-significant. Only the `Email` segment contains a single case (`email-enron`) with an opposite trend, which, as discussed earlier (e.g., Figure 4a), exhibits a highly non-monotonic U-shaped pattern. This overall trend consistency supports our earlier findings in Subsection 3.1, where limited variability of correlation coefficients within segments, especially under bipartite representation, led to relatively high η^2 values.

Positive monotonic trends dominate several hypergraph categories, including `Physical Contact`, `User-Thread`, `Tag-Question`, and `Drugs`. Focusing on the first two, we can offer interpretable, domain-informed explanations for why increasing relationships between hyperedge size and node degree are expected. In the **Physical Contact** segment (e.g., `contact-high-school`, `contact-primary-school`, `Malawi-village`, `Hypertext-conference`), nodes represent individuals equipped with sensors, and hyperedges correspond to physical group interactions over brief time intervals. An individual participating in larger group interactions is likely to engage with more people overall, thus appearing in more interactions. For example, a student present in large classroom settings will tend to accumulate more contacts than a student mostly present in one-on-one or small group interactions. Hence, larger hyperedges naturally imply higher node degrees, leading to a positive correlation.

Similarly, in the **User-Thread** segment (e.g., `threads-ask-ubuntu`, `threads-math-sx`), hyperedges represent discussion threads on Q&A forums, and nodes are users contributing to those threads. Larger threads typically attract more engaged or experienced users who tend to participate in many discussions. Conversely, users who are active across multiple threads are more likely to contribute to longer, multi-user conversations. Therefore, there is a natural expectation of a positive relationship between thread size and user activity levels, consistent with the increasing trend found in the data.

Several hypergraph segments exhibit a dominant *negative* monotonic trend between hyperedge size and node degree, particularly the `Political`, `Participant-Conference`, and `Person-Place` categories. In the **Political** segment, such as `house-bills` and `house-committees`, nodes represent political actors (e.g., members of Congress), and hyperedges represent either legislative bills or committee memberships. A negative relationship in this context indicates that politicians involved in large coalitions (e.g., large bills with many cosponsors) tend to participate in fewer overall bills. This aligns with political specialization: high-frequency participants may focus on narrow, small-scale initiatives, while those contributing to large, broad coalitions do so more occasionally. Moreover, committee memberships are often limited in number due to institutional constraints, and members of large committees may serve on fewer committees overall, reinforcing the observed inverse relationship.

In the **Participant-Conference** segment (e.g., `Hypertext-conference`, `SFHH-conference`), hyperedges represent group interactions at specific time intervals, and nodes are attendees. A negative correlation in this setting suggests that participants found in large group gatherings (e.g., plenary sessions) are less likely to be involved in many distinct interactions across time. In contrast, those who accumulate many contacts tend to do so through repeated small-group interactions (e.g., informal meetings or hallway conversations), leading to lower average hyperedge sizes. Similarly, in the **Person-Place** segment (e.g., `got`, `evernote-places`), hyperedges are shared scenes or event venues, while nodes are characters or individuals. The negative trend reflects that individuals who frequently appear in scenes or attend events often do so in small settings, supporting either focused plotlines in narrative data (like `got`) or niche appearances in real-world event data (like `evernote-places`). In contrast, characters or artists appearing in large group events typically do so less frequently. These observations align well with empirical social and narrative dynamics and support the statistical findings.

Summary This interpretive analysis demonstrates that the direction of monotonic relationships observed between hyperedge size and node degree is not random but meaningfully structured across semantic hypergraph

segments. Positive monotonic trends are prevalent in settings where participation in larger groups is associated with higher overall engagement, such as physical contact networks or user interaction threads, reflecting natural social dynamics and patterns of individual activity. In contrast, negative trends emerge in domains where involvement in large groups tends to limit broader participation, due to constraints or role specialization, as seen in political affiliations, conference attendance patterns, or media appearances. The strong within-segment consistency in monotonic direction, along with interpretable domain-based rationales, supports the validity of our classification procedure and reinforces the value of semantic segmentation for interpreting structure-function relationships in empirical hypergraphs.

3.2.4 Robustness Check: Logarithmic Feature Transformation

An important design decision we investigated in this study concerns whether to apply feature engineering transformations to the variables of interest, particularly natural logarithm transformations of either the hyperedge size or node degree, or both, when computing correlations. Logarithmic transformation is a widely used technique in network science due to its ability to compress skewed degree distributions, which often follow heavy-tailed or power-law forms. This transformation facilitates clearer visualization and improves the stability of statistical modelling [93, 14]. In economics, it is commonly employed in the estimation of Cobb–Douglas production functions, where taking logarithms transforms multiplicative models into additive linear ones, stabilizes variance, and allows for coefficients to be interpreted as elasticities or percentage changes [73, 133, 54].

Intuitively, one might expect that monotonic transformations like the logarithm would not affect non-parametric correlation coefficients such as Spearman or Kendall, nor the shape of non-parametric models such as GAMs. This expectation holds, but only under specific conditions. When the predictor variable X is transformed using a logarithm, non-parametric measures and models remain numerically unchanged, as their ranks and monotonicity are preserved. However, when the dependent variable Y is log-transformed, all models and correlation coefficients, including non-parametric ones, may be affected. This is because $\mathbb{E}[\log(Y|X)] \neq \log(\mathbb{E}[Y|X])$, and hence the underlying Conditional expectation function (CEF), see [8], is altered in a way that affects model fit and statistical outputs across the board.

The key takeaway is that applying a logarithmic transformation to the predictor X has no impact on non-parametric measures or GAMs, but will affect parametric measures such as Pearson correlation and linear regression. In contrast, log-transforming the dependent variable Y has consequences for all modelling approaches and thus warrants careful consideration. Accordingly, sensitivity analyses should be conducted when log-transformations are applied to Y, which we report below.

Despite the theoretical implications outlined above, our empirical evaluation shows that the actual impact of logarithmic transformation on results is quantitatively minor and does not alter the qualitative conclusions of our study. Specifically, the Pearson correlations between outcomes obtained from logarithmic and non-logarithmic setups exceed 96% across all cases. Therefore, for clarity and consistency, we focus our main results on the non-logarithmic setup. Nevertheless, the primary conclusions remain robust and fully extend to the logarithmic case. The decision to apply a logarithmic transformation should be guided more by interpretability, e.g., the need of switching from additive to multiplicative interpretations, than by model performance.

Summary This robustness check confirms that logarithmic transformations, while theoretically impactful, especially when applied to the dependent variable, have only minimal empirical effect on our results. The high concordance between log- and non-log-transformed outcomes (Pearson correlations exceeding 96%) indicates that our classification and correlation patterns are stable. We recommend that log-transformations be considered primarily for interpretability purposes rather than performance gains, and sensitivity checks should accompany any such transformation, particularly when applied to the dependent variable.

3.2.5 Summary and Recommendations

To summarize, although all three correlation coefficients: Pearson, Spearman, and Kendall, show comparable segment-level alignment as measured by η^2 (ranging between 0.66 and 0.68 for bipartite representation; see Table 1), Pearson stands out for its consistency in capturing the direction of the global relationship between hyperedge size and node degree. Specifically, Pearson achieves 89% alignment with the monotonicity direction inferred from shape-constrained GAMs at a stringent $\alpha = 0.00001$, improving to 97% at $\alpha = 0.05$ (see Table 2). In contrast, Spearman and Kendall coefficients align with GAM trends in only 67% of cases and are more prone to directional misclassification, including sign reversals.

Several problematic cases such as: `email-eu`, `house-bills`, `kaggle-whats-cooking`, `SFHH-conference`, and `dblp`, illustrate scenarios where Spearman and Kendall coefficients suggest the opposite trend from Pearson. In all these instances, visual inspection reveals either a clear global downward trend or complex multimodal patterns where Pearson better captures the overall direction. Moreover, in cases involving U-shaped or weakly curved relationships (e.g., `email-enron`, `diseasome`), Pearson returns near-zero values, consistent with the lack of a strong directional trend, whereas non-parametric coefficients often report a misleadingly strong positive or negative association.

These findings indicate that while non-parametric methods like Spearman and Kendall are well-suited for detecting monotonic but nonlinear relationships, they are more sensitive to local structure and can be misled by multi-modality or curvature, sometimes even reversing the sign of the association. Pearson, although traditionally associated with linearity, proves more robust in capturing the global relationship direction, especially in complex empirical settings.

Therefore, we recommend the following:

- Use the bipartite representation as the default preprocessing method for quantifying the relationship between hyperedge size and node degree, due to its superior η^2 scores and segment-level alignment.
- When selecting a single correlation coefficient, prefer Pearson over Spearman and Kendall, even in nonlinear settings, because it more reliably reflects the direction of the global relationship, particularly in the presence of multimodal or weakly U-shaped trends.
- Be cautious with non-parametric correlations in the presence of complex structures, as they may produce inflated coefficients and misleading signs that do not correspond to the dominant trend in the data.

In summary, while non-parametric coefficients offer value in detecting monotonic trends, Pearson is more robust for summarizing complex real-world patterns, making it the preferred choice for global characterization. The next subsection explores more nuanced characterizations of the hyperedge size–degree relationship beyond a single numerical summary.

3.3 Identifying Relationship Types Between Hyperedge Size and Node Degree

In the previous two Subsections 3.1 and 3.2, we focused on identifying optimal design choices, i.e., the data preprocessing strategy and correlation coefficient, that produce a single numerical indicator well aligned with both the semantic segment of the hypergraph and the global trend structure identified via shape-constrained GAMs. These indicators aim to reflect dominant relationships while being robust to local fluctuations in complex patterns. However, summarizing even a two-dimensional relationship with a single number can be overly reductive. As demonstrated in earlier sections, the relationships between hyperedge size and node degree can exhibit diverse and often complex forms.

The goal of this subsection is to shift the focus from a quantitative summary to a qualitative understanding of the relationship types observed across empirical hypergraphs. Rather than measuring strength, we classify each of the 36 empirical hypergraphs into one of four qualitative categories that characterize the nature of the relationship between hyperedge size and node degree: (1) non-monotonic (including unimodal or multimodal patterns), (2) monotonic, (3) linear, and (4) no relationship. This classification is carried out using the statistical procedure introduced in Subsection 2.3, which combines models' fitting with sequential hypothesis testing.

In Subsection 3.3.1, we present one illustrative example from each category to clarify how the classification is determined in practice, and to better understand the advantages and limitations of the proposed procedure. Then, in Subsection 3.3.2, we move beyond individual examples to summarize the empirical distribution of relationship types and perform Bayesian inference to generalize findings to the broader population of hypergraphs, beyond the 36 examples analyzed here. Finally, Subsection 3.3.3 presents robustness checks to assess the sensitivity of our classifications to modelling assumptions, such as variable directionality and feature engineering. Throughout this section, we restrict our analysis to the bipartite representation, which was previously shown in Subsection 3.1 to be the most informative and stable preprocessing strategy for analyzing hyperedge size–degree relationships.

3.3.1 Examples of Identified Relationship Types

In this subsection, we present four representative hypergraphs, each illustrating one of the relationship types identified in our classification scheme. We begin with the most complex case of non-monotonic relationship

and proceed through monotonic and linear examples, concluding with a case of no apparent relationship. These examples are paired with the statistical test results introduced in Subsection 2.3, providing concrete illustrations of how the classification procedure operates, including its strengths and limitations. In the next subsection, Subsection 3.3.2, we transition from individual cases to population-level insights about the distribution of relationship types.

Non-monotonic Relationship This relationship type has already been illustrated in Figure 4a for the `email-eu` hypergraph and discussed in detail in Subsection 3.2.2. In this case, non-parametric correlation coefficients such as Spearman (0.037) and Kendall (0.024) suggest a weak positive trend due to an initial upward pattern. However, this masks a stronger global downward trend, more accurately captured by the negative Pearson correlation (-0.084).

To formalize this classification, we focus on the top-left inset of Figure 4a, which displays two ANOVA tests and one F-test as per our methodology. The first ANOVA test compares an unconstrained GAM (blue line) against a monotonic GAM (green line), with the null hypothesis assuming equivalence. A p-value near zero indicates strong evidence that the unconstrained model fits the data significantly better, revealing a non-monotonic (e.g., unimodal or multimodal) pattern. The unrestricted GAM clearly fluctuates, and its narrow confidence intervals at $\alpha = 0.00001$ support the conclusion that these deviations are statistically significant. As the result of this first test is decisive, the procedure terminates at this step, classifying the relationship as non-monotonic. The other two test results are reported for completeness but do not affect the classification.

Monotonic Relationship Figure 4b shows an example of a monotonic relationship in the `email-W3C` hypergraph. Visually, the unconstrained GAM (blue line) closely follows the monotonic GAM (green line), both showing a steep initial decline followed by a gradual flattening. The visual similarity and overlapping confidence intervals suggest that the additional flexibility of the unconstrained GAM is unnecessary. This is confirmed by the first ANOVA test, which yields a high p-value, indicating no significant difference between the models.

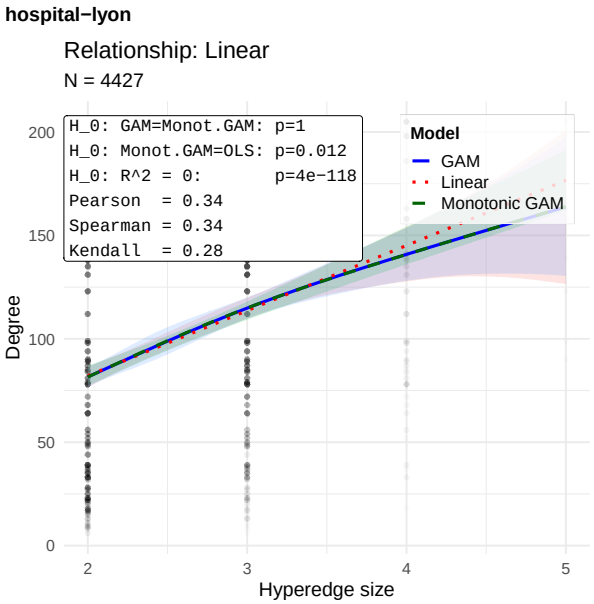
Following our procedure, we proceed to test whether a simple linear regression (dotted red line) could adequately describe the data. Here, the regression line visibly diverges from the monotonic GAM fit across most of the hyperedge size range, and the corresponding ANOVA test produces an extremely small p-value (1.5×10^{-26}), confirming that a linear model is insufficient. Therefore, the final classification is monotonic. No further testing (e.g., the F-test) is required.

Although `email-enron`, `email-eu`, and `email-W3C` all belong to the same semantic category of email-based hypergraphs, their structural patterns differ notably. Both `email-enron` and `email-eu` show an initial increase in average node degree up to a hyperedge size of five, followed by a gradual decline, suggesting that moderate group sizes are associated with higher engagement, while very large groups dilute individual participation. In contrast, `email-W3C` shows a steady, monotonic decline starting from the smallest hyperedge sizes, likely reflecting the dynamics of mailing lists, where communications tend to be broadcast-like and less reciprocal.

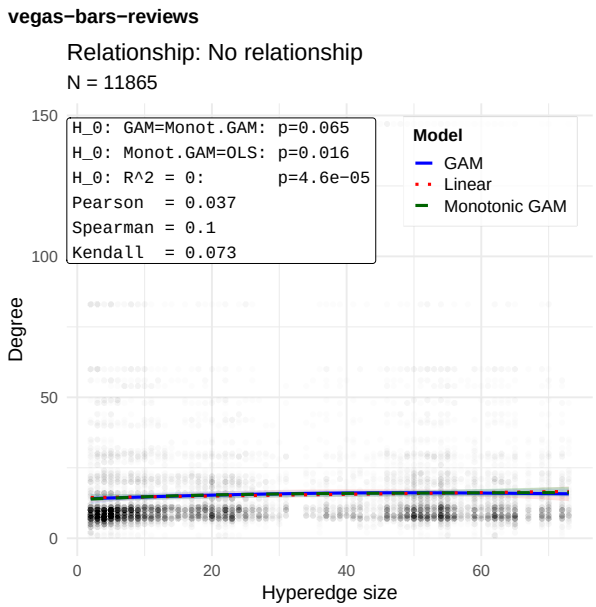
These differences can be attributed to the nature of the datasets. The Enron corpus consists of internal corporate emails exchanged primarily among upper management between 1999 and 2002 [76], while the EU dataset originates from email interactions within a European research institution [138], likely reflecting team-based coordination. In both cases, email exchanges are driven by project work and organizational structures that foster high interaction in small to medium groups. Conversely, the W3C emails were collected from public mailing lists used for technical discussions and announcements in the broader web standards community. Such lists typically feature broadcast-style communications with minimal back-and-forth, which explains the sharp and steady decline in individual involvement as hyperedge size increases. These observations highlight that structural patterns in hypergraph data are strongly shaped by the institutional, cultural, and functional context of the underlying communication systems.

Linear Relationship The `hospital-lyon` hypergraph, depicted in Figure 5a, provides a clear example of a linear relationship. As summarized in Table 5, this dataset captures group interactions among healthcare workers and patients in a Lyon hospital ward. The hyperedge sizes are limited to 2–5 participants, with most interactions involving groups of size 2 or 3. This restricted domain naturally favours simpler functional forms, such as linear models.

Model fits show a consistent positive trend: individuals involved in larger group interactions tend to accumulate more total interactions. This aligns with intuition in the hospital setting, where participation in larger gatherings, such as rounds or shift changes, implies broader involvement across the ward.



(a) Linear relationship in *hospital-lyon* hypergraph between the number of hospital staff and patients attending a gathering (hyperedge size) and the total number of gatherings attended by each staff member or a patient (degree).



(b) No relationship in *vegas-bars-reviews* hypergraph between the number of reviews for a bar in a given month (hyperedge size) and the number of reviews submitted by each user (degree).

Figure 5: Examples of a linear relationship (left) and no discernible relationship (right) between hyperedge size and node degree in two empirical hypergraphs.

While the fit of unrestricted GAM (blue line) suggests a slight concave curvature at hyperedge size 4, this is based on sparse data and is not statistically robust. Formal testing supports the linear interpretation. The first ANOVA test comparing the unconstrained GAM and a monotonic GAM yields a p-value close to 1, suggesting no evidence against monotonicity. A second ANOVA test comparing the monotonic GAM to a linear model produces a p-value of 0.012. Under our stringent significance threshold of $\alpha = 0.00001$, this result is not sufficient to reject the linear model. Additionally, a comparison between the linear fit and a constant mean model strongly favours the former (with a p-value of 4×10^{-118}), reinforcing the classification as a linear relationship.

Two factors could lead to reclassification. First, using a relaxed significance threshold (e.g., $\alpha = 0.02$ or 0.05) would favour the monotonic GAM due to the minor curvature. Second, reversing the model direction (hyperedge size as response, degree as predictor) reveals a wider range of degrees (6–205, see Table 6), and as checked in Section 3.3.3, this leads to reclassification as monotonic.

No Relationship Figure 5b illustrates a case of no significant relationship between node degree and hyperedge size in the `vegas-bars-reviews` hypergraph. This dataset captures Yelp users (nodes) who reviewed the same bar in Las Vegas within a one-month period (hyperedges). As shown in Table 7, the number of reviews per bar per month (i.e., hyperedge size) ranges from 2 to 73, while individual users have submitted between 1 and 147 reviews overall, with an average of 9.6. Despite this wide variability, the relationship between the number of reviews a bar receives in a given month and the total number of reviews submitted by its reviewers appears to be flat, with no clear trend.

This is confirmed through a sequence of statistical tests. The first ANOVA test comparing an unconstrained GAM to a monotonic GAM yields a p-value of 0.065, suggesting limited evidence against monotonicity. The second test comparing the monotonic GAM to a linear model produces a p-value of 0.016, indicating that the linear model fits the data sufficiently well. However, an F-test comparing the linear model to a constant mean returns a p-value of 4.6×10^{-5}, leading to the rejection of the linear model under our strict significance threshold of $\alpha = 0.00001$. Thus, despite a slight upward trend suggested by the positive Pearson correlation

($r = 0.037$), the final classification for this configuration is "no relationship." It is worth noting that under a more conventional threshold (e.g., $\alpha = 0.01$), the relationship would be classified as linear.

From an interpretive standpoint, one might expect a weak positive relationship in this setting. Bars that attract many reviewers in a given month may be more popular or prominent, and such venues are likely to be visited and reviewed by more active Yelp users, who tend to submit reviews more frequently overall. However, the noisy nature of user behaviour, varying reviewing habits, and the casual context of online review platforms likely dilute any clear structural trend, resulting in only a mild correlation that fails to reach significance under strict criteria.

Interestingly, the result is not robust to a reversal of the variables. When hyperedge size and node degree are swapped (i.e., node degree as the predictor and hyperedge size as the response), the model is reclassified as "non-monotonic." This reversal highlights the sensitivity of the classification to model direction and supports a more nuanced interpretation of the data. Additionally, the earlier p-value of 0.065 already suggests some deviation from monotonicity even in the original configuration.

As will be further discussed in Section 3.3.2, only three out of 36 empirical hypergraphs fall into the "no relationship" category: `vegas-bars-reviews`, `amazon`, and `InVS13`. Each exhibits unique structural characteristics that help explain this classification. In the case of `amazon` and `InVS13`, the lack of a relationship is more robust and can be attributed to the highly limited range of hyperedge sizes, only 3 distinct values (2, 3, 4) for `InVS13`, and just 1 to 6 for `amazon`, as well as low node degree variability. For example, `amazon` users have degrees ranging only from 1 to 4. These constraints naturally bias the analysis toward simpler models. Moreover, both datasets are relatively small (5,112 nodes for `InVS13` and 19,380 for `amazon`), which, under the strict $\alpha = 0.00001$ threshold, reduces the power to detect more subtle effects. Notably, both of these hypergraphs remain classified as "no relationship" even when the predictor and response variables are swapped, further confirming the robustness of their categorization.

Summary The four examples presented above: non-monotonic (`email-eu`), monotonic (`email-W3C`), linear (`hospital-lyon`), and no relationship (`vegas-bars-reviews`), illustrate the diversity of patterns that can emerge between node degree and hyperedge size in empirical hypergraphs. These case studies not only demonstrate the behaviour of different model fits but also highlight the strengths and limitations of the statistical procedure introduced in Subsection 2.3. They show how both statistical evidence and domain knowledge contribute to classification outcomes, and how modelling choices, such as the selection of the response variable or the choice of significance threshold, can affect the resulting interpretation. These examples serve as a foundation for understanding the broader conclusions drawn in the next subsection. The remaining 32 empirical hypergraphs, along with one synthetic hypergraph generated using the ABCD-h model, are visualized in the same manner in Appendix Subsection A.4.3.

In the following section, we move beyond individual cases to examine the overall distribution of relationship types across the full set of 36 hypergraphs and population of hypergraphs. This analysis allows us to identify common structural patterns, assess how frequently each relationship type occurs, and explore what characteristics may be associated with different classes of relationships.

3.3.2 Distribution of Relationship Types

This subsection reports the distribution of the four identified relationship types across the 36 empirical hypergraphs and provides inferential insight into what might be expected in the broader population of hypergraphs.

Table 9 in the Appendix presents, for each hypergraph, the p-values from the three statistical tests and the relationship type classification described in Subsection 2.3, based on a conservative significance level of $\alpha = 0.00001$. To summarize these detailed results, Table 3 shows the empirical distribution of relationship types across the dataset. Among the 36 hypergraphs, only 3 (8.3%) exhibit no discernible relationship between hyperedge size and node degree. A majority of hypergraphs (18/36, 50%) show a monotonic relationship (including linear), while the remaining 15 (41.7%) demonstrate a non-monotonic relationship. This distribution already provides strong empirical evidence against no relationship in real-world hypergraph data. Moreover, the use of a stringent α level reduces the likelihood of false positives, reinforcing the robustness of our classifications in the presence of large datasets.

Given this empirical evidence, it would be misleading to assume that no relationship exists between hyperedge size and node degree in a randomly chosen hypergraph. On the contrary, one should generally expect at least a monotonic relationship, if not a more complex, non-monotonic pattern. These findings are particularly relevant

Relationship Type	Count	Share (%)	Cumulative Share (%)
No relationship	3	8.3	8.3
Linear	6	16.7	25.0
Monotonic	12	33.3	58.3
Non-monotonic	15	41.7	100.0

Table 3: Distribution of four identified relationship types—non-monotonic, monotonic, linear, and no relationship—across 36 empirical hypergraphs.

for the design of generative models for hypergraphs, which ought to account for such structural dependencies rather than assuming independence between hyperedge size and node degree.

To generalize beyond our finite sample, we conduct Bayesian inference using non-informative uniform priors and derive posterior Beta distributions for selected proportions. Specifically, we compute Bayesian Credible Intervals (BCIs) for the fractions of hypergraphs showing no relationship and those showing non-monotonic relationships. For the "no relationship" category (3 out of 36), the posterior mean is 10.5%, with a 95% BCI of $(3.0\%, 21.9\%)$ and a more conservative 99% BCI of $(1.9\%, 26.6\%)$. These intervals suggest that, in the general hypergraph population, the fraction of cases with no relationship is likely below 25%, reinforcing the idea that such cases are relatively rare.

Conversely, the proportion of non-monotonic relationships is estimated at 42.1% (posterior mean), with a 95% BCI of $(27.1\%, 57.9\%)$ and a 99% BCI of $(22.0\%, 62.7\%)$. This indicates that non-monotonic relationships may not only be common but could even constitute the majority in the broader population. Taken together, these findings highlight the prevalence and diversity of structural dependencies in real-world hypergraphs and call into question modeling assumptions that treat group size and individual connectivity as unrelated.

Summary This analysis establishes that relationships between hyperedge size and node degree are pervasive in empirical hypergraphs, with non-monotonic and monotonic patterns dominating the landscape. The scarcity of hypergraphs with no detectable relationship, both in sample and in posterior inference, underscores the importance of incorporating these structural regularities into modelling frameworks. In the next subsection, we test the robustness of these classifications by examining the sensitivity of results to changes in modelling assumptions, such as reversing the direction of the dependent variable.

3.3.3 Robustness Check: the Choice of X and Y Axis

One of the key modelling choices in our classification procedure was to treat hyperedge size as the predictor (X-axis) and node degree as the dependent variable (Y-axis) when fitting GAM models. While this decision does not influence the outcome of the F-test used for comparing linear regression with a constant model, it can substantially affect the fit and flexibility of nonparametric models like GAMs, which in turn may influence the final classification into one of the four relationship types. To assess the robustness of our conclusions, we repeated the entire classification procedure after switching the roles of the two variables.

Table 4 presents a confusion matrix comparing relationship classifications under the original (columns) and reversed (rows) configurations. A notable trend is the increase in the number of hypergraphs classified as non-monotonic under the reversed setup: 22 out of 36, compared to 15 in the original. Meanwhile, the number of linear classifications drops from 6 to just 1, and the count of "no relationship" cases remains unchanged at 3. This confirms our earlier finding from Subsection 3.3.2 that hypergraphs without any significant structural relationship are rare. The shift toward more complex categories upon reversing axes suggests that our original classifications may, if anything, understate the prevalence of non-monotonic patterns.

A closer inspection reveals that 21 out of 36 hypergraphs (58%) are classified identically in both settings, while only 2 hypergraphs (5.5%) are placed in entirely different categories (e.g., from "no relationship" to "non-monotonic"). Of the 15 cases that were differently reclassified, 12 shifted toward a more complex relationship type, most frequently from monotonic to non-monotonic, while only 3 moved to a simpler category.

This tendency toward increased complexity is expected. In most hypergraphs, node degree tends to have a higher variance than hyperedge size. Since GAMs are more flexible when the predictor variable spans a wide range, reversing the axes effectively exposes the model to greater variation, allowing it to detect more nuanced,

		X: Hyperedge size, Y: Degree				
		None	**Linear**	**Monotonic**	**Non-mon.**	**Total**
	None	**2**	0	0	1	**3**
X: Degree,	**Linear**	0	**1**	0	0	**1**
Y: Hyperedge size	**Monotonic**	0	2	**6**	2	**10**
	Non-mon.	1	3	6	**12**	**22**
	Total	**3**	**6**	**12**	**15**	**36**

Table 4: Confusion matrix comparing relationship classifications with original variable assignment (columns) versus reversed (rows).

non-monotonic patterns. This reinforces the idea that the relationship between degree and hyperedge size is often complex and context-dependent.

Summary Reversing the roles of predictor and response variables confirms the robustness of our core findings while also highlighting an important asymmetry: more complex, non-monotonic relationships become even more prevalent when degree is used as the predictor. While over half of the classifications remain unchanged, most discrepancies result in a shift toward greater complexity, not simplification. This suggests that our main conclusions about the widespread and intricate nature of degree–hyperedge size relationships are, if anything, conservative.

3.3.4 Summary and Recommendations

This section consolidates the findings of our investigation into the types of relationships between hyperedge size and node degree across 36 empirical hypergraphs and provides actionable recommendations for researchers analyzing such data. Our classification procedure, grounded in a sequence of nested statistical tests applied to shape-constrained GAMs and linear models, revealed that structural dependencies between these two quantities are widespread, often non-linear, and in many cases non-monotonic. This holds true even under stringent significance thresholds and is robust to modelling assumptions, such as the choice of predictor and response variable.

Only a small minority (3 out of 36) of hypergraphs exhibited no detectable relationship. The majority fell into the monotonic (including linear) or non-monotonic categories, with non-monotonic relationships constituting 42% of cases. Bayesian inference confirms that such patterns likely generalize beyond our finite dataset, with the fraction of hypergraphs showing no relationship unlikely to exceed 25%. These results challenge common modeling assumptions of independence or linearity between hyperedge size and node degree and call for more nuanced representations in both descriptive and generative settings.

We recommend that analysts avoid relying solely on a single correlation coefficient, especially rank-based ones such as Spearman or Kendall, as these may fail to capture underlying complexity, particularly in the presence of U-shaped or multimodal trends. Instead, we suggest a model-based classification approach, such as the one used here, which compares the fit of multiple nested models and provides interpretable outcomes grounded in statistical evidence. When a single summary metric is needed, we recommend using Pearson correlation on data preprocessed via the bipartite representation, as this approach showed the strongest alignment with model-based trends across semantic segments.

In sum, we recommend that researchers treating hypergraph data, whether for descriptive analysis, predictive modelling, or generative simulation, take into account the existence and complexity of the hyperedge size–degree relationship. This dependency is both widespread and interpretable, varies systematically across semantic domains, and should be incorporated into statistical modelling and synthetic data generation to ensure more faithful and functionally relevant representations of hypergraph structure.

4 Discussion: On the Impact of Degree–Hyperedge Size Correlation in Social Dynamics

The primary motivation for this study is to identify and quantify the relationship between two fundamental structural properties of hypergraphs: hyperedge size and node degree. This relationship is of particular interest

due to its potential influence on emergent behaviours and dynamical processes that unfold on such higher-order structures. In the context of classical pairwise networks, extensive research has shown that structural features, such as degree distribution, clustering, and degree-degree correlations, can critically affect dynamics including epidemic spreading [104, 97], diffusion [71, 29], and the emergence of cooperation [117, 107].

Although the literature on dynamical processes in hypergraphs is still emerging, it has already produced a growing body of work demonstrating the importance of higher-order interactions in shaping collective behaviour. This section reviews selected processes modelled on hypergraphs, such as social contagion, influence diffusion, and multiplayer cooperation, and proposes hypotheses on how structural correlations between hyperedge size and node degree might impact these dynamics. By doing so, we aim to highlight the broader relevance of our empirical findings to modelling and understanding complex systems through a higher-order lens.

4.1 Spreading Dynamics of Social Contagions and Epidemics

Understanding how behaviours, ideas, and infectious diseases propagate across social systems is a central topic in network science and social network analysis. Traditional models of contagion dynamics, such as SIS or SIR (Susceptible-Infected-Susceptible / Susceptible-Infected-Recovered), have been extensively studied in the context of pairwise networks, where each edge represents a dyadic interaction between individuals [6, 58, 74]. In such settings, it is well-established that community structure, clustering, and degree heterogeneity strongly influence both the epidemic threshold and the eventual prevalence of contagion [87, 70, 101]. More recently, researchers have extended these models to higher-order structures, particularly hypergraphs, which allow for the direct modelling of group interactions that cannot be reduced to dyads [113, 21, 121, 34]. Unlike simplicial complexes, which require inclusion of all lower-order simplices, hypergraphs offer greater flexibility in representing complex social contexts, where group interactions do not always entail all possible sub-interactions [49, 100]. This makes them particularly well-suited for modelling settings such as classrooms, households, or workplaces, where contagion spreads through collective exposure rather than simple pairwise contact.

One of the earliest contributions to modelling spreading dynamics in higher-order social systems is the work by [21], summarized in [17]. In this framework, individuals are modelled as nodes, and each hyperedge represents a shared social context such as a household or workplace [60, 13]. The authors analyzed an SIS model governed by a continuous-time Markov chain in which the infection and recovery processes follow Poisson dynamics. Notably, the infection probability of a susceptible individual v depends on higher-order group structures. Specifically, the probability of infection over a small time interval Δt is given by:

$$P_{\text{infect}} = 1 - \exp\left(-\tau \Delta t \sum_{v \in e} f(i_e)\right),$$

where the summation is over all hyperedges e containing the susceptible node v, i_e is the number of infected nodes in hyperedge e, and $f(i_e)$ is a non-decreasing function capturing the infection pressure. While prior work explored various forms of f, including linear and tangent functions, [21] showed that much of the qualitative behaviour of the system could be reproduced with a simple piecewise-linear function:

$$f(x) = \begin{cases} x, & \text{if } 0 \leqslant x \leqslant c \\ c, & \text{if } x > c. \end{cases}$$

Here, the parameter c serves as a saturation threshold: once the number of infected individuals in a group exceeds c, additional infections no longer increase the infection rate. This diminishing returns property, frequently observed in socio-economic systems [126, 84, 63], adds realism to the modelling of social contagions.

Simulations using this model revealed that both hyperedge size and structural heterogeneity significantly influence the contagion dynamics. Larger average hyperedge sizes accelerate the early stages of spread, while greater heterogeneity in group sizes leads to faster initial outbreaks but lower long-term infection prevalence. Degree heterogeneity, the variation in the number of groups an individual belongs to, further modulates outcomes: homogeneous degree distributions yield slower onset but higher final infection levels, whereas heterogeneous configurations enable rapid early transmission that later plateaus at lower steady-state levels.

However, this line of research has not yet addressed the potential role of correlation between hyperedge size and node degree. In scenarios where both node degree and hyperedge size are fixed or drawn independently, the correlation between them is by design zero. But it remains unclear whether the heterogeneous configurations

used in simulations by [21] unintentionally introduced non-zero correlations due to the design of generative configuration model [30]. If such correlations were present, they may have contributed to the observed acceleration in early contagion spread. Disentangling the effects of degree/size heterogeneity from the correlation between them requires simulation experiments on hypergraphs where both marginal distributions and correlations can be independently controlled.

We hypothesize that a positive correlation between hyperedge size and node degree, empirically observed in many social hypergraphs as demonstrated in this paper, could amplify the early phase of spreading even further. Intuitively, individuals who are more socially active tend to participate in larger group settings (e.g., conferences, social events), making them likely early spreaders. This structural coupling concentrates initial exposure in hubs of group activity, facilitating rapid diffusion of behaviours or infections. However, as shown in [21], such accelerating effects may also lead to earlier saturation or reduced steady-state levels, especially if bottlenecks form. These dynamics carry important implications for understanding not only epidemic outbreaks but also the spread of social behaviours, norms, or innovations within networked populations.

4.2 Social Influence Diffusion Process on Hypergraphs

Social influence refers to the capacity of individuals to affect others' beliefs, attitudes, or behaviours, often through mechanisms such as imitation, persuasion, or peer pressure [32]. With the advent of digital platforms and online social media, word-of-mouth diffusion, a form of peer-to-peer information spread, has become a dominant mechanism of influence, widely exploited in domains such as viral marketing [79] or recommender systems [137]. The core computational challenge in this context is known as Social Influence Maximization (SIM): identifying a small seed set of influential individuals in a network whose activation leads to maximal spread of influence [82].

Traditionally, SIM and its variants have been studied on graphs where nodes represent individuals and edges represent pairwise interactions. One prominent formulation is the Target Set Selection (TSS) problem [71], which seeks the smallest set of initially activated nodes that can eventually activate the entire network under a diffusion model, typically the linear threshold (LT) model [53, 116]. In this model, each node has a threshold and becomes active when the sum of influence weights from active neighbours exceeds it.

In reality, however, social interactions are often group-based rather than pairwise. Examples include participation in online communities, co-authorships, or collaborative projects. Hypergraphs, where hyperedges can connect any number of nodes, offer a natural and lossless representation of these many-to-many relationships. Extensions of SIM and TSS to hypergraphs are therefore crucial for modelling higher-order social influence.

One such extension is the Target Set Selection on Hypergraphs (TSSH) problem introduced in [10], where influence diffuses not only from node to node but through the entire structure of the hypergraph, accounting for both node and hyperedge thresholds. The diffusion process is defined over the incidence graph of the hypergraph, a bipartite representation connecting nodes and hyperedges. At each discrete step, influence alternates between nodes and hyperedges: nodes activate hyperedges if enough of their members are active, and hyperedges, in turn, activate nodes based on their thresholds.

This alternating, bipartite mechanism allows for modelling realistic social scenarios where an individual may be influenced by the collective stance of a group. The TSSH problem seeks the smallest seed set of nodes S such that, through this process, all nodes eventually become influenced. Several greedy heuristics have been developed to approximate TSSH [9], and empirical results show the model's effectiveness on both synthetic and real-world data.

An open question, however, concerns how structural properties of the hypergraph, particularly the correlation between node degrees and hyperedge sizes, impact the required size of the seed set. Since the diffusion operates on the incidence bipartite graph, a positive correlation between node degree and hyperedge size corresponds to positive assortativity in the bipartite structure. Prior work in network epidemiology shows that positive degree–degree correlations can slow diffusion [104, 96]. Hence, we hypothesize that in hypergraphs with positive node–hyperedge size correlation, the TSSH seed set must be larger to achieve full diffusion. Conversely, in negatively correlated structures, fewer influential nodes may suffice to trigger widespread influence. This insight underlines the importance of measuring and understanding degree–size correlations in empirical hypergraphs.

4.3 Cooperation in the Public Goods Game on Hypergraphs

The public goods game is a model in evolutionary game theory that extends the prisoner's dilemma to group interactions [117, 107]. In its simplest form, each of the G players in a group decides whether to contribute a token (cooperate) or not (defect). The total contributions are then multiplied by a synergy factor $r > 1$ and equally distributed among all group members, regardless of their strategy. If N_c players cooperate, then a cooperator receives $\pi_C = rN_c/G - t$ (paying cost t), and a defector receives $\pi_D = rN_c/G$. Thus, the game captures the essential social dilemma: while defection is individually rational, collective cooperation yields the highest group payoff.

Multiplayer games such as the public goods game are typically studied on classical graphs by randomly selecting an edge, i.e., a pair of neighbouring players (i, j). Each node participates in $k + 1$ games: one as the focal player and k as a co-player in the games initiated by its neighbours, where k denotes the node's degree. After payoffs are computed, player i may adopt the strategy of player j with a probability that depends on the relative difference in their payoffs. Two cost-allocation schemes are commonly considered: (i) fixed cost per game ($t = \text{const}$), in which the total cost incurred by a player increases linearly with degree, and (ii) fixed cost per individual ($t \propto \frac{1}{k+1}$), which distributes the cost evenly across all games, keeping the total cost constant regardless of degree.

Simulations on lattices showed that cooperation is sustained for the synergy factor below the critical condition, i.e. $r < G$ [122, 25]. Additionally, scale-free networks networks enhance cooperation more under the fixed cost per individual setup (rather than fixed cost per game) due to the disproportionately high payoffs collected by high-degree nodes [115]. This example is particularly relevant to our study, as it highlights the crucial role of degree–degree correlations in shaping cooperative dynamics. Specifically, it is shown that positive degree correlations tend to suppress cooperation [114].

A more realistic formulation by playing the public goods game on bipartite networks that explicitly represent the group structure was introduced by [51]. In their model, one node set represents individuals and the other set represents groups (e.g., papers in scientific collaborations). They showed that cooperation is systematically enhanced when the game is implemented on the bipartite network rather than its one-mode projection, for both cost schemes. Moreover, under the fixed cost per individual setup, cooperation tends to increase among players who are involved in fewer but tighter groups, i.e., when there is a positive correlation between hyperedge size and node degree. This result is particularly relevant to our study, as it highlights the potential influence of the observed correlation on the emergence of cooperative behaviour. Interestingly, increasing group size generally reduces cooperation levels in both formulations.

Building on this foundation, [4] developed a public goods game directly on hypergraphs, representing groups as hyperlinks and individuals as nodes. In a newly introduced formulation, each individual i accumulates payoffs from all hyperlinks they participate in and imitates the strategy of its best performing neighbour j with a probability which depends on $\pi_j - \pi_i$. They studied this model on both uniform and heterogeneous random hypergraphs, as well as empirical ones. Key findings include that larger group sizes promote cooperation in harsh conditions (low r), and that heterogeneous hypergraphs allow for nuanced control over the critical threshold for cooperation and the speed of convergence to steady states.

These findings prompt two hypotheses that directly relate to our empirical observations. First, the positive correlation between hyperedge size and node degree observed in many real-world hypergraphs of physical contact, see Figure 3b, is likely to promote cooperation, particularly under the fixed cost per individual scheme, as suggested by [51]. Intuitively, individuals who participate in many groups (i.e., high-degree nodes) also tend to be part of larger groups (i.e., large hyperedges), enabling them to access greater pooled benefits at relatively lower individual cost. Second, the seemingly contradictory results regarding the effect of group size on cooperation between [51] and [4], the former reporting decreased cooperation with increasing group size, and the latter showing enhanced cooperation under similar conditions, could potentially be reconciled by accounting for the correlation between group size and node degree. Future theoretical work and simulations that allow for independent control of both node degree and group size distributions, as well as their correlation, may clarify whether it is heterogeneity alone or its alignment across structural features that drives cooperation in higher-order systems.

5 Conclusions and Further Research

The structural relationship between hyperedge size and node degree is a fundamental yet understudied property of hypergraphs. Understanding whether and how these two features co-vary is essential not only for characterizing

real-world hypergraph data but also for designing more realistic generative models and interpreting dynamical processes that unfold over such structures. This section summarizes our key findings and offers guidance for future research and tool development aimed at incorporating these insights into empirical analysis and synthetic graph construction.

Conclusions This study systematically investigated the empirical relationship between hyperedge size and node degree across 36 real-world hypergraphs, using linear models, non-parametric correlations, and flexible Generalized Additive Models (GAMs). We classified each hypergraph according to the complexity and direction of its underlying trend: linear, monotonic, non-monotonic, or absent, using a sequence of statistical tests. Our results reveal that such relationships are not only widespread but also often complex: nearly 42% of hypergraphs exhibit non-monotonic patterns, and only a small minority (3 out of 36) show no significant dependency. These findings directly challenge common assumptions of structural independence in generative models and downstream hypergraph applications.

An important practical takeaway for data analysts and hypergraph modellers is the critical role of data preprocessing. Among the three examined strategies: node-centric, edge-centric, and bipartite representation, we recommend the bipartite projection as the default. It consistently exhibited the lowest within-segment variability and highest η^2 effect sizes.

Moreover, we found that the sign of classical correlation coefficients, especially Pearson, aligns well with the direction of the dominant trend estimated by GAMs, particularly when statistical significance is evaluated at conventional levels ($\alpha = 0.05$). However, we also showed that Spearman and Kendall coefficients can misrepresent global trends in the presence of multimodal or U-shaped relationships, often due to data density and structure-induced dependencies.

Importantly, our analysis uncovered strong consistency in the direction of monotonicity within semantically defined hypergraph categories, such as Physical Contact, User-Thread, or Political. These patterns were not only statistically robust but also interpretable based on the domain-specific meaning of nodes and hyperedges. For example, a positive relationship is expected in contact networks where individuals participating in larger groups naturally accumulate more contacts, while negative correlations in political networks align with institutional constraints or specialization effects.

Although our primary focus was on empirical hypergraphs, we also applied our analytical and statistical procedures to a well-known class of synthetic hypergraphs, namely, the h–ABCD synthetic model [69]. While detailed results are omitted here, it can be shown analytically that h–ABCD hypergraphs exhibit no structural relationship between hyperedge size and node degree. Consistent with this, our full statistical pipeline, including Pearson, Spearman, and Kendall correlations, as well as model-based classification, correctly identified these cases as exhibiting no relationship. This serves as an important validation of our methodology, demonstrating that the proposed procedure does not spuriously detect structure where none exists.

Taken together, this work highlights that the relationship between hyperedge size and node degree is not a marginal feature but a structurally and semantically meaningful property of empirical hypergraphs. It should therefore be measured carefully, interpreted in context, and considered when designing models, generating synthetic data, or studying dynamics on hypergraph-structured systems.

Further Research Directions A direct application of this work is to inform the next generation of generative models for synthetic hypergraphs, which typically neglect the rich, empirically observed relationships between hyperedge size and node degree. Many existing generative models either ignore these correlations entirely or assume overly simplistic linear dependencies. One promising approach is to leverage algorithms designed for degree–degree correlations in bipartite graphs, such as the method proposed by Xulvi-Brunet and Sokolov [136, 68], and adapt them for hypergraph construction via bipartite representations. Since Pearson correlations between hyperedge size and node degree translate directly to degree–degree correlations in bipartite graphs, steering such dependencies during bipartite construction enables generation of synthetic hypergraphs with empirically realistic structure.

However, as our findings show, linear measures like Pearson often fail to capture the full complexity of these relationships, many of which are non-linear and non-monotonic. A promising direction is to model this complexity explicitly by introducing dependencies into the data-generation process. For instance, by duplicating hyperedges according to simulated weights drawn from power-law distributions, one could mimic the empirical fluctuations and multimodal structures observed in real hypergraphs. These patterns often reflect structural

dependencies and repeated groupings (e.g., recurring meetings or standard mailing lists), and embedding such mechanisms into generative models could substantially increase their realism.

Another critical avenue for future research is to examine how the strength and form of the correlation between hyperedge size and node degree affect dynamical processes on hypergraphs. A particularly relevant testbed for this inquiry is the Target Set Selection problem, discussed in Subsection 4.2, which seeks the smallest set of initially activated nodes that can trigger full diffusion under a specified activation rule. Based on the discussion in Section 4, we hypothesize that stronger positive correlations may hinder diffusion by concentrating activation capacity in a subset of high-degree nodes participating in large hyperedges, potentially requiring a larger seed set to achieve full coverage. However, testing this hypothesis empirically is complicated by the presence of confounding structural features that co-vary with degree–size correlation across real datasets. To overcome this, a promising direction is the development of generative hypergraph models that can vary the node degree–hyperedge size correlation in a controlled manner while holding other properties constant. Such models would enable rigorous ceteris paribus experiments to isolate the causal effects of this correlation on influence diffusion and related dynamics.

Finally, the statistical procedure developed in this study, implemented in R, provides an end-to-end framework for assessing the relationship type in a collection of empirical datasets. Beyond hypergraph science, such a tool has potential applications in many disciplines that require robust classification of structural dependencies in large, noisy datasets. Future research could extend the tool in several directions:

- enhancing statistical robustness by accounting for correlated or clustered observations (e.g., repeated measures),
- enabling richer feature engineering, such as logarithmic, Box–Cox, or user-defined transformations,
- packaging the method into accessible software libraries in R, Python, or Julia, and
- disseminating the software and methodology through publications, tutorials, and conference workshops to promote its cross-disciplinary adoption.

By combining rigorous empirical analysis, domain-informed interpretation, and practical tooling, this study lays the foundation for future theoretical, algorithmic, and applied research into the structural dependencies that underlie real-world hypergraph data.

Acknowledgements

Mateusz Zawisza gratefully acknowledges Jordan Barrett (Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada) for sharing initial versions of code and datasets, as well as for valuable discussions.

He also thanks the participants of the Modelling and Mining Complex Networks as Hypergraphs Workshop in 2024 at Toronto Metropolitan University for their questions and feedback, as well as seminar participants at Toronto Metropolitan University, Department of Mathematics, in particular, François Théberge (Tutte Institute for Mathematics and Computing), Austin Eide, Lourens Touwen, and Ash Dehghan, for insightful discussions. Additional thanks go to the seminar audience at the University of Cassino and Southern Lazio, Department of Economics and Law, for their comments and engagement.

CRediT authorship contribution statement

Bogumił Kamiński: Conceptualization, Methodology, Project administration, Supervision, Writing: review & editing. **Paweł Prałat**: Conceptualization, Methodology, Project administration, Supervision, Writing: review & editing. **Aleksander Wojnarowicz**: Data curation, Formal analysis, Investigation, Software, Validation, Roles/Writing: original draft. **Mateusz Zawisza**: Conceptualization, Formal analysis, Methodology, Investigation, Software, Validation, Visualization, Writing: original draft, Writing: review & editing.

Declaration of Interest

The authors declare no competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI) in order to assist with language editing, code documentation, and the formulation of section summaries. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Sameer Agarwal, Kristin Branson, and Serge Belongie. Higher order learning with graphs. In *Proceedings of the 23rd international conference on Machine learning*, pages 17–24, 2006.

[2] Sinan G Aksoy, Cliff Joslyn, Carlos Ortiz Marrero, Brenda Praggastis, and Emilie Purvine. Hypernetwork science via high-order hypergraph walks. *EPJ Data Science*, 9(1):16, 2020.

[3] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

[4] Unai Alvarez-Rodriguez, Federico Battiston, Guilherme Ferraz de Arruda, Yamir Moreno, Matjaž Perc, and Vito Latora. Evolutionary dynamics of higher-order interactions in social networks. *Nature Human Behaviour*, 5(5):586–595, 2021.

[5] Ilya Amburg, Nate Veldt, and Austin Benson. Clustering in graphs and hypergraphs with categorical edge labels. In *Proceedings of the web conference 2020*, pages 706–717, 2020.

[6] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1991.

[7] Miroslav Andjelković, Bosiljka Tadić, Slobodan Maletić, and Milan Rajković. Hierarchical sequencing of online social graphs. *Physica A: Statistical Mechanics and its Applications*, 436:582–595, 2015.

[8] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2009.

[9] Alessia Antelmi, Gennaro Cordasco, Carmine Spagnuolo, and Przemysław Szufel. Information diffusion in complex networks: a model based on hypergraphs and its analysis. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 36–51. Springer, 2020.

[10] Alessia Antelmi, Gennaro Cordasco, Carmine Spagnuolo, and Przemysław Szufel. Social influence maximization in hypergraphs. *Entropy*, 23(7):796, 2021.

[11] Alessia Antelmi, Daniele De Vinco, and Carmine Spagnuolo. Hypergraphrepository: a community-driven and interactive hypernetwork data collection. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 159–173. Springer, 2024.

[12] Alessia Antelmi, Daniele De Vinco, and Carmine Spagnuolo. Hypergraphrepository: A community-driven and interactive hypernetwork data collection. In Megan Dewar, Bogumił Kamiński, Daniel Kaszyński, Łukasz Kraiński, Paweł Prałat, François Théberge, and Małgorzata Wrzosek, editors, *Modelling and Mining Networks*, pages 159–173, Cham, 2024. Springer Nature Switzerland.

[13] Frank Ball, Tom Britton, Thomas House, Valerie Isham, Denis Mollison, Lorenzo Pellis, and Gianpaolo Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, 2015.

[14] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016. Chapter 4.

[15] J. Barrett, P. Prałat, A. Smith, and F. Theberge. Counting simplicial pairs in hypergraphs. *Journal of Complex Networks*, 2025. Accepted for publication, 39 pages.

[16] M. Barthelemy. A class of models for random hypergraphs. *Physical review. E*, 106 6-1:064310, 2022.

[17] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92, 2020.

[18] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, 115(48):E11221–E11230, 2018.

[19] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[20] Claude Berge. *Hypergraphs: combinatorics of finite sets*, volume 45. Elsevier, 1984.

[21] Ágnes Bodó, Gyula Y Katona, and Péter L Simon. Sis epidemic propagation on hypergraphs. *Bulletin of mathematical biology*, 78(4):713–735, 2016.

[22] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182, 1987.

[23] Phillip Bonacich. Simultaneous group and individual centralities. *Social networks*, 13(2):155–168, 1991.

[24] Phillip Bonacich, Annie Cody Holdren, and Michael Johnston. Hyper-edges and multidimensional centrality. *Social networks*, 26(3):189–203, 2004.

[25] Hannelore Brandt, Christoph Hauert, and Karl Sigmund. Punishment and reputation in spatial public goods games. *Proceedings of the royal society of London. Series B: biological sciences*, 270(1519):1099–1104, 2003.

[26] Alain Bretto. *Hypergraph Theory: An Introduction*. Springer, 2013.

[27] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

[28] John M. Chambers and Trevor J. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, CA, 1992.

[29] Ning Chen. On the approximability of influence in social networks. *SIAM Journal on Discrete Mathematics*, 23(3):1400–1415, 2009.

[30] Philip S Chodrow. Configuration models of random hypergraphs. *Journal of Complex Networks*, 8(3):cnaa018, 2020.

[31] Philip S Chodrow, Nate Veldt, and Austin R Benson. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances*, 7(28):eabh1303, 2021.

[32] Robert B Cialdini et al. *Influence: Science and practice*, volume 4. Pearson education Boston, 2009.

[33] Christophe Croux and Catherine Dehon. Influence functions of the spearman and kendall correlation measures. *Statistical methods & applications*, 19:497–515, 2010.

[34] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. Social contagion models on hypergraphs. *Physical Review Research*, 2(2):023032, 2020.

[35] Manh Tuan Do, Se-eun Yoon, Bryan Hooi, and Kijung Shin. Structural patterns and generative models of real-world hypergraphs. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 176–186, 2020.

[36] Patrick Doreian. On the evolution of group and network structure. *Social Networks*, 2(3):235–252, 1979.

[37] Matt Dowle and Arun Srinivasan. data.table: Extension of data.frame. *R package version 1.12.8*, 2019.

[38] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, New York, 2010.

[39] F. Essam, H. El, and S. R. H. Ali. A comparison of the pearson, spearman rank and kendall tau correlation coefficients using quantitative variables. *Asian Journal of Probability and Statistics*, 2022.

[40] Ernesto Estrada and Juan A Rodríguez-Velázquez. Subgraph centrality and clustering in complex hyper-networks. *Physica A: Statistical Mechanics and its Applications*, 364:581–594, 2006.

[41] Katherine Faust. Centrality in affiliation networks. *Social networks*, 19(2):157–191, 1997.

[42] David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, 2009.

[43] John E. Freund and Benjamin M. Perles. *Statistics: A First Course*. Pearson, 9th edition, 2014.

[44] Andrew Gelman. Analysis of variance—why it is more important than ever. *The Annals of Statistics*, 33(1):1–53, 2005.

[45] Mathieu Génois and Alain Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):1–18, 2018.

[46] Mathieu Génois and Alain Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):1–18, 2018.

[47] Mathieu Génois, Christian L Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Science*, 3(3):326–347, 2015.

[48] Gourab Ghoshal and M E J Newman. Random hypergraphs and their applications. *Physical Review E*, 79(6):061109, 2009.

[49] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[50] Kwang-Il Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, 2007.

[51] Jesus Gomez-Gardenes, Miguel Romance, Regino Criado, Daniele Vilone, and Angel Sánchez. Evolutionary games defined at the network mesoscale: The public goods game. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1), 2011.

[52] Peter Gould and Anthony Gatrell. A structural analysis of a game: the liverpool v manchester united cup final of 1977. *Social Networks*, 2(3):253–273, 1979.

[53] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[54] William H. Greene. *Econometric Analysis*. Pearson Education, 7th edition, 2012. Chapter 9.

[55] Jacopo Grilli, György Barabás, Matthew J Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210–213, 2017.

[56] Trevor Hastie and Robert Tibshirani. *Generalized Additive Models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, 1990.

[57] Xie He, Philip S. Chodrow, and Peter J. Mucha. Hypergraph link prediction via hyperedge copying. *ArXiv*, abs/2502.02386, 2025.

[58] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

[59] Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric Statistical Methods*. Wiley, 3rd edition, 2013.

[60] Thomas House and Matt J Keeling. Deterministic epidemic models with explicit household structure. *Mathematical biosciences*, 213(1):29–39, 2008.

[61] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.

[62] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.

[63] Matthew O Jackson et al. *Social and economic networks*, volume 3. Princeton university press Princeton, 2008.

[64] Bogumił Kamiński, Paweł Misiorek, Paweł Prałat, and François Théberge. Modularity based community detection in hypergraphs. *Journal of Complex Networks*, 12(5):cnae041, 2024.

[65] Bogumił Kamiński, Valérie Poulin, Paweł Prałat, Przemysław Szufel, and François Théberge. Clustering via hypergraph modularity. *PloS one*, 14(11):e0224307, 2019.

[66] Bogumił Kamiński, Paweł Prałat, and François Théberge. Community detection algorithm using hypergraph modularity. In *International Conference on Complex Networks and Their Applications*, pages 152–163. Springer, 2020.

[67] Bogumił Kamiński, Paweł Prałat, and François Théberge. Artificial benchmark for community detection (abcd)—fast random graph model with community structure. *Network Science*, 9(2):153–178, 2021.

[68] Bogumil Kaminski, Pawel Prałat, and François Théberge. *Mining complex networks*. Chapman and Hall/CRC, 2021.

[69] Bogumił Kamiński, Paweł Prałat, and François Théberge. Hypergraph artificial benchmark for community detection (h–abcd). *Journal of Complex Networks*, 11(4):cnad028, 2023.

[70] Brian Karrer and Mark EJ Newman. Random graphs containing arbitrary distributions of subgraphs. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(6):066118, 2010.

[71] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[72] Maurice G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.

[73] Peter Kennedy. A guide to econometrics. *Wiley-Blackwell*, 2008.

[74] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[75] Jung-Ho Kim and K. Goh. Higher-order components dictate higher-order contagion dynamics in hypergraphs. *Physical review letters*, 132 8:087401, 2022.

[76] Bryan Klimt and Yiming Yang. Introducing the enron corpus. In *CEAS*, volume 2004, 2004.

[77] Nicholas W Landry and Juan G Restrepo. The effect of heterogeneity on hypergraph contagion models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10), 2020.

[78] Geon Lee, Fanchen Bu, Tina Eliassi-Rad, and Kijung Shin. A survey on hypergraph mining: Patterns, tools, and generators. *ACM Computing Surveys*, 57(8):1–36, 2025.

[79] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5–es, 2007.

[80] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

[81] N. Levshina. Relationships between two quantitative variables: Correlation analysis with elements of linear regression modelling. *Quantitative Methods in Linguistics*, 2015.

[82] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872, 2018.

[83] B. Luo. eventernote-places. Zenodo, May 2024.

[84] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[85] Stan Matwin, Aristides Milios, Paweł Prałat, Amilcar Soares, and François Théberge. *Generative Methods for Social Media Analysis*. Springer, 2023.

[86] J Miller McPherson. Hypernetwork sampling: Duality and differentiation among voluntary organizations. *Social Networks*, 3(4):225–249, 1982.

[87] Joel C Miller. Percolation and epidemics in random clustered networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 80(2):020901, 2009.

[88] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[89] Staša Milojević. Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences*, 111(11):3984–3989, 2014.

[90] Danielle Navarro. *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6)*. University of New South Wales, Sydney, Australia, 2015. R package version 0.5.1.

[91] Saket Navlakha and Carl Kingsford. The power of protein interaction networks for associating genes with diseases. *Bioinformatics*, 26(8):1057–1063, 2010.

[92] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[93] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351, 2005.

[94] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(Suppl 1):2566–2572, 2002.

[95] Mark Newman. *Networks*. Oxford university press, 2018.

[96] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.

[97] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[98] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

[99] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[100] Mark EJ Newman and Juyong Park. Why social networks are different from other types of networks. *Physical review E*, 68(3):036122, 2003.

[101] David JP O'Sullivan, Gary J O'Keeffe, Peter G Fennell, and James P Gleeson. Mathematical modeling of complex contagion on clustered networks. *Frontiers in Physics*, 3:71, 2015.

[102] Laura Ozella, Daniela Paolotti, Guilherme Lichand, Jorge P Rodríguez, Simon Haenni, John Phuka, Onicio B Leal-Neto, and Ciro Cattuto. Using wearable proximity sensors to characterize social contact patterns in a village of rural malawi. *EPJ Data Science*, 10(1):46, 2021.

[103] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):236, 2019.

[104] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

[105] Alice Patania, Giovanni Petri, and Franco Vaccarino. The shape of collaborations. *EPJ Data Science*, 6(1):18, 2017.

[106] Karl Pearson. Mathematical contributions to the theory of evolution. iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A*, 187:253–318, 1896.

[107] Matjaž Perc, Jesús Gómez-Gardenes, Attila Szolnoki, Luis M Floría, and Yamir Moreno. Evolutionary dynamics of group interactions on structured populations: a review. *Journal of the royal society interface*, 10(80):20120997, 2013.

[108] Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I Furlong. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic acids research*, 48(D1):D845–D855, 2020.

[109] Alexis Pister and Marc Barthelemy. Stochastic block hypergraph model. *Physical Review E*, 110(3):034312, 2024.

[110] N. Pya and S.N. Wood. Shape constrained additive models. *Statistics and Computing*, 25(3):543–559, 2015.

[111] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.

[112] John T. E. Richardson. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2):135–147, 2011.

[113] Martin Ritchie, Luc Berthouze, Thomas House, and Istvan Z Kiss. Higher-order structure and epidemic dynamics in clustered networks. *Journal of Theoretical Biology*, 348:21–32, 2014.

[114] Zhihai Rong and Zhi-Xi Wu. Effect of the degree correlation in public goods game on scale-free networks. *Europhysics letters*, 87(3):30001, 2009.

[115] Francisco C Santos and Jorge M Pacheco. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical review letters*, 95(9):098104, 2005.

[116] Thomas C Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[117] Karl Sigmund. *The calculus of selfishness*. Princeton University Press, 2010.

[118] Charles Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.

[119] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, et al. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9(1):87, 2011.

[120] Hanlin Sun and G. Bianconi. Higher-order percolation processes on multiplex hypergraphs. *Physical review. E*, 104 3-1:034306, 2021.

[121] Qi Suo, Jin-Li Guo, and Ai-Zhong Shen. Information spreading dynamics in hypernetworks. *Physica A: Statistical Mechanics and its Applications*, 495:475–487, 2018.

[122] György Szabó and Christoph Hauert. Phase transitions and volunteering in spatial public goods games. *Physical review letters*, 89(11):118101, 2002.

[123] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

[124] Edwin van den Heuvel and Zhuozhao Zhan. Myths about linear and monotonic associations: Pearson'sr, spearman's ρ, and kendall's τ. *The American Statistician*, 76(1):44–52, 2022.

[125] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS one*, 8(9):e73970, 2013.

[126] Hal R Varian. *Microeconomic analysis*, volume 3. Norton New York, 1992.

[127] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.

[128] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Cambridge university press, 1994.

[129] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[130] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[131] Hadley Wickham, Romain Francois, Lionel Henry, and Kirill Müller. dplyr: A grammar of data manipulation. *R package version 1.0.0*, 2020.

[132] S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.

[133] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2nd edition, 2010.

[134] W. Xu, C. Chang, Y. S. Hung, and S. K. Kwan. Order statistics correlation coefficient as a novel association measurement with applications to biosignal analysis. *IEEE Transactions on Biomedical Engineering*, 2007.

[135] Weichao Xu, Yunhe Hou, YS Hung, and Yuexian Zou. A comparative analysis of spearman's rho and kendall's tau in normal and contaminated normal models. *Signal Processing*, 93(1):261–276, 2013.

[136] Ramón Xulvi-Brunet and Igor M Sokolov. Changing correlations in networks: assortativity and dissortativity. *Acta Physica Polonica B*, 36(5):1431–1455, 2005.

[137] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation: a generative model approach. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 671–680, 2012.

[138] Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, 2017.

[139] Yuanzhao Zhang, M. Lucas, and F. Battiston. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. *Nature Communications*, 14, 2022.

A Appendix

A.1 Empirical Hypergraph Datasets: Domains and Descriptive Statistics

The dataset analyzed in this study consists of 36 empirical hypergraphs drawn from a wide variety of domains, encompassing both physical and digital social networks, biological networks and political structures. This breadth includes physical contact networks (e.g., `hospital-lyon`, `contact-primary-school`), online user interactions (e.g., `threads-math-sx`, `tags-ask-ubuntu`), institutional affiliations (e.g., `house-committees`, `senate-bills`),

and domain-specific scientific data such as drug composition (`NDC-substances`) or disease-gene associations (`diseasome`, `disgenenet`). All 36 hypergraphs are organized into semantically coherent and internally homogeneous segments, such as User-Answer, Physical Contact, Email, and others. The complete list of datasets, along with their assigned segments and detailed interpretations of both nodes and hyperedges, is provided in Table 5. This interpretability is critical, as it allows us to meaningfully analyze and interpret the relationship between node degree and hyperedge size. The node and hyperedge semantics ensure that our calculated correlations are not just statistical artifacts but reflect domain-relevant structural patterns. All datasets are publicly available through their respective sources, and we additionally host them in our GitHub repository: `https://github.com/AleksanderWWW/hypergraph-properties`.

Hypergraph name	**Segment**	**Node interpretation**	**Hyperedge interpretation**
algebra [11]	User-Answer	users of mathoverflow.net	users who answered a particular type of question about algebra within a month
amazon [11]	Product-Category	products reviewed by users on Amazon	groups of similar items
contact-high-school [11]	Physical Contact	people at a high school	interactions at a resolution of 20 seconds
contact-primary-school [11]	Physical Contact	people at a primary school	interactions at a resolution of 20 seconds
dblp [11]	Part-Whole	DBLP paper authors	documents published between January and May 2017
diseasome [50]	Diseases and Gene	diseases	genes associated with diseases
disgenenet [108]	Diseases and Gene	genes associated with diseases	diseases
email-enron [11]	Email	email addresses at Enron	sender and all recipients of the email
email-eu [18] [138] [80]	Email	email addresses at a European research institution	sender and all receivers grouped by timestamp
email-W3C [11]	Email	email addresses on W3C mailing lists	set of email addresses on the same email
geometry [12]	User-Answer	users of mathoverflow.net	sets of users who answered a certain question category about geometry
got [11]	Person-Place	GoT characters	GoT scenes linking characters appearing in the same scene
hospital-lyon [125]	Physical Contact	patients and health-care workers in a hospital ward in Lyon, France	group interactions
music-blues-reviews [11]	User-Review	Amazon users	users who reviewed a blues music product within a month
nba [12]	Part-Whole	NBA players	players involved in a match up to 2012
NDC-classes [12]	Drugs	class labels applied to drugs	drugs
NDC-substances [12]	Drugs	substances making up a drug	drugs
restaurant-reviews [11]	User-Review	Yelp users	users who reviewed restaurants in Madison, WI within a month
tags-ask-ubuntu [18]	Tag-Question	tags	sets of tags applied to questions on askubuntu.com
tags-math-sx [18]	Tag-Question	tags	sets of tags applied to questions on math.stackexchange.com
threads-ask-ubuntu [11]	User-Thread	users on askubuntu.com	users participating in a thread lasting $\leqslant$ 24 hours
threads-math-sx [11]	User-Thread	users on math.stackexchange.com	users participating in a thread lasting $\leqslant$ 24 hours
twitter [11]	User-Thread		
vegas-bars-reviews [11]	User-Thread	Yelp users	users who reviewed the same bar in Las Vegas within a month
evernote-places [83]	Person-Place	artists or artist groups	places where idol/voice actor events took place
house-bills (House) [31]	Political	political affiliation	bill cosponsorship in the US House of Representatives
house-committees (House) [31]	Political	political affiliation	committee membership in the US House of Representatives
Hypertext-conference [61]	Participant-Conference	conference attendees	face-to-face interactions over 2.5 days
InVS13 [47], InVS15 [45], science-gallery [62]	Physical Contact	participants with sensors	snapshots of groups present at specific times
kaggle-whats-cooking [5]	Part-Whole	ingredients	dishes comprising those ingredients

Malawi-village [102]	Physical Contact	individuals living in a village	interactions in a rural Malawi village
house-bills (Senate) [31]	Political	political affiliation	bill cosponsorship in the US Senate
house-committees (Senate) [31]	Political	political affiliation	committee membership in the US Senate
SFHH-conference [46, 119, 27]	Participant-Conference	conference attendees	face-to-face contacts every 20 seconds

Table 5: List of hypergraph datasets with their sources, assigned semantic segments, and interpretations of nodes and hyperedges.

Beyond their domain diversity, the hypergraphs analyzed in this study exhibit substantial variation in structural characteristics, particularly in the size and distribution of node degrees and hyperedge sizes. This variability is essential for assessing the correlation and relationship between these two quantities. In contrast, classical graphs with binary edges lack such variability, as edge size is fixed at 2. Consequently, in standard graphs, the notion of a relationship between edge size and node degree is either undefined or trivially zero. Tables 6 and 7 provide detailed summary statistics of node degrees and hyperedge sizes, including the number of observations (`n`), average value (`avg`), standard deviation (`sd`), skewness (`skew`), and observed range (`range`).

name	**n**	**avg**	**sd**	**range**	**skew**
algebra	423	19.53	34.01	1–375	5.03
amazon	4989	1.02	0.18	1–4	8.98
contact-high-school	327	55.63	27.06	2–148	0.48
contact-primary-school	242	126.98	55.15	28–261	0.31
dblp	71116	1.24	0.80	1–25	7.33
diseasome	516	2.15	2.15	1–22	3.84
disgenenet	12368	9.09	16.87	1–377	6.67
email-enron	143	32.33	24.26	2–118	1.22
email-eu	1005	88.96	116.35	1–918	2.52
email-W3C	5601	2.39	11.43	1–282	17.23
geometry	580	21.53	36.26	1–260	3.72
got	577	20.99	59.79	1–632	5.83
hospital-lyon	75	59.03	48.99	6–205	1.22
music-blues-reviews	1106	9.49	10.72	1–127	3.25
nba	2191	293.95	308.26	1–1476	1.12
NDC-classes	1161	134.53	402.96	1–5357	7.88
NDC-substances	5311	10.08	35.11	1–579	8.85
restaurant-reviews	565	8.14	7.22	1–59	3.51
tags-ask-ubuntu	3029	164.84	606.11	1–12931	10.31
tags-math-sx	1629	364.10	1039.61	1–13950	6.80
threads-ask-ubuntu	125602	2.76	20.78	1–2332	51.55
threads-math-sx	176445	9.13	92.98	1–12511	59.98
twitter	22964	2.21	4.61	1–266	18.03
vegas-bars-reviews	1234	9.62	7.37	1–147	7.85
eventernote-places	71890	9.92	25.02	1–421	5.70
house-bills	1494	835.79	815.06	1–6220	2.10
house-committees	1290	9.18	7.09	1–44	1.16
Hypertext-conference	113	345.56	304.16	2–1446	1.52
InVS13	92	210.65	193.14	5–1089	2.13
InVS15	217	691.01	488.80	1–3192	1.50
kaggle-whats-cooking	6714	63.78	388.31	1–18048	22.99
Malawi-village	86	2338.01	1780.06	12–7636	0.63
Science-Gallery	10972	65.41	56.06	1–486	1.66
senate-bills	294	789.62	640.09	1–3514	1.31
senate-committees	282	19.18	14.85	1–63	0.85

name	n	avg	sd	range	skew
SFHH-conference	403	289.42	311.67	2–1960	2.64

Table 6: Node degree distribution

name	n	avg	sd	skew	range
algebra	1268	6.52	6.58	6.32	2–107
amazon	1176	4.35	2.27	-0.71	1–6
contact-high-school	7818	2.33	0.53	1.38	2–5
contact-primary-school	12704	2.42	0.55	0.88	2–5
dblp	25624	3.45	2.12	5.24	1–69
diseasome	481	2.31	1.50	1.95	1–11
disgenenet	2069	54.36	169.31	7.38	1–2453
email-enron	1514	3.05	2.29	5.92	1–37
email-eu	25148	3.56	3.40	4.51	1–40
email-W3C	6000	2.23	0.99	10.11	2–23
geometry	1193	10.47	15.65	4.11	2–230
got	4165	2.91	2.35	2.33	0–24
hospital-lyon	1824	2.43	0.56	0.92	2–5
music-blues-reviews	694	15.13	14.71	1.81	2–83
nba	31686	20.33	1.89	0.18	14–28
NDC-classes	49724	3.14	2.10	2.66	1–24
NDC-substances	9906	5.40	5.78	1.49	1–25
restaurant-reviews	601	7.66	7.28	1.90	2–43
tags-ask-ubuntu	147222	3.39	1.03	0.04	1–5
tags-math-sx	170476	3.48	0.97	0.02	1–5
threads-ask-ubuntu	192947	1.80	0.80	1.29	1–14
threads-math-sx	719792	2.24	1.04	1.48	1–21
twitter	4065	12.51	16.90	3.40	1–207
vegas-bars-reviews	1194	9.94	13.82	2.65	2–73
eventernote-places	19033	37.48	185.35	12.84	0–6420
house-bills	60987	20.47	33.83	4.27	2–399
house-committees	341	34.73	21.39	-0.03	1–81
Hypertext-conference	19036	2.05	0.24	5.51	2–6
InVS13	9644	2.01	0.10	10.72	2–4
InVS15	73822	2.03	0.18	5.50	2–4
kaggle-whats-cooking	39774	10.77	4.43	0.86	1–65
Malawi-village	99942	2.01	0.11	9.18	2–4
Science-Gallery	338765	2.12	0.35	3.13	2–5
senate-bills	29157	7.96	10.27	3.23	2–99
senate-committees	315	17.17	6.79	-0.53	4–31
SFHH-conference	54305	2.15	0.50	5.34	2–9

Table 7: Hyperedge size distribution

This structural heterogeneity, combined with interpretability and semantic clarity, makes our dataset particularly suitable for a robust investigation of correlations between hyperedge size and node degree. The richness of the dataset ensures that the findings are not limited to a single domain or structure, while the semantic interpretability allows us to validate the significance of results in real-world terms. Altogether, this provides a strong foundation for generalizable and meaningful analysis.

A.2 Computational Implementation and Complexity

In this section, we outline the implementation details behind the analysis hypergraph properties employed in this paper. The key components involve:

- data ingestion from various source formats,
- construction of efficient data structures for hypergraph representation,
- tools used to optimize computations on hypergraphs.

The data used for the process was obtained from different sources and was therefore stored in diverse file formats. Those included JSON (JavaScript Object Notation), HGF (Hypergraph format), XGI (CompleX Group Interactions) and plain text (.txt). A different strategy was necessary for each. Additionally, for all but for JSON the code for line-by-line reading had to be crafted (Python's built-in `json` library handled JSON files without the need of custom reading and parsing implementations).

Based on the loaded file contents, an instance of a sparse matrix was created. Choosing this type of data structure allowed for efficient storage of large hypergraphs (dense matrices would quickly drain memory resources and cause crashes in the processing pipeline), while remaining on par with the representation used in literature. The latter significantly simplified the translation from theory into software implementation.

For the JSON files, it was possible to use the `scipy.sparse.coo_array` object and construct the entire sparse matrix in one function call. In the other cases, an incremental line-by-line approach was needed. For synthetic hypergraphs generated by the h–ABCD synthetic benchmark [69], `scipy.sparse.lil_array` was the most efficient type for construction, whereas for empirical hypergraphs, `scipy.sparse.dok_array` performed best. Based on this observation, we recommend further research into a potential relationship between a hypergraph type, and the optimal sparse matrix type for incremental construction. In all cases, the constructed object was converted in the `scipy.sparse.csr_array` type, as it is best suited for fast data indexing, crucial to the calculations performed in the following parts of the process.

The achieved space complexity was $O(nnz) + O(n)$, where nnz is the number of non-zero elements and n is the number of rows in the matrix. This is more efficient for hypergraphs, where the nnz will typically be substantially smaller than the dense matrices' $O(n \cdot m)$ complexity (m being the number of columns).

Upon receiving the object representing a hypergraph, the downstream tasks in the pipeline utilized `numpy`'s array and `scipy`'s sparse array methods optimized for fast vector computations, to calculate correlations and descriptive statistics of the data. Initially, `numba` was employed in hopes of taking advantage of the JIT (Just-in-time) compilation. This approach, however, yielded no significant improvements in the processing speed, while increasing the complexity of the implementation details.

Predominant operations in the process of correlation computation were summing over rows, summing over columns and indexing non-zero elements of the CSR matrix. The first two operations are of $O(nnz)$ time complexity, and the indexing of $O(1)$. Those characteristics allowed for efficient data processing even for large hypergraph files.

All statistical analyses, including correlation computations, model fitting, and figure generation, were performed using the R programming language [111]. Correlation measures such as Pearson's r, Spearman's ρ, and Kendall's τ were calculated using base R functions, while statistical modeling was conducted using Generalized Additive Models (GAMs) and shape-constrained additive models (SCAMs). Specifically, unrestricted GAMs were fitted using the `mgcv` package [132], and monotonic (increasing or decreasing) GAMs were implemented with the `scam` package [110], which extends `mgcv` to support monotonicity constraints.

Model comparisons were carried out using ANOVA F-tests from base R functions [28]. For data wrangling and summarization, we employed the `dplyr` [131] and `data.table` [37] packages. Figures were produced using `ggplot2` [130], with `ggrepel` for improved label placement and `ggpubr` for consistent theming. Supplementary LaTeX-ready tables were generated using `xtable`, and eta-squared (η^2) statistics were calculated with the `lsr` package [90]. Altogether, the R ecosystem provided a flexible and reproducible framework for executing the statistical pipeline described in this study.

All reproducible code used in this study, including both Python and R scripts for data processing, statistical analysis, and figure generation, is available in the public repository: `https://github.com/AleksanderWWW/hypergraph-properties`.

A.3 Correlation Measures for Hyperedge Size and Node Degree

To characterise the statistical association between hyperedge size and node degree, we employ three standard correlation measures: Pearson's r, Spearman's ρ, and Kendall's τ. Each captures a different notion of dependence and responds differently to nonlinearity, outliers, and the shape of the relationship. In this subsection, we briefly define each measure, discuss their strengths and limitations, and outline when their use is most appropriate.

Pearson Correlation Pearson's correlation coefficient r quantifies the strength and direction of a linear relationship between two continuous variables: x and y [106]. It is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

Pearson's r assumes that both variables are linearly related. It is sensitive to outliers and may be misleading in the presence of nonlinear or monotonic but non-linear relationships [43]. However, recent work by van den Heuvel and Zhan (2022) challenges the conventional wisdom distinguishing Pearson's r for linear relationships and Spearman's ρ or Kendall's τ for nonlinear monotonic associations. They argue that "Pearson's correlation coefficient should not be ruled out a priori for measuring nonlinear monotonic associations," and further demonstrate via counterexamples that Pearson's r can be preferred over Spearman's ρ and Kendall's τ in testing dependency even when the association is monotonic but nonlinear [124]. Therefore, Pearson's r tends to be more robust and interpretable in contexts where the global trend described by a single summary of direction and strength is desired.

Spearman Correlation Spearman's rank correlation coefficient ρ is a non-parametric measure that assesses the strength of a monotonic relationship between two variables [118]. It is defined as the Pearson correlation between the ranks of the variables:

$$\rho = \frac{\sum_{i=1}^{n}(R(x_i) - \bar{R}_x)(R(y_i) - \bar{R}_y)}{\sqrt{\sum_{i=1}^{n}(R(x_i) - \bar{R}_x)^2}\sqrt{\sum_{i=1}^{n}(R(y_i) - \bar{R}_y)^2}} \tag{3}$$

where $R(x_i)$ and $R(y_i)$ are the ranks of x_i and y_i, respectively. Spearman's ρ is less sensitive to outliers and appropriate when the relationship is monotonic but not necessarily linear [127]. It provides a more flexible summary than Pearson's r but is less interpretable in terms of raw variable scales.

Kendall Correlation Kendall's tau (τ) is another non-parametric measure of monotonic association, based on the number of concordant and discordant pairs in the data [72]. It is defined as:

$$\tau = \frac{C - D}{\binom{n}{2}} \tag{4}$$

where C is the number of concordant pairs and D the number of discordant pairs. A pair of observations (x_i, y_i) and (x_j, y_j) is said to be *concordant* if the ranks of both elements agree in direction: that is, either $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$. Conversely, the pair is *discordant* if the ranks disagree: one variable increases while the other decreases (e.g., $x_i > x_j$ but $y_i < y_j$). Ties may be handled differently in various adjusted versions of Kendall's tau, but in the basic form shown above, tied pairs are typically excluded from C and D.

Kendall's τ is often considered more conservative than Spearman's ρ in the sense that it yields smaller values in absolute magnitude, particularly in small samples. This makes it less prone to detecting spurious associations, but potentially less sensitive to weak monotonic trends [59]. It is particularly well-suited to ordinal data and robust against anomalies [59], but may lack sensitivity to more subtle trends in large-scale, noisy data.

Comparison and Application In the context of our hypergraph analysis, Pearson's r offers a direct assessment of global trends between hyperedge size and node degree and is meaningful when such trends are present. Spearman's ρ and Kendall's τ, on the other hand, are more appropriate when the relationship is suspected to be nonlinear but monotonic, especially common in empirical network data. While Spearman tends to be more sensitive, Kendall is more statistically robust and better suited to small or highly discrete datasets.

Non-parametric correlation coefficients such as Spearman's ρ and Kendall's τ are designed to capture rank-based, monotonic associations and are known for their robustness to outliers and non-normal distributions [124, 39]. These coefficients rely on relative ordering rather than the actual magnitudes of data, making them less sensitive to the shape and sign of complex, nonlinear relationships when compared to Pearson's r, which directly assesses covariance between variable magnitudes [81, 134].

The statistical literature supports three relevant findings regarding the use of Pearson's r, Spearman's ρ, and Kendall's τ for measuring association. First, although non-parametric measures like ρ and τ are valued for their robustness, they exhibit higher variance and bias than Pearson's r even under non-normal, contaminated, or curved distributions [135]. Second, Pearson's r remains the most statistically efficient estimator even when the underlying relationship is approximately linear or near-normal, conditions that are frequently approximated in large empirical datasets [33]. Third, while Spearman's and Kendall's measures are designed to detect monotonicity, they may fail to reflect the dominant global trend direction, especially when the relationship is weakly monotonic or contains local non-monotonic variations [124]. In this paper, we further investigate this third point by directly comparing the signs of Pearson's r, Spearman's ρ, and Kendall's τ to the global direction of association estimated by monotonic Generalized Additive Models (GAMs) [56, 132]. This allows us to evaluate how well each correlation measure captures the overarching trend in the data, even when local fluctuations or curvature are present.

A.4 Supplementary Analyses Referenced in Main Text

A.4.1 Optimal Choice of Hypergraph Preprocessing Strategy and Correlation Coefficients

In our experiments, all three correlation coefficients are evaluated across different data preprocessing strategies, as defined in Section 2.1 and analyzed in Section 3.1. In particular Figure 6 presents the variability of correlation values for six selected combinations of data preprocessing method and correlation coefficient, specifically, all pairings of Pearson and Spearman correlations with the three preprocessing strategies: node-centric, edge-centric, and bipartite representation. It is discussed and referenced in subsection 3.1.2.

Figure 7 presents Pearson correlation values between hyperedge size and node degree for all hypergraphs, computed under three data preprocessing strategies: node-centric, edge-centric, and bipartite representation. The hypergraphs are sorted by decreasing Pearson correlation under bipartite representation. This ordering allows to visually identify clusters of hypergraphs that exhibit similar correlation structure, that is discussed and interpreted in detail in subsection 3.1.4.

The combination of preprocessing method and correlation measure that best aligns with the structural distinctions between semantic hypergraph segments is selected in Section 3.1, based on the η^2 criterion introduced in Section 2.2. In addition, we consider the alignment between the sign of each correlation coefficient and the monotonicity direction inferred from GAM models in Section 3.2. This analysis draws extensively on Table 8, which reports Pearson, Spearman, and Kendall correlations (sorted by decreasing Pearson) for all 36 empirical hypergraphs, along with the monotonicity direction of the fitted monotonic GAM in the bipartite representation. This allows us to compare the sign of each coefficient with the inferred trend direction. A condensed summary of this alignment is presented in Table 2, and both tables are discussed in detail in Subsection 3.2.2.

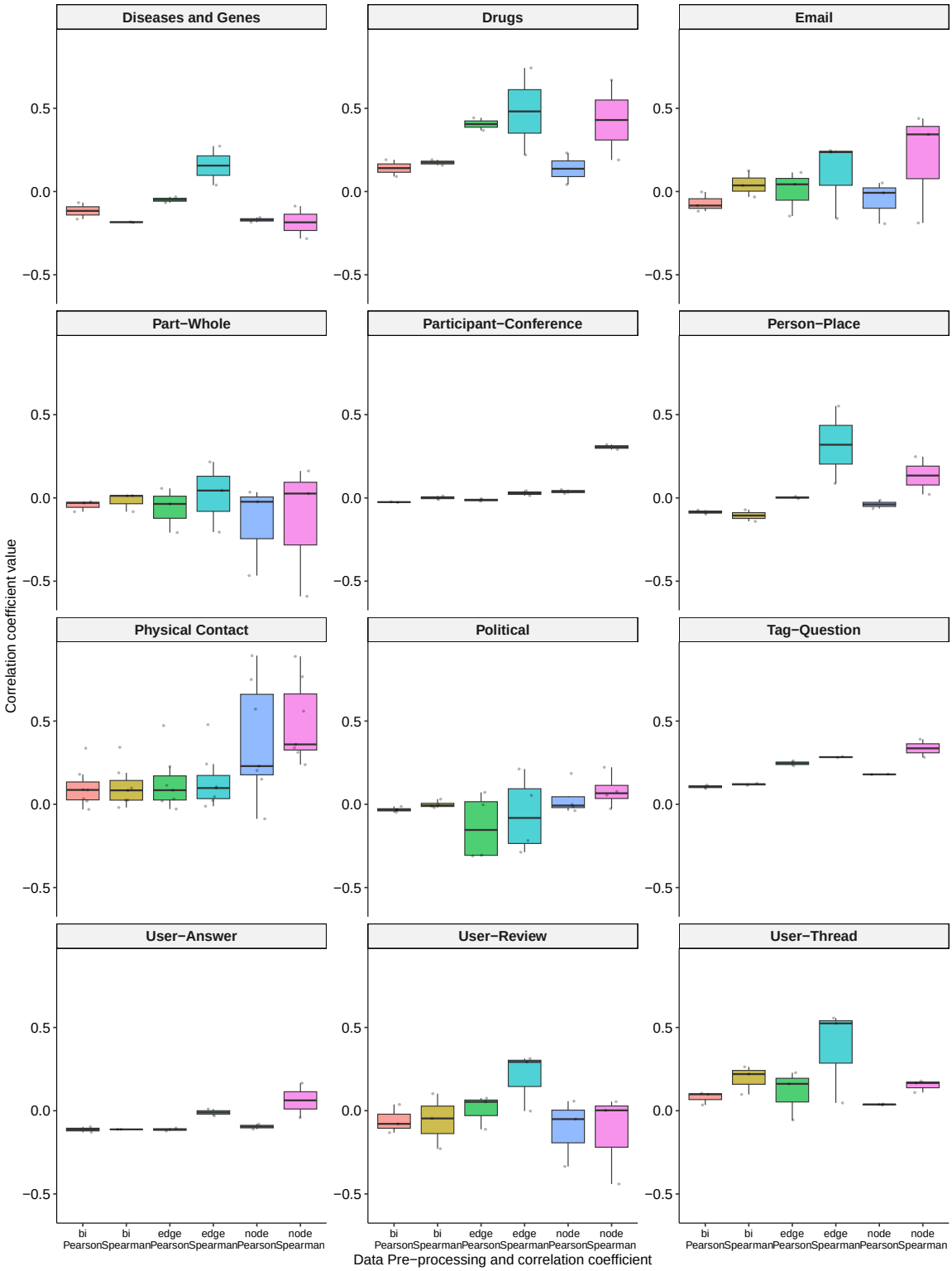


Figure 6: Box plots with points of correlation values across preprocessing methods and correlation types, grouped by hypergraph segments. Labels `bi`, `edge`, and `node` denote bipartite, edge-, and node-centric strategies.

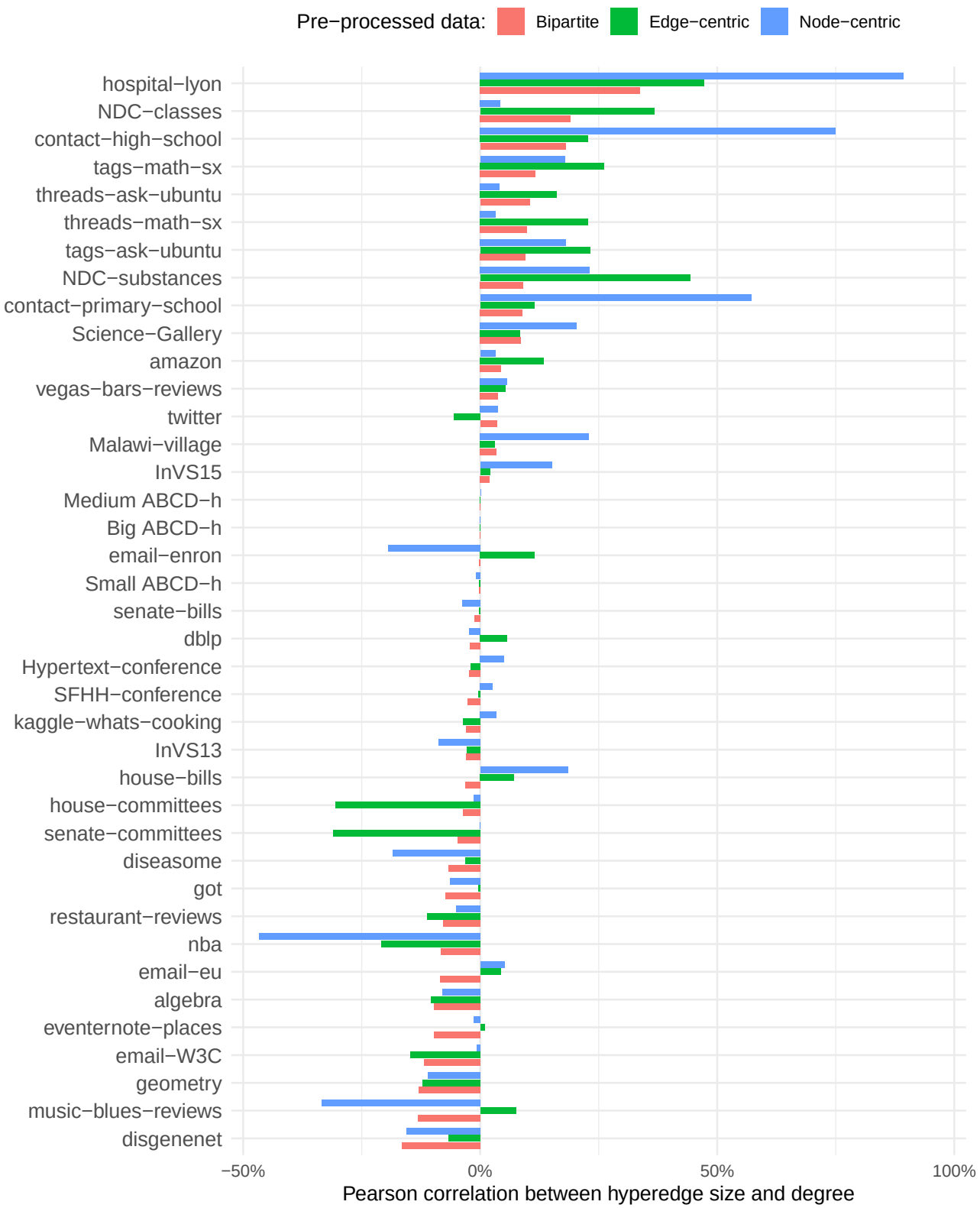


Figure 7: Pearson correlations between hyperedge size and node degree calculated for three datasets pre-processed according to: (1) bipartite representation, (2) edge-centric, (3) node-centric way.

Hypergraph	Pearson	Spearman	Kendall	GAM Monotonicity
hospital-lyon	0.337***	0.343***	0.278***	Inc.
NDC-classes	0.191***	0.157***	0.119***	Inc.
contact-high-school	0.180***	0.189***	0.152***	Inc.
tags-math-sx	0.116***	0.114***	0.086***	Inc.
threads-ask-ubuntu	0.104***	0.263***	0.208***	Inc.
threads-math-sx	0.099***	0.220***	0.165***	Inc.
tags-ask-ubuntu	0.095***	0.126***	0.095***	Inc.
NDC-substances	0.090***	0.191***	0.129***	Inc.
contact-primary-school	0.089***	0.084***	0.068***	Inc.
Science-Gallery	0.086***	0.097***	0.079***	Inc.
amazon	0.044**	0.041**	0.040**	Non-sign.
vegas-bars-reviews	0.037**	0.102***	0.073***	Non-sign.
twitter	0.035***	0.098***	0.073***	Inc.
Malawi-village	0.034***	0.025***	0.021***	Inc.
InVS15	0.020***	0.025***	0.020***	Inc.
email-enron	-0.002	0.125***	0.093***	Inc.
senate-bills	-0.013***	-0.003	-0.002	Dec.
dblp	-0.022***	0.013**	0.011**	Dec.
Hypertext-conference	-0.023***	-0.010*	-0.008*	Dec.
SFHH-conference	-0.027***	0.012**	0.010**	Dec.
InVS13	-0.030**	-0.019**	-0.016**	Non-sign.
kaggle-whats-cooking	-0.030***	0.014***	0.009***	Dec.
house-bills	-0.031***	0.031***	0.020***	Dec.
house-committees	-0.036**	-0.011	-0.007	Dec.
senate-committees	-0.048**	-0.019	-0.013	Dec.
diseasome	-0.067*	-0.186***	-0.138***	Dec.
got	-0.073***	-0.141***	-0.100***	Dec.
restaurant-reviews	-0.079***	-0.046**	-0.032**	Dec.
nba	-0.083***	-0.083***	-0.059***	Dec.
email-eu	-0.084***	0.037***	0.024***	Dec.
algebra	-0.097***	-0.112***	-0.078***	Dec.
eventernote-places	-0.097***	-0.071***	-0.048***	Dec.
email-W3C	-0.118***	-0.033**	-0.029**	Dec.
geometry	-0.129***	-0.112***	-0.078***	Dec.
music-blues-reviews	-0.132***	-0.228***	-0.159***	Dec.
disgenenet	-0.166***	-0.182***	-0.123***	Dec.

Table 8: Correlation measures (sorted by decreasing Pearson) for bipartite representation with significance stars (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.00001$) and the monotonicity direction of fitted monotonic GAM.

A.4.2 Identification of Relationship Types: Results of Statistical Tests

Table 9 reports, for each hypergraph, the p-values from the three nested statistical tests and the resulting classification of the relationship type, following the procedure outlined in Subsection 2.3. The classification is based on a conservative significance threshold of $\alpha = 0.00001$, ensuring robustness against spurious detections. The overall distribution of relationship types across the 36 empirical hypergraphs is summarized in Table 3 and discussed in Subsection 3.3.2.

A.4.3 Identification of Relationship Types: Visual Inspection

Subsection 3.3.1 introduced and discussed four representative examples (out of 36) of hypergraphs, each illustrating one of the four identified relationship types: non-monotonic, monotonic, linear, and no relationship. These examples were visualized in Figures 4 and 5. The remaining 32 empirical hypergraphs are presented in

Hypergraph	**Relationship**	**p-val for H_0 : monotonic**	**p-val for H_0 : linear**	**p-val for H_0 : $R^2 = 0$**	**N**
hospital-lyon	Linear	1	0.012	4.00E-118	4,427
NDC-classes	Non-monotonic	0	0	0	156,185
contact-high-school	Monotonic	0.19	3.00E-07	1.90E-132	18,192
tags-math-sx	Monotonic	0.088	0	0	593,121
threads-ask-ubuntu	Monotonic	0.26	1.10E-214	0	346,537
threads-math-sx	Monotonic	1	0	0	1,610,393
tags-ask-ubuntu	Monotonic	0.2	0	0	499,298
NDC-substances	Non-monotonic	2.40E-104	0	1.10E-97	53,528
contact-primary-school	Linear	0.016	0.016	8.80E-55	30,729
Science-Gallery	Non-monotonic	2.10E-32	9.60E-292	0	717,690
amazon	No relationship	0.081	0.077	0.0015	5,112
vegas-bars-reviews	No relationship	0.065	0.016	4.60E-05	11,865
twitter	Non-monotonic	0	4.70E-105	3.70E-15	50,850
Malawi-village	Linear	0.01	0.01	2.00E-52	201,069
InVS15	Linear	0.28	0.0051	3.80E-14	149,949
Medium ABCD-h	No relationship	1	0.056	0.2	1,079,154
Big ABCD-h	No relationship	1	0.14	0.72	2,000,000
email-enron	Non-monotonic	1.80E-09	5.40E-27	0.87	4,623
Small ABCD-h	No relationship	0.057	0.057	0.33	107,960
senate-bills	Non-monotonic	1.40E-06	0.00096	9.00E-10	232,147
dblp	Non-monotonic	1.30E-109	1.50E-31	2.30E-11	88,458
Hypertext-conference	Linear	0.13	0.12	6.50E-06	39,048
SFHH-conference	Non-monotonic	1.30E-31	2.80E-57	8.40E-21	116,636
InVS13	No relationship	0.0099	0.0099	2.50E-05	19,380
kaggle-whats-cooking	Non-monotonic	2.90E-35	6.10E-34	6.90E-87	428,249
house-bills	Non-monotonic	0	0	1.70E-260	1,248,666
house-committees	Non-monotonic	1.10E-21	9.30E-31	7.10E-05	11,843
senate-committees	Monotonic	0.056	2.00E-21	0.00038	5,408
diseasome	Monotonic	0.34	6.00E-06	0.026	1,109
got	Monotonic	1.20E-05	2.10E-42	9.80E-16	12,114
restaurant-reviews	Linear	0.0044	0.0044	8.20E-08	4,601
nba	Non-monotonic	4.90E-37	2.50E-87	0	644,051
email-eu	Non-monotonic	0	2.90E-103	7.80E-141	89,409
algebra	Monotonic	0.11	7.00E-07	1.20E-18	8,262
eventernote-places	Non-monotonic	2.40E-13	0	0	713,400
email-W3C	Monotonic	1	1.50E-26	1.20E-42	13,361
music-blues-reviews	Non-monotonic	7.70E-33	3.60E-19	9.00E-42	10,499
geometry	Monotonic	0.00078	1.20E-06	8.10E-48	12,485
disgenenet	Monotonic	0.89	1.50E-112	0	112,471

Table 9: Results of statistical tests (p-values) identifying one of four relationship types, i.e. non-monotonic, monotonic, linear, no relationship, by running following three statistical tests: (1) ANOVA test comparing two models with $H_0 : unrestrictedGAM = monotonicGAM$, (2) ANOVA test comparing two models with $H_0 monotonicGAM = OLS$, (3) F-test with $H_0 : R^2 = 0$. Significance level $\alpha = 10^{-5}$.

Figures 8, 9, 10, 11, 12, and 13, along with an additional synthetic hypergraph generated using the ABCD-h algorithm, shown in Figure 9.

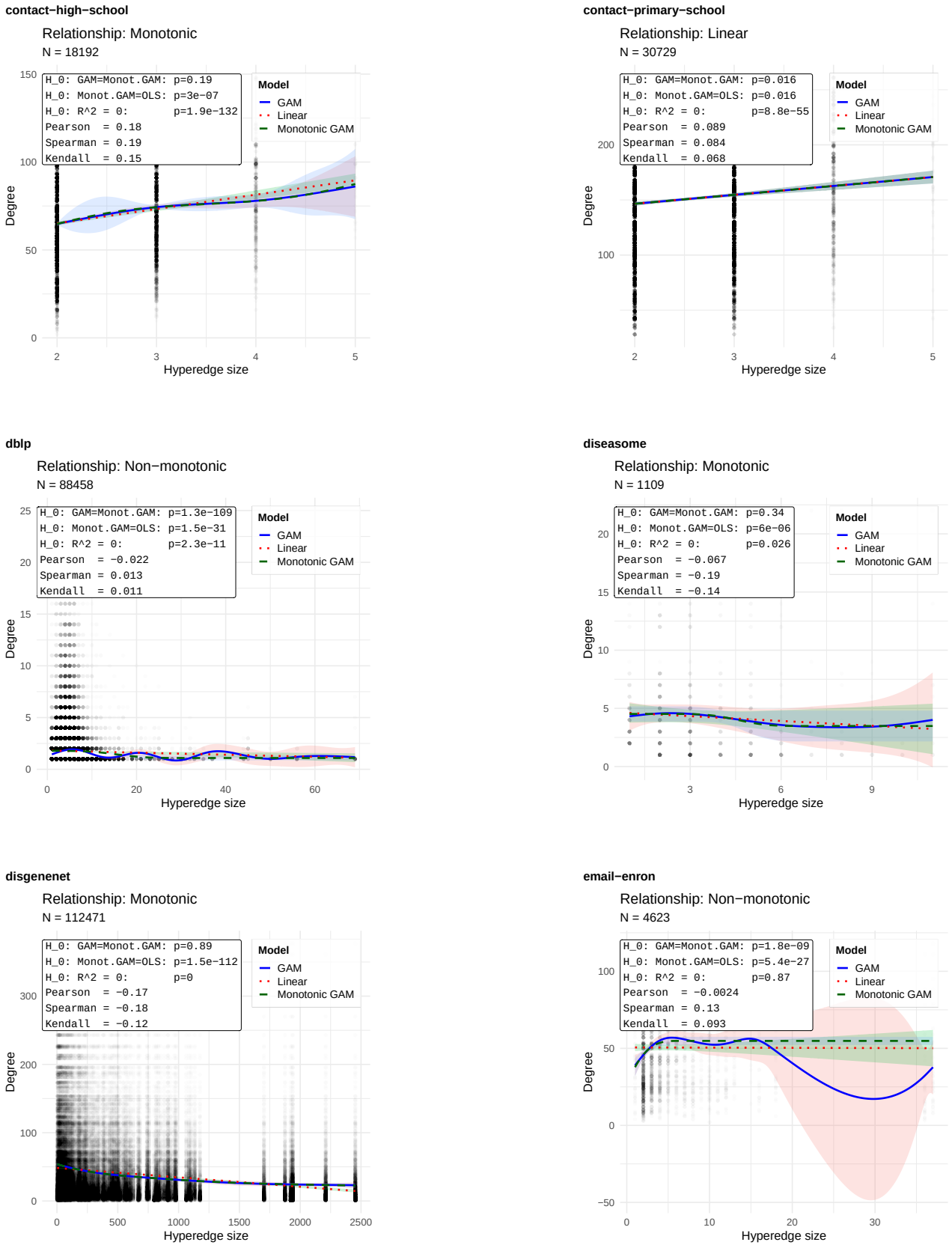


Figure 8: Scatterplots of node degree vs. hyperedge size (bipartite) with GAM, monotonic-GAM and OLS fits (99.999% CIs) for `contact-high-school`, `contact-primary-school`, `dblp`, `diseasome`, `disgenenet` and `email-enron`.

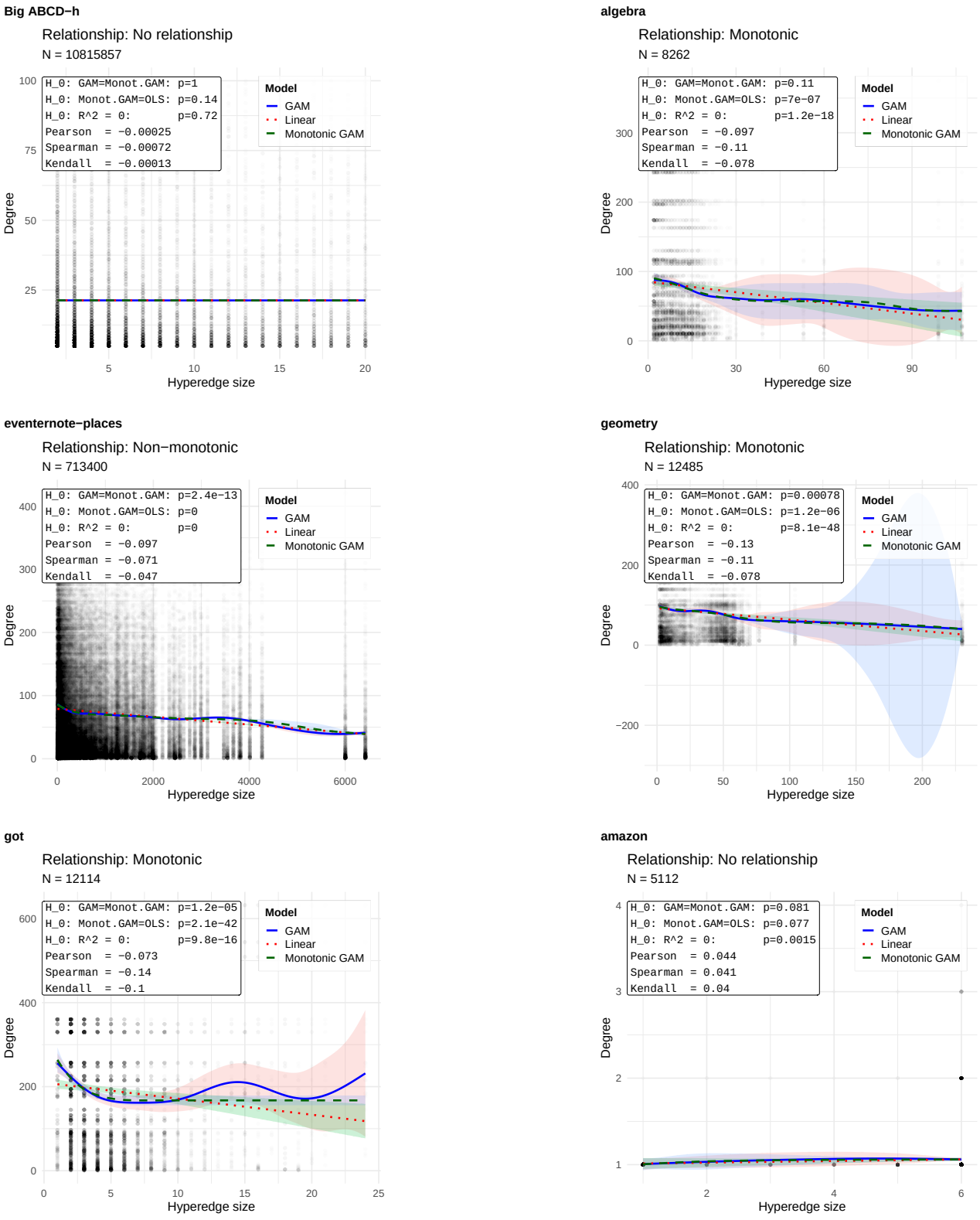


Figure 9: Scatterplots of node degree vs. hyperedge size with GAM, monotonic-GAM and OLS fits (99.999% CIs) for synthetic hypergraph generated by the ABCD-h algorithm and empirical hypergraphs: `algebra`, `eventernote-places`, `geometry`, `got` and `amazon`.

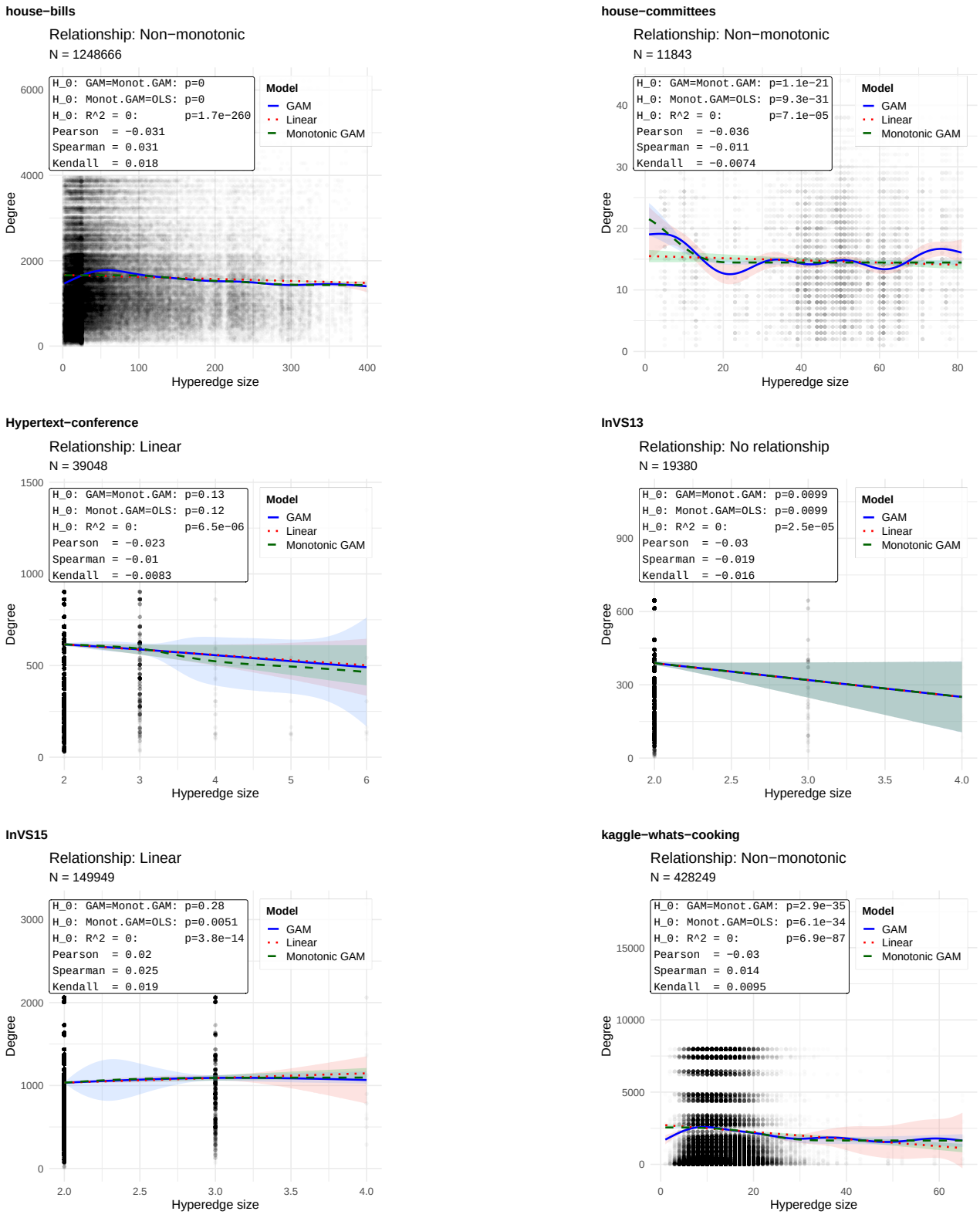


Figure 10: Scatterplots of node degree vs. hyperedge size (bipartite) with GAM, monotonic-GAM and OLS fits (99.999% CIs) for `house-bills`, `house-committees`, `Hypertext-conference`, `InVS13`, `InVS15` and `kaggle-whatscooking`.

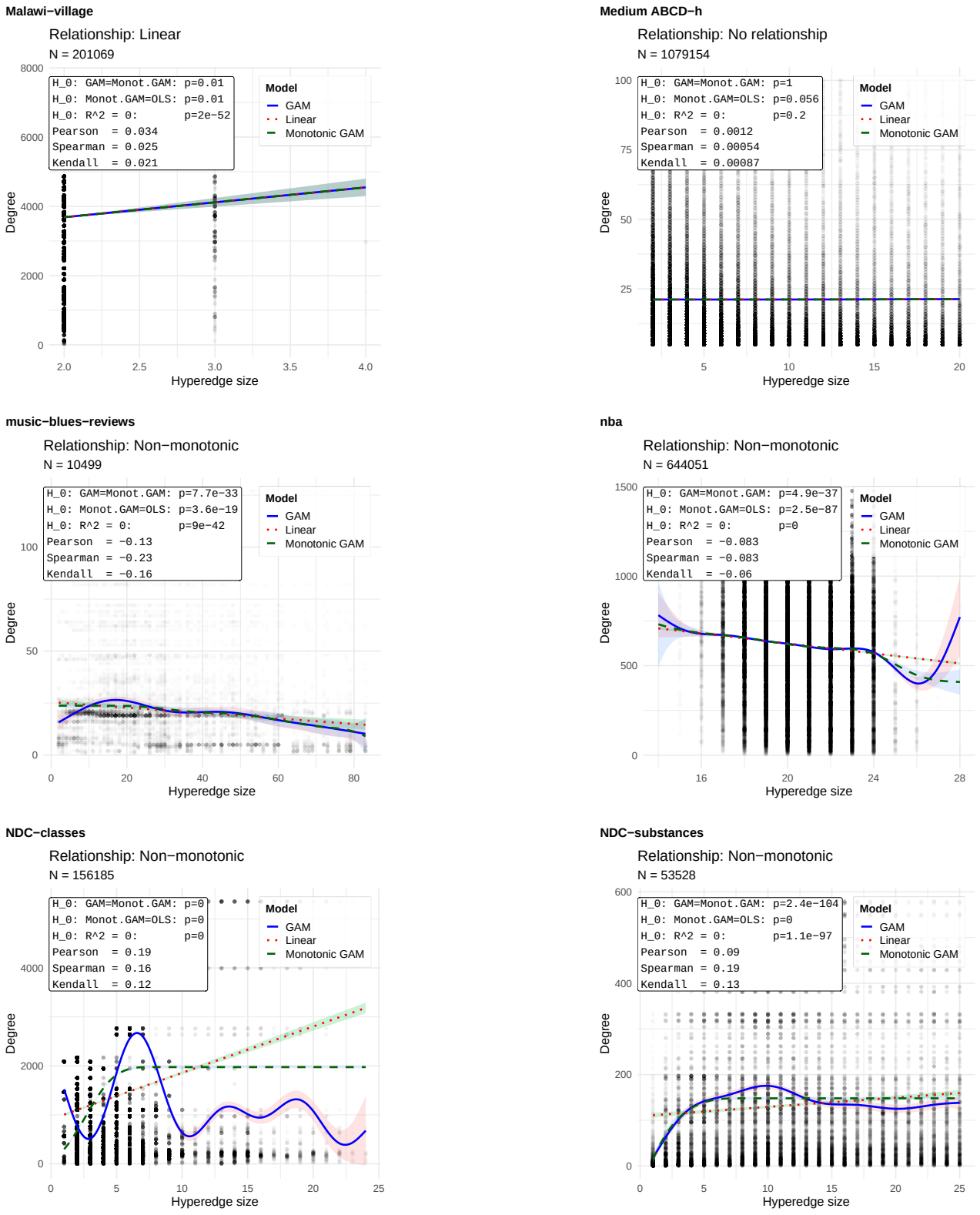


Figure 11: Scatterplots of node degree vs. hyperedge size (bipartite) with GAM, monotonic-GAM and OLS fits (99.999% CIs) for `Malawi-village`, `Medium ABCD-h`, `music-blues-reviews`, `nba`, `NDC-classes` and `NDC-substances`.

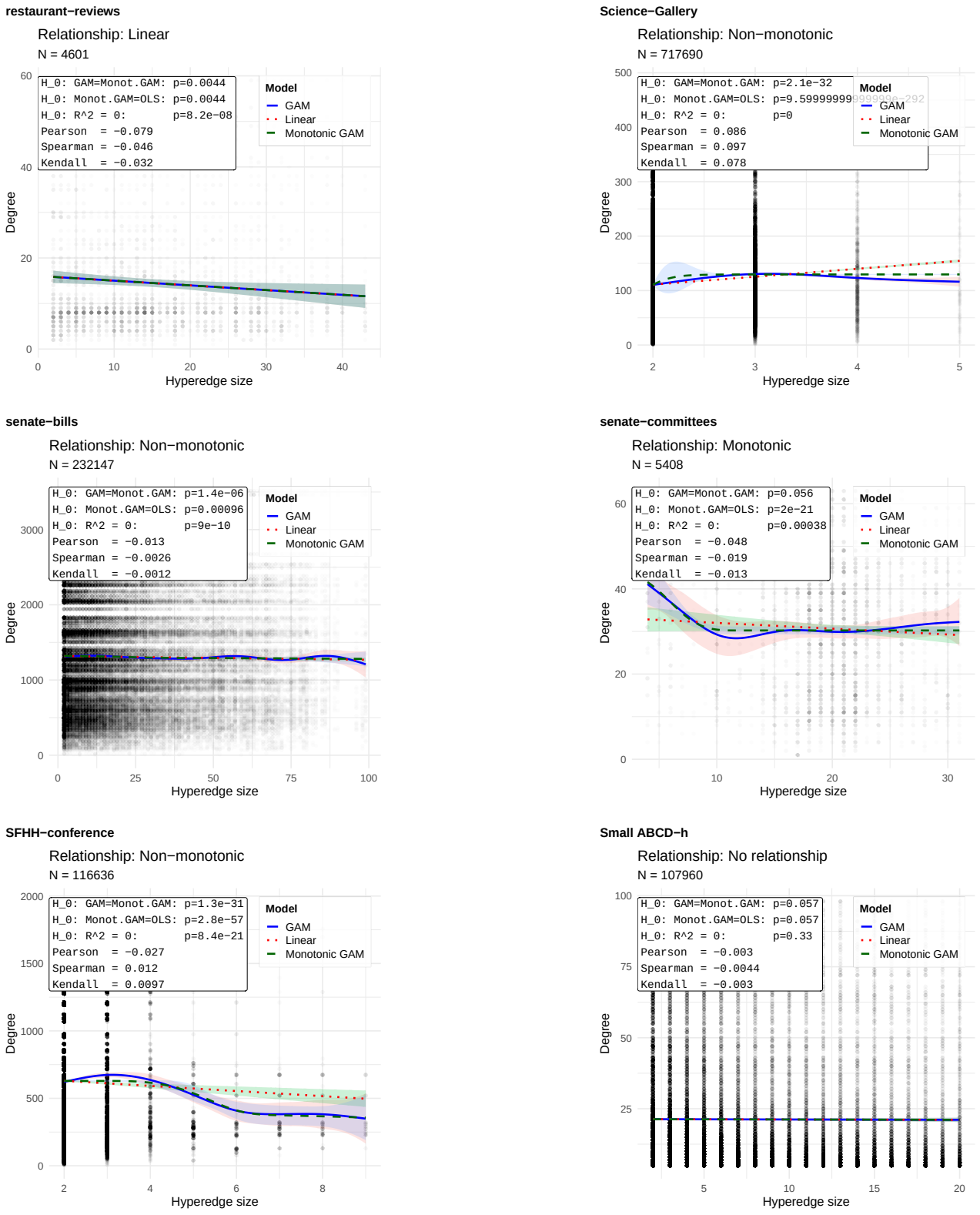


Figure 12: Scatterplots of node degree vs. hyperedge size (bipartite) with GAM, monotonic-GAM and OLS fits (99.999% CIs) for `restaurant-reviews`, `Science-Gallery`, `senate-bills`, `senate-committees`, `SFHH-conference` and `Small ABCD-h`.

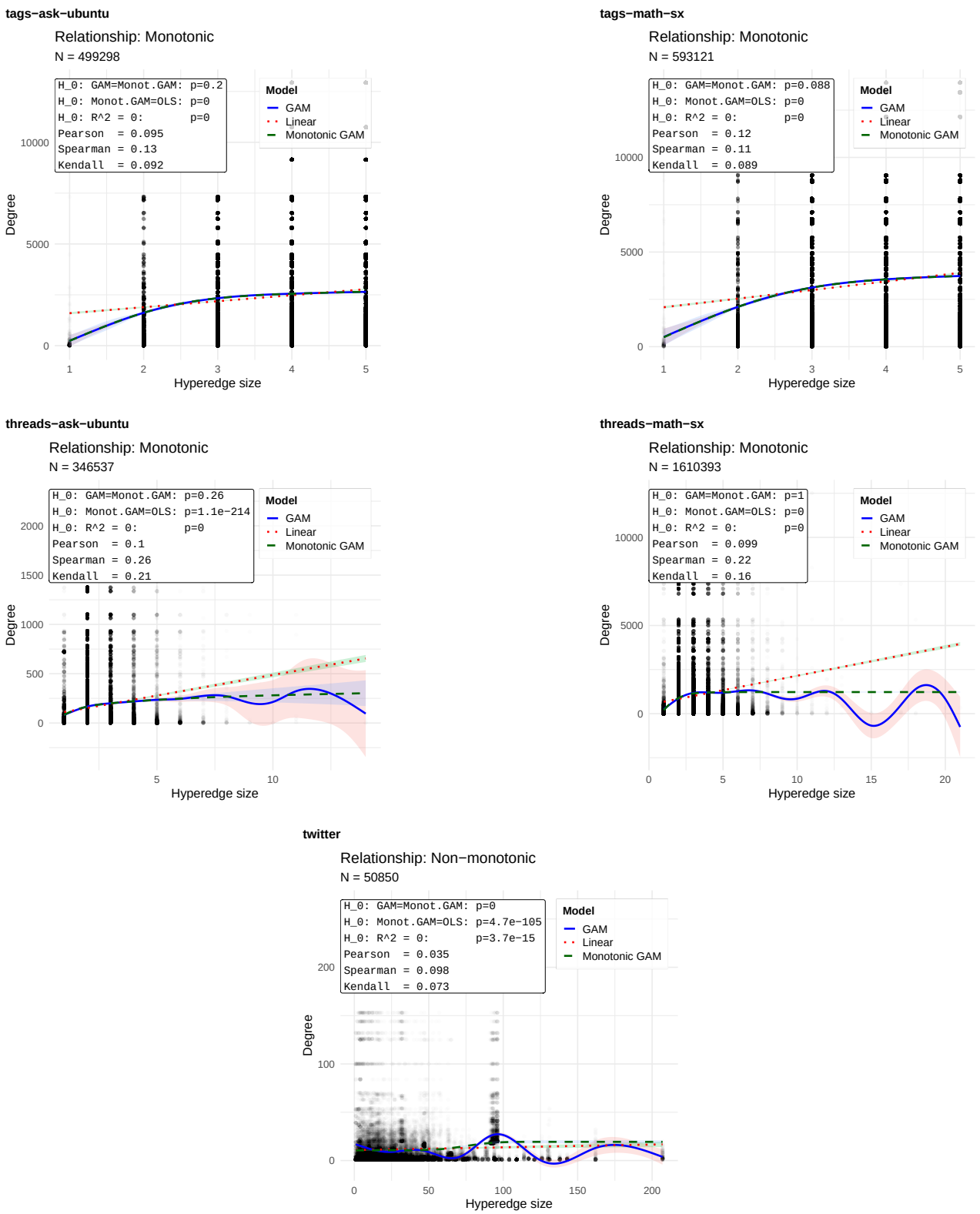


Figure 13: Scatterplots of node degree vs. hyperedge size (bipartite) with GAM, monotonic-GAM and OLS fits (99.999% CIs) for `tags-ask-ubuntu`, `tags-math-sx`, `threads-ask-ubuntu`, `threads-math-sx` and `twitter`.