

Perplexity-Homophily Index: Homophily through Diversity in Hypergraphs

Gaurav Kumar

IISER Pune

India

gaurav2k29@gmail.com

Akrati Saxena

LIACS, Leiden University

The Netherlands

a.saxena@liacs.leidenuniv.nl

Chandrakala Meena

IISER Pune

India

chandrakala@iiserpune.ac.in

Abstract

Real-world complex systems are often better modeled as hypergraphs, where edges represent group interactions involving multiple entities. Understanding and quantifying homophily (similarity-driven association) in such networks is essential for analyzing community formation and information flow. We propose a hyperedge-centric framework to quantify homophily in hypergraphs. Each interaction is represented as a hyperedge, and its interaction perplexity measures the effective number of distinct attributes it contains. Comparing this observed perplexity with a degree-preserving random baseline defines the diversity gap, which quantifies how diverse an interaction is than expected by chance. The global homophily score for a network, called “Perplexity-Homophily Index”, is computed by averaging the normalized diversity gap across all hyperedges. Experiments on synthetic and real-world datasets show that the proposed index captures the full distribution of homophily and reveals how homophilic and heterophilic tendencies vary with interaction size in hypergraphs.

CCS Concepts

• Information Systems → Social Networks; • Mathematics of Computing → Hypergraphs.

Keywords

Hypergraphs, Homophily, Diversity, Perplexity

1 Introduction

Traditionally, complex systems are modeled as graphs with edges capturing pairwise interactions. However, many real-world interactions involve multiple nodes (groups) rather than pairs, such as co-authors on a paper, items bought together, or legislators co-sponsoring a bill. These systems are better captured by hypergraphs, where hyperedges represent group interactions. Real-world networks, driven by human behavior, are evolved using homophily [10] that shows the tendency of entities to interact preferentially with others sharing similar attributes or interests (e.g., age, beliefs, nationality). Measuring homophily is essential for understanding community formation, information flow [5, 12], and the emergence of structural inequalities in social systems [13].

A widely used measure of homophily in pairwise graphs is Newman’s assortativity coefficient [11], which quantifies the deviation of a graph’s structure from random mixing and enables comparisons across networks. Despite extensive work on homophily in pairwise graphs [14], such measures for hypergraphs remain limited. Veldt et al. [17] characterize the homophily types for a specific class (e.g., majority or minority preference) but provide no aggregate score for the whole hypergraph and assumes a k -uniform hypergraph. Li

et al. [8] estimate node and hyperedge homophily by counting the fraction of same-class pairs and aggregating it for the whole hypergraph, while Telyatnikov et al. [16] use message passing to define node-level homophily. However, all existing approaches lack an appropriate null model, and their scores are not directly comparable across hypergraphs.

In this work, we introduce a novel, hyperedge-centric framework to measure homophily in hypergraphs. We first define the interaction perplexity $D(e)$, which quantifies the “effective number of distinct attributes” within a hyperedge. Perplexity is an order-1 Hill number [6] originally used to measure the diversity of species in a region. In hypergraphs, hyperedges can vary from pure (all nodes have the same attribute) to balanced (attributes are evenly distributed). Interaction perplexity captures this full spectrum of hyperedge compositions. We compare the observed perplexity of each hyperedge to a random baseline that accounts for both the global attribute distribution and the degree distribution. The homophily score $\phi(e)$ for a hyperedge e is defined as the normalized difference of the observed and the baseline perplexity. The Perplexity-Homophily Index for a hypergraph, $\Phi(H)$, is the average over homophily scores of all hyperedges. This index is comparable across datasets with differing hyperedge sizes and attribute distributions.

We validate our method on various real-world hypergraphs, and the results show that $\Phi(H)$ is able to characterize homophily across a wide range of datasets. Moreover, we highlight how homophilic or heterophilic tendencies vary with interaction size, such as the stronger homophily in small co-sponsorship groups and near-random mixing in larger ones for US Congressional bill co-sponsorship dataset.

2 The Proposed Homophily Measure

A hypergraph H is represented as (V, E) , where V and E are the set of nodes and hyperedges, respectively. A hyperedge $e \in E$ is a non-empty subset of V and every node $v \in V$ is associated with an attribute from a set of possible attributes M . Based on the distribution of attributes of nodes in a hyperedge, it can be classified as pure if all nodes have the same attribute, balanced if all attributes are in equal proportion, and biased if one or more attributes dominate. A k -uniform hypergraph is one where each hyperedge has exactly k nodes.

2.1 Interaction Perplexity

Let $p_{i,e}$ denote the fraction of nodes with attribute i in hyperedge e . We now define the interaction perplexity $D(e)$ as the perplexity calculated over the distribution of attributes $\{p_{i,e}\}_{i \in M}$:

$$D(e) = 2^{-\sum_{i \in M} p_{i,e} \log_2 p_{i,e}} \quad (1)$$

$D(e)$ ranges from 1 to the number of attributes in e , m_e . A pure interaction will have $D(e) = 1$, while a balanced interaction will have $D(e) = m_e$. Biased interactions will have a perplexity between 1 and m_e . Thus, $D(e)$ captures the full spectrum of attribute dominance within H , which is essential for accurately quantifying the homophily of the hypergraph.

Example. Consider a hyperedge with four researchers from three departments: Physics (2), CS (1), and Math (1), giving the distribution (2, 1, 1). While three departments are present, Physics dominates, resulting in lower effective diversity. The interaction perplexity for this distribution is 2.83, meaning the group behaves as if it effectively contains about 2.83 equally represented departments. If all researchers were from Physics, the perplexity would be 1, and if each belonged to a different department, it would be 4.

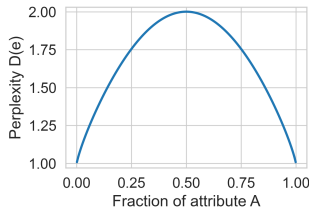


Figure 1: Perplexity $D(e)$ for a hyperedge with two attributes A and B .

2.2 Baseline and Diversity Gap

To assess homophily, the observed perplexity must be compared to a null model that accounts for both the global attribute distribution and degree distribution of nodes. In a degree-preserving randomized k -uniform hypergraph, the baseline perplexity B_k represents the expected number of distinct attributes in a hyperedge of size k under random mixing.

A hypergraph with a highly skewed global attribute distribution (e.g., 90% of nodes sharing one attribute) will naturally have a low average interaction perplexity, since most interactions will be dominated by the majority group. Similarly, high-degree nodes are more likely to appear in sampled interactions, increasing the expected prevalence of their attributes.

We estimate the baseline using a Monte Carlo procedure: for each sample, we draw k nodes without replacement, with sampling probability proportional to their k -degree (the number of size- k hyperedges the node belongs to), and then compute the interaction perplexity of the resulting hyperedge. For a nonuniform hypergraph, each k -uniform sub-hypergraph will have its own baseline. The underlying distribution of the null model is a multivariate non-central hypergeometric distribution. Computing the expected value of the baseline perplexity for a hypergraph would require summing over a large state space of all possible combinations of attributes, which is not feasible as the size of the hypergraph increases.

We define the Diversity Gap $\Delta(e)$ of an interaction as the difference between the observed perplexity and its baseline.

$$\Delta(e) = B_{|e|} - D(e)$$

When $\Delta(e) > 0$, the observed hyperedge is less diverse than the expected diversity (high homophily). Similarly, $\Delta(e) < 0$ shows

that the hyperedge is more diverse than expected (heterophily). A gap near zero suggests that the diversity of the hyperedge is as expected, and interactions are consistent with the null model.

The Diversity Gap is maximum for a pure hyperedge, since $D(e) = 1$:

$$\Delta_{\max}(e) = B_{|e|} - 1$$

and minimum for a balanced hyperedge:

$$\Delta_{\min}(e) = B_{|e|} - m_e$$

2.3 Perplexity-Homophily Index

To obtain a homophily score that is comparable across hyperedges, we normalize the diversity gap by its maximum value:

$$\phi(e) = \frac{\Delta(e)}{\Delta_{\max}(e)} = \frac{B_{|e|} - D(e)}{B_{|e|} - 1} \quad (2)$$

This normalization provides a clear interpretation for measuring homophily, as it represents the fraction of the maximum possible reduction in diversity relative to the baseline. The maximum score of +1 corresponds to a fully pure hyperedge, while $\phi(e) = 0.8$ means that the hyperedge is 80% less diverse than expected, or 80% of the way being completely homophilic. The minimum score depends on the number of attributes in the hyperedge, $\phi_{\min}(e) = \Delta_{\min}(e)/\Delta_{\max}(e)$, providing a meaningful way to compare homophily across hyperedges with different attribute compositions.

The Perplexity-Homophily Index of a hypergraph H is defined as the average homophily score over all hyperedges:

$$\Phi(H) = \frac{1}{|E|} \sum_{e \in E} \phi(e) \quad (3)$$

Similarly, $\Phi(H_k)$ measures the homophily of k -uniform sub-hypergraphs (H_k) present in a hypergraph H (Fig. 3). For $k = 2$, the index closely approximates Newman’s assortativity coefficient.

3 Results

We first analyze our method on synthetic hypergraphs generated using a Hypergraph Stochastic Block Model (HSBM) [3]. The HSBM produces a k -uniform hypergraph with $|M|$ equal-sized partitions, each of size $|V|/|M|$ (assuming $|V|$ is divisible by $|M|$). For any subset of k nodes, it creates a hyperedge with probability p if all

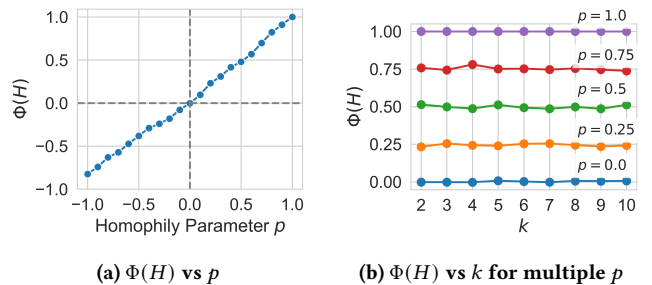
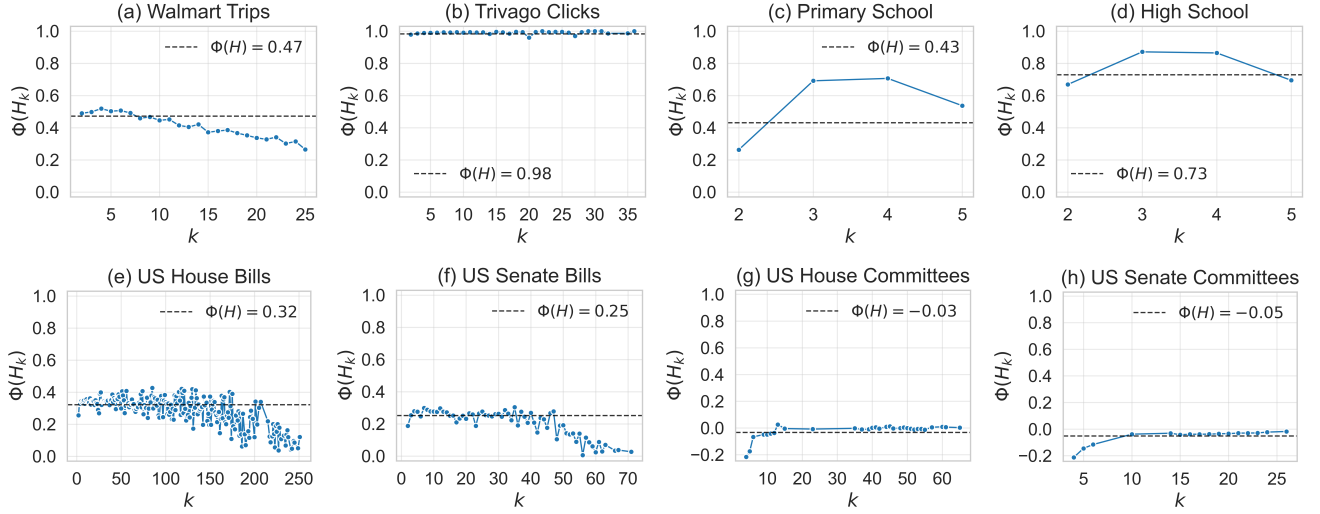


Figure 2: (a). Homophily vs. p for 10-uniform hypergraph with 1000 nodes and 10 evenly distributed attributes, and (b). Comparison of homophily scores for k -uniform hypergraphs with varying p on 1000 nodes network with 10 evenly distributed attributes.

Table 1: Summary of real-world hypergraph datasets used in the analysis.

Dataset (H)	Nodes	$ V $	Hyperedges	$ E $	Attribute	$ M $	$\Phi(H)$
Walmart Trips [1]	Products	88,860	Shopping carts	65,979	Product category	11	0.47
Trivago Clicks [2]	Hotels	172,738	Browsing sessions	220,971	Country	160	0.98
Primary School [2]	Students/Teachers	242	Proximity groups	12,704	Class / Teacher	11	0.43
High School [2, 9]	Students	327	Proximity groups	7,818	Class	9	0.73
US House Bills [2, 4]	Legislators	1,494	Bills	54,933	Political party	2	0.32
US Senate Bills [2, 4]	Legislators	294	Bills	21,721	Political party	2	0.24
US House Committees [2, 15]	Legislators	1,290	Committees	336	Political party	2	-0.03
US Senate Committees [2, 15]	Legislators	282	Committees	301	Political party	2	-0.05

**Figure 3: $\Phi(H_k)$ vs. k across all datasets.**

nodes belong to the same partition, and with probability q otherwise. For our experiments, we modify this model in such a way that it generates a pure hyperedge with probability p when the homophilic parameter $p > 0$, a balanced hyperedge with probability $|p|$ when $p < 0$, and a random hyperedge otherwise. Fig. 2 presents results for a 10-uniform hypergraph with 10 attributes and 1000 nodes. As shown in Fig. 2(a), the Perplexity-Homophily Index $\Phi(H)$ increases linearly as p varies from -1 to $+1$, matching the model’s intended structure. Fig. 2(b) shows $\Phi(H)$ for different values of k and p ; the index accurately captures the intended homophily level.

Table 1 summarizes the real-world datasets used in our analysis. For any hypergraph H , let H_k be the sub-hypergraph of H obtained by retaining only hyperedges of size k . In Fig. 3, for each dataset H , we plot and compare $\Phi(H_k)$ for all k , where the black dashed line shows the value of $\Phi(H)$ for reference, and blue dots show the value of $\Phi(H_k)$ for each k . It shows that the homophily $\Phi(H_k)$ varies substantially with interaction size k . We first observe that the method accurately identifies extreme cases of homophily. The Trivago Clicks hypergraph exhibits near-perfect homophily with $\Phi(H) = 0.98$, reflecting highly focused user browsing behavior. As shown in Fig. 3(b), $\Phi(H_k) \approx 1$ across all session lengths. Fig. 4(b) compares the observed perplexity with the baseline perplexity for each value of k . Although the baseline perplexity varies with k ,

suggesting that a random session would check out hotels in multiple countries, the observed perplexity remains constant at $D(e) = 1$. The decrease in baseline is due to the fact that a limited number of countries participate and are available for baseline calculation in large sessions. These findings confirm that users browse hotels exclusively within a single country, regardless of session length. In contrast, the US Congress committees dataset exhibits homophily scores consistently close to zero (Figs. 3(g), 3(h)), reflecting that the committees are intentionally constructed to maintain balanced representation across political parties. As a result, their attribute composition is identical (Fig. 4(g)) or even more diverse from that of the null model (Fig. 4(h)).

A strong dependence on interaction size emerges in datasets describing legislative co-sponsorship. In both the US House and Senate, homophily decreases sharply once the number of co-sponsors exceeds a threshold (Figs. 3(e), 3(f)). Small bills exhibit high homophily, whereas large bills have $\Phi \approx 0$, indicating that bipartisan support is necessary for attracting many co-sponsors. Correspondingly, the perplexity of large bills becomes indistinguishable from the null model (Figs. 4(e), 4(f)). A similar but more gradual pattern appears in the Walmart shopping data: small carts are strongly homophilic, but homophily decreases with cart size (Fig. 3(a)). Observed perplexity saturates to a value of ≈ 2.5 as k increases, while

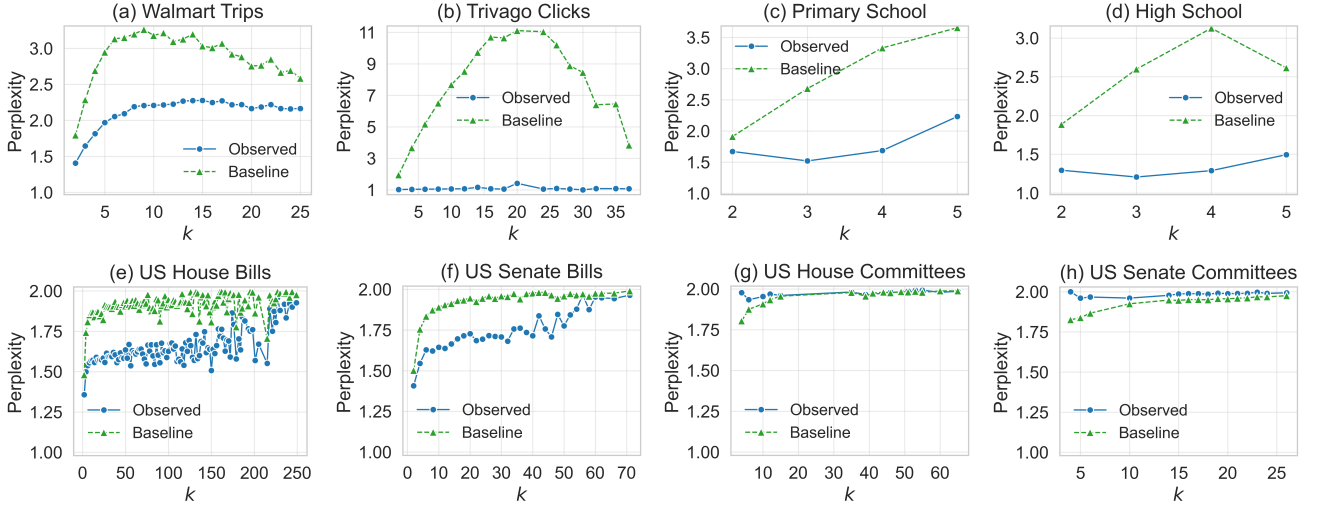


Figure 4: Perplexity vs. hyperedge size (k).

baseline perplexity rises to 11 then falls back to 4, making the largest carts match more closely with the null model (Fig. 4(a)).

Finally, in the high school and primary school contact networks, the strongest homophily occurs for groups of size $k = 3$ and $k = 4$, with lower homophily for pairs and larger groups. High school students show stronger class-based segregation than primary school students, reflected in the higher overall homophily ($\Phi(H) = 0.73$ vs. 0.43). The curve for baseline is consistently above the observed perplexity in Figs. 4(c) and 4(d). For all group sizes, the observed perplexity remains below 1.5 for High School, while it stays above 1.5 for Primary School. Overall, these results show that the Perplexity-Homophily Index captures both overall homophily and its dependence on interaction size.

4 Discussion and Conclusion

The proposed framework provides a general and flexible approach for quantifying homophily in hypergraphs. By computing the diversity of each hyperedge, comparing it with the expected diversity under a degree- and attribute-aware null model, and defining homophily as a normalized deviation between the two, our method extends the concept of homophily well beyond the limitations of pairwise networks. Traditional graph-based measures, which only distinguish between same-attribute and different-attribute edges, are insufficient in higher-order settings where interactions may involve several nodes and multiple attributes. In contrast, the use of perplexity naturally incorporates both the number of distinct attributes and their proportions, making it well-suited for capturing the richness of higher-order interactions.

It is important to emphasize that perplexity is just one member of a larger family of diversity indices, specifically the order-1 Hill number [6]. The framework is not restricted to this choice. For example, the order-2 Hill number (the inverse Simpson index), widely known as the Laakso-Taagepera index to measure “effective number of parties” [7] places greater emphasis on majority attributes and discounts minority ones. In an attribute composition such as

(8, 1, 1), the inverse Simpson index yields significantly lower diversity (≈ 1.52) than perplexity (≈ 1.89) because the squared term strongly amplifies the dominant class. By using perplexity, the present framework maintains sensitivity to the full distribution of attributes and provides a more balanced assessment of diversity within a hyperedge.

Overall, the framework is adaptable and can be paired with any suitable diversity index, depending on the application and the desired sensitivity to minority attributes. This flexibility, together with its natural extension of homophily to higher-order structures, makes the framework broadly applicable for analyzing complex datasets where interactions extend beyond simple pairwise relations. Future research may extend this framework to temporal and multilayer hypergraphs and explore alternative diversity indices to tailor sensitivity to specific applications. Integrating the method with generative higher-order models would further enable parameter inference and realistic benchmark construction.

References

- [1] Ilya Amburg, Nate Veldt, and Austin R. Benson. 2020. Clustering in graphs and hypergraphs with categorical edge labels. In *Proceedings of the Web Conference*.
- [2] Philip S Chodrow, Nate Veldt, and Austin R Benson. 2021. Hypergraph clustering: from blockmodels to modularity. *Science Advances* (2021).
- [3] Sam Cole and Yizhe Zhu. 2020. Exact Recovery in the Hypergraph Stochastic Block Model: A Spectral Algorithm. *Linear Algebra Appl.* 593 (May 2020), 45–73.
- [4] James H. Fowler. 2006. Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis* 14, 04 (2006), 456–487.
- [5] Benjamin Golub and Matthew O. Jackson. 2009. How Homophily Affects Diffusion and Learning in Networks. arXiv:0811.4013 [physics] doi:10.48550/arXiv.0811.4013
- [6] M. O. Hill. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* 54, 2 (1973), 427–432. jstor:1934352
- [7] Markku Laakso and Rein Taagepera. 1979. “Effective” Number of Parties: A Measure with Application to West Europe. *Comparative Political Studies* 12, 1 (1979), 3–27.
- [8] Ming Li, Yongchun Gu, Yi Wang, Yujie Fang, Lu Bai, Xiaosheng Zhuang, and Pietro Liò. 2025. When Hypergraph Meets Heterophily: New Benchmark Datasets and Baseline. *Proceedings of the AAAI Conference on Artificial Intelligence* 39, 17 (April 2025), 18377–18384. doi:10.1609/aaai.v39i17.34022
- [9] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. 2015. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors,

- Contact Diaries and Friendship Surveys. *PLOS ONE* 10, 9 (2015), e0136497.
- [10] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, Volume 27, 2001 (Aug. 2001), 415–444. doi:10.1146/annurev.soc.27.1.415
- [11] M. E. J. Newman. 2003. Mixing Patterns in Networks. *Physical Review E* 67, 2 (Feb. 2003), 026126. doi:10.1103/PhysRevE.67.026126
- [12] Robert Rai, Rajesh Sharma, and Chandrakala Meena. 2025. IPSR Model: Misinformation Intervention through Prebunking in Social Networks. *Authorea Preprints* (2025).
- [13] Akрати Saxena, George Fletcher, and Mykola Pechenizkiy. 2024. Fairsna: Algorithmic fairness in social network analysis. *Comput. Surveys* 56, 8 (2024), 1–45.
- [14] Akрати Saxena, Gaurav Kumar, and Chandrakala Meena. 2025. Homophily in Complex Networks: Measures, Models, and Applications. *arXiv preprint arXiv:2509.18289* (2025).
- [15] Charles Stewart III and Jonathan Woon. 2017. Congressional Committee Assignments, 103rd to 114th Congresses, 1993–2017.
- [16] Lev Telyatnikov, Maria Sofia Bucarelli, Guillermo Bernardez, Olga Zaghen, Simone Scardapane, and Pietro Lio. 2023. Hypergraph Neural Networks through the Lens of Message Passing: A Common Perspective to Homophily and Architecture Design. <https://arxiv.org/abs/2310.07684v3>.
- [17] Nate Veldt, Austin R. Benson, and Jon Kleinberg. 2023. Combinatorial Characterizations and Impossibilities for Higher-Order Homophily. *Science Advances* 9, 1 (Jan. 2023), eabq3200. doi:10.1126/sciadv.abq3200