Revisiting Graph Homophily Measures

Mikhail Mironov
Yandex Research
mironov.m.k@gmail.com

Liudmila Prokhorenkova
Yandex Research
ostroumova-la@yandex-team.ru

Abstract

Homophily is a graph property describing the tendency of edges to connect similar nodes. There are several measures used for assessing homophily but all are known to have certain drawbacks: in particular, they cannot be reliably used for comparing datasets with varying numbers of classes and class size balance. To show this, previous works on graph homophily suggested several properties desirable for a good homophily measure, also noting that no existing homophily measure has all these properties. Our paper addresses this issue by introducing a new homophily measure — *unbiased homophily* — that has all the desirable properties and thus can be reliably used across datasets with different label distributions. The proposed measure is suitable for undirected (and possibly weighted) graphs. We show both theoretically and via empirical examples that the existing homophily measures have serious drawbacks while unbiased homophily has a desirable behavior for the considered scenarios. Finally, when it comes to directed graphs, we prove that some desirable properties contradict each other and thus a measure satisfying all of them cannot exist.

1 Introduction

Graphs serve as a natural data structure in many areas: social networks, molecules, databases, road traffic, citation networks, etc. Graph nodes often come with labels that characterize them and divide them into several classes. For instance, in a molecular graph, these labels could represent different types of atoms, while in a citation network, they might indicate the scientific disciplines to which each paper belongs. Based on the relationship between graph edges and node labels, graphs are typically categorized as either homophilic or heterophilic [1]. A graph is called *homophilic* if its nodes are more likely to connect with nodes having the same label. For example, a citation network where papers are annotated with their research areas is a homophilic graph since papers tend to cite other papers from the same research area. In contrast, a graph is called *heterophilic* if its nodes are more likely to be connected to nodes with different labels. An example of a heterophilic graph is a supply chain network where suppliers tend to connect to manufacturers but not to other suppliers.

Graph neural networks (GNNs) are the primary modern machine learning tools for working with graphs. The performance of GNNs can vary depending on whether they are applied to homophilic or heterophilic graphs. Earlier studies on GNNs primarily concentrated on homophilic graphs; however, there is a growing debate regarding the necessity of developing specialized models tailored for heterophilic settings [2–5]. Thus, the notion of homophily and heterophily is important for machine learning on graphs, and it is important to be able to measure how homophilic or heterophilic each graph is.

A *homophily measure* is a graph characteristic that indicates the level of homophily of a given graph. For undirected graphs, many homophily measures were constructed in previous works, e.g., edge homophily, node homophily, class homophily, and adjusted homophily. Since these measures may often disagree on which graph is more homophilic, being able to decide which measure is more reliable is important. In a recent paper, Platonov et al. [6] propose the following approach for comparing different homophily measures: first, formulate desirable properties that a reliable measure is expected to satisfy, and then check what measures have these properties. The authors formulate

Mikhail Mironov and Liudmila Prokhorenkova, Revisiting Graph Homophily Measures. *Proceedings of the Third Learning on Graphs Conference (LoG 2024)*, PMLR 269, Virtual Event, November 26–29, 2024.

five such desirable properties. In a nutshell, these properties set the range of a homophily measure values, ensure that homophily increases when we add a homophilic or remove a heterophilic edge, and guarantee that the measure is not biased towards a particular number of classes or their size balance. Platonov et al. [6] prove that none of the known homophily measures has all five desirable properties and recommend using adjusted homophily since it dominates all other measures. Finally, they pose an open question regarding the existence of a measure that has all five properties.

We solve this open problem by proposing a new homophily measure — *unbiased homophily*. It has all five desirable properties and thus is not biased towards a particular number of classes or their size balance. Hence, unbiased homophily can be reliably used across different datasets. To additionally illustrate this, we conduct experiments showing that all known measures have their failure cases, while unbiased homophily works well in all the considered scenarios. The proposed measure can be applied to any undirected graphs, including weighted with non-negative weights.

Then, we extend our research to directed graphs. We note that all the considered desirable properties of a homophily measure can be naturally formulated for the directed case. However, we prove that in this case, the desirable properties contradict each other, thus no homophily measure for directed graphs can have all five of them. This opens a direction for future research — we need to rethink the list of desirable properties for the directed case and find a reliable measure.

To sum up, our paper answers the open question of whether there exists a homophily measure satisfying all the desirable properties: positively for undirected and negatively for directed graphs.

2 Background on Homophily Measures

Our work focuses on *homophily measures* that quantify how well graph structure agrees with node labels. In a homophilic graph, edges tend to connect nodes with the same label. In graph analysis literature, *homophily* is usually referred to as *label assortativity* (we refer to Appendix A for a discussion of assortativity measures).

Let us now define homophily measures used in previous studies. We start with notation. Suppose we are given a graph $G = (V, E)$ with nodes V, $|V| = n$, and edges E. For now, we assume that G is undirected but may include self-loops and multiple edges. For simplicity, we assume that G is unweighted, but all the results and measures easily generalize to the weighted graphs by replacing edge indicators with edge weights. Each node v has a class label $y_v \in \{1, \dots, m\}$. We denote by $d(v)$ the degree of v and by $N(v)$ the multiset of neighbors of v, that is $d(v) = |N(v)|$. We denote by n_k the size of the k-th class, i.e., $n_k = |\{v : y_v = k\}|$. Finally, by D_k we denote the total degree of the k-th class, i.e., $D_k = \sum\limits_{v:y_v=k} d(v)$.

Edge homophily [2, 7] is simply the fraction of homophilic edges:

$$h_{edge} = \frac{|\{\{u, v\} \in E : y_u = y_v\}|}{|E|}.$$

Node homophily [8] computes how homophilic each node is and then averages the values over all nodes:

$$h_{node} = \frac{1}{n} \sum_{v \in V} \frac{|\{u \in N(v) : y_u = y_v\}|}{d(v)}.$$

Class homophily [9] sums positive excess homophily in every class:

$$h_{class} = \frac{1}{m-1} \sum_{k=1}^{m} \left[\frac{\sum\limits_{v:y_v=k} |\{u \in N(v) : y_u = y_v\}|}{D_k} - \frac{n_k}{n} \right]_+,$$

where $[x]_+ = \max\{x, 0\}$. The factor $\frac{1}{m-1}$ scales h_{class} to the interval $[0, 1]$.

Adjusted homophily [6] (also known as *assortativity coefficient* [10]) is defined as:

$$h_{adj} = \frac{h_{edge} - \sum\limits_{k=1}^{m} D_k^2/(2|E|)^2}{1 - \sum\limits_{k=1}^{m} D_k^2/(2|E|)^2}. \tag{1}$$

Here the numerator is the difference between the observed fraction of homophilic edges and its expected value assuming that the edge endpoints are connected randomly. The denominator scales the obtained value so that the maximum achievable homophily equals one. Platonov et al. [6] recommend using adjusted homophily instead of other homophily measures since it is the only measure that is unbiased towards particular class size distributions (as it has the constant baseline property, see Section 4 for the discussion).

To outline the scope of the paper, we first note that our work solely focuses on graph-label relations. In other words, we do not consider node features to be a part of a homophily measure. Also, in graph machine learning literature, the term *homophily* is sometimes used with a different meaning [4, 11]: that a good homophily measure is expected to correlate well with GNN performance. However, Platonov et al. [6] suggest separating these concepts since both highly homophilic and highly heterophilic datasets are easy for GNNs and thus other types of measures should be applied for evaluating the simplicity of a dataset for GNNs [4, 11, 12]. In this work we also stick to the standard definition of a homophily measure: it evaluates whether similar nodes are connected. In particular, this implies that we do not evaluate and compare homophily measures based on how well they correlate with GNN performance. Instead, we rely on the theoretical approach based on properties that a good homophily measure is expected to satisfy.

3 Inconsistency of Homophily Measures

It turns out that homophily measures discussed in the previous section are often inconsistent with each other. That is, given two graphs G_1 and G_2, the graph G_1 can be more homophilic than G_2 according to one homophily measure, while G_2 more homophilic than G_1 according to another homophily measure.

To illustrate this inconsistency, we conducted an experiment on the AIDS molecular dataset [13]. This dataset contains molecular graphs and we consider the types of atoms as node classes. We sampled 1000 pairs of graphs. For each pair and each homophily measure, either the first graph is more homophilic, the second graph is more homophilic, or they are equally homophilic according to this measure. Thus, for two homophlily measures, we can count the percentage of pairs where the measures agree on what graph is more homophilic. The results are shown in Table 1. As we can see, while edge and node homophily usually agree with each other, the consistency of all other pairs of measures is below 75%, which means that in many cases the measures disagree on which graph is more homophilic.

Table 1: Agreement of homophily measures on the AIDS dataset

	h_{edge}	h_{node}	h_{class}	h_{adj}
h_{edge}	-	91%	60%	73%
h_{node}	91%	-	59%	74%
h_{class}	60%	59%	-	71%
h_{adj}	73%	74%	71%	-

We conducted similar experiments for the PROTEINS and MUTAG molecular datasets [13], synthetic graphs, and real node classification datasets, see Appendix B. Interestingly, for the MUTAG dataset, the agreement of class homophily with the other measures is below 50% which means that in most of the cases, the measures are inconsistent. These experiments clearly demonstrate the disagreements between existing homophily measures and thus motivate our study on the comparative analysis of these measures.

4 Desirable Properties of Homophily Measures

Our experiments in the previous section show that known homophily measures are often inconsistent. So how to choose which ones to trust? In other words, can we say that some measures are better at quantifying the level of homophily? To answer this question, we follow previous studies [6] and formulate what properties a good homophily measure should have. Then, by checking which measures satisfy which properties we can compare these measures.

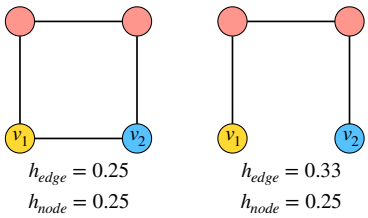
In this section, we motivate and discuss such desirable properties. All of these properties (excluding class symmetry) were introduced by Platonov et al. [6]. We give an intuitive explanation for each property in this section and refer to Appendix C for formal definitions and deeper discussion. To illustrate why each property is needed, we motivate it with a simple example showing desirable

and undesirable behavior of the existing homophily measures. Further in the text, we call an edge *homophilic* if it connects nodes with the same label and *heterophilic* otherwise.

Property 1. *Monotonicity* requires that if we add a homophilic edge or remove a heterophilic edge, homophily of a graph must increase.

This property is natural to require as it formalizes the meaning of the term 'homophily'.

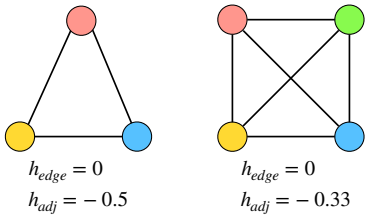
Example. Consider a graph G and two of its nodes v_1 and v_2 that are connected and belong to different classes. Assume that v_1 and v_2 are fully heterophilic, i.e., are not connected to any node of its class. Then, deleting the edge between v_1 and v_2 does increase edge homophily, but does not change node homophily, see the illustration on the right. We consider the behavior of node homophily to be undesirable in this case.



A more problematic situation is when deleting a heterophilic edge leads to a *decrease* of a homophily measure. We show such an example for adjusted homophily in Appendix F.

Property 2. *Minimal agreement* requires that graphs with only heterophilic edges achieve constant lower bound R_{min} of the homophily measure.

Example. Consider graphs G_1 with 3 nodes and G_2 with 4 nodes. Both graphs are complete and the labels of all nodes are different. Clearly, both G_1 and G_2 are fully heterophilic and we expect a homophily measure to have the lowest possible (and equal) values on these graphs. Edge homophily, node homophily, and class homophily do just that: they take the value 0 on both these graphs. Yet adjusted homophily takes the value -0.50 on G_1 and -0.33 on G_2, which we consider as undesirable behavior.



Property 3. *Maximal agreement* requires that graphs with only homophilic edges achieve constant upper bound R_{max} of the homophily measure.

Similarly to minimal agreement, this property is natural to require for any good homophily measure. We note that all measures listed in Section 2 have maximal agreement.

Property 4. *Empty class tolerance* requires that if we add a new dummy label that is not present in the graph, the graph homophily does not change.

Example. Consider a graph and suppose we add a new dummy label that is not present in the graph. That is, the graph does not change, but the formal number of classes m increases by 1. We expect graph homophily to not change. This property is essential for being able to compare the values of a homophily measure for datasets with different numbers of classes. Clearly, edge homophily, node homophily, and adjusted homophily are not affected by the addition of a dummy label. However, class homophily becomes undefined when we make such an addition since we get $D_k = 0$ in the denominator of the corresponding term.

Property 5. *Constant baseline* requires that if the graph structure is independent of node labels, the homophily measure should be equal to some constant R_{base}. We formalize and discuss this property in more detail in Appendix C.

Example. Suppose we have two graphs G_1 and G_2 with 100 nodes each. Each graph has two classes, where the class balance of G_1 is $50 : 50$, and the class balance of G_2 is $98 : 2$. Suppose we draw edges according to the Erdős–Rényi model (with self-loops): every two nodes are connected with probability 0.5 independently of their labels and other edges. Such graphs should be considered neutral with respect to homophily, i.e., neither homophilic nor heterophilic. Thus, we expect both graphs to have the same homophily value (in expectation). The expected value of adjusted homophily is 0 for both G_1 and G_2, which we consider as desirable behavior. Yet, the expected value of edge homophily is 0.5 for G_1 and 0.9608 for G_2, which we clearly consider as undesirable behavior.

Finally, let us introduce an additional simple property.

Property 6. *Class symmetry* requires that if we reorder (or rename) the classes, the value of a homophily measure should not change.

Note that all measures listed in Section 2 have class symmetry.

The problem raised in previous studies [6] is that no known measure has all the above properties. Indeed, only adjusted homophily has constant baseline, but it is not monotone for some values of h_{adj}, which is a critical drawback. In the next section, we construct a measure that has all the properties.

5 Homophily Measure Having All Desirable Properties

In this section, we construct a measure that has all the desirable properties discussed above. Moreover, we require one additional property — continuity — and we explain why it is necessary below. To formulate this property, we need to first define edge-wise scale-invariant measures.

5.1 Edge-Wise and Scale-Invariant Measures

In this section, we say that a *homophily measure* is a function h from the set of all undirected unweighted graphs with labeled nodes to $\mathbb{R}$. A homophily measure is called *edge-wise* if it is a function of the *class adjacency matrix* that we define below.

Definition 1. For a graph G, we define an $m \times m$ *class adjacency matrix* L_G as follows:

$$l_{ii} = 2|\{\{u,v\} \in E : y_u = y_v = i\}|\,,$$
$$l_{ij} = |\{\{u,v\} \in E : \{y_u, y_v\} = \{i,j\}\}| \text{ for } i \neq j\,.$$

In other words, an edge-wise homophily measure depends on the number of edges between nodes of each pair of classes (or within one class), but it does not take into account how the edges are distributed among individual nodes. Considering the measures listed in Section 2, edge homophily and adjusted homophily are edge-wise.

Definition 2. An edge-wise homophily measure h is called *scale-invariant* if $h(L_G) = h(L_{G'})$ for any graphs G, G' such that $L_G = \beta L_{G'}$ for some $\beta > 0$.

Note that edge homophily and adjusted homophily are scale-invariant.

By definition, edge-wise scale-invariant homophily measures maintain their value under scaling transformations of the class adjacency matrix L_G. Let us then normalize L_G as follows.

Definition 3. For a graph G, we define $m \times m$ *normalized class adjacency matrix* $C_G = \frac{1}{2|E|} L_G$:

$$c_{ii} = \frac{|\{\{u,v\} \in E : y_u = y_v = i\}|}{|E|}\,,$$
$$c_{ij} = \frac{|\{\{u,v\} \in E : \{y_u, y_v\} = \{i,j\}\}|}{2|E|} \text{ for } i \neq j\,.$$

For any nonempty graph G, its matrix C_G is symmetric, non-negative, and its elements sum to 1.

Clearly, every edge-wise scale-invariant homophily measure can be viewed as a function of the normalized class adjacency matrix C_G. Sometimes (when it is clear from the context), we omit the index G and denote the normalized class adjacency matrix C_G by C. Also, we further assume that there are at least two non-zero elements in C, that is, we consider graphs with at least two non-degenerate classes.

For edge-wise scale-invariant homophily measures, we require one more property to hold (in addition to those listed in Section 4).

Property 7. *Continuity* requires an edge-wise scale-invariant homophily measure h to be a continuous function of C.

This property ensures that small changes in the graph structure (i.e., in the normalized class adjacency matrix) result in small changes in the homophily measure, which is a natural requirement. Note that edge homophily and adjusted homophily are continuous. To show the importance of this property, in Appendix D we give an example of a homophily measure that has all desirable properties excluding continuity but demonstrates unwanted behavior, thus proving that continuity must be required.

Formal definition of Properties 1–6. We formally define all the properties in terms of the normalized class adjacency matrix in Appendix C.

5.2 Unbiased Homophily

In this subsection, we construct a homophily measure that satisfies all the desirable properties. We name this new measure *unbiased homophily*, since by having all the desirable properties it is not biased towards a particular number of classes or their size balance. Hence, unbiased homophily can be reliably used across different datasets: the value of the measure is calibrated for fully homophilic, fully heterophilic, and fully randomized graphs and behaves as desired (due to monotonicity) between these values.[1]

We define *unbiased homophily* as:

$$h^{\alpha}_{unb}(C) := \frac{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} - c_{ij})}{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} + c_{ij})} + \alpha \min\left(\sum_i \sqrt{c_{ii}}, 1\right), \tag{2}$$

where $\alpha > 0$ is any positive constant. Note that the denominator of the first term is always greater than zero, thus h^{α}_{unb} is well-defined.

Interpretation. First, consider the first term of (2) since, as discussed below, it is the main ingredient of the measure. Note that if edges were drawn independently of classes, we expect c_{ij} to be equal to $\sqrt{c_{ii}c_{jj}}$ (see the definition of $\mathrm{rand}(C)$ in Appendix C for the details). So, $\sqrt{c_{ii}c_{jj}} - c_{ij}$ can be interpreted as the difference between the expected and observed fraction of heterophilic edges between the classes i and j (up to a factor of 2 since the fraction of such heterophilic edges equals $2c_{ij}$). Summing such differences for all pairs of classes, we interpret the numerator as the difference between the expected and observed fraction of heterophilic edges in the graph (again up to a factor of 2). The denominator normalizes the obtained value to the interval [-1;1]: when a graph is fully homophilic (i.e., $c_{ij} = 0\ \forall i, j$) the first term equals $\frac{\sum\limits_{i<j}\sqrt{c_{ii}c_{jj}}}{\sum\limits_{i<j}\sqrt{c_{ii}c_{jj}}} = 1$, while for fully heterophilic graphs we get $\frac{\sum\limits_{i<j}-c_{ij}}{\sum\limits_{i<j}c_{ij}} = -1$.

As we discuss below, the second term $\alpha \min\left(\sum\limits_i \sqrt{c_{ii}}, 1\right)$ is needed to cover some rare special cases, and in practice, we advise using $\alpha = 0$. So essentially h^{α}_{unb} should be interpreted as the scaled difference between the expected and observed fraction of heterophilic edges.

Theoretical guarantees. The following theorem holds (the proof can be found in Appendix E).

Theorem 5.1. *For $\alpha > 0$, the measure $h^{\alpha}_{unb}(C)$ is continuous, has all the desirable properties listed in Section 4, and its values $R_{max}, R_{base}, R_{min}$ are equal to $1+\alpha, \alpha, -1$, respectively.*

Note that the desirable properties of unbiased homophily (2) hold for any $\alpha > 0$. To apply this measure in practice, one has to choose the value of α, since different choices may potentially give different results. As we show in the proof of the theorem, the first term of (2) by itself satisfies all the desirable properties in most cases, and the second term is added to resolve sensitivity issues in some rare special cases. Thus, in practice we recommend taking $\alpha = 0$ as the default value and use:

$$h_{unb}(C) := h^{0}_{unb}(C) = \frac{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} - c_{ij})}{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} + c_{ij})}. \tag{3}$$

Theorem 5.2. *The measure $h_{unb}(C)$ is continuous and differentiable and it has all the desirable properties listed in Section 4 except some special cases that only occur when there is at most one non-zero element on the diagonal (i.e., only one class has intra-edges). The values $R_{max}, R_{base}, R_{min}$ are equal to $1, 0, -1$, respectively.*

[1] Note that we do not claim that unbiased homophily is the only possible homophily measure having all the desirable properties. Our goal was to construct at least one measure satisfying these conditions since all the measures from previous studies do not have at least one property.

Table 2: Properties of homophily measures; ✓* means that the property is satisfied always except a rare special case when only one class has intra-edges

	continuity	max agreement	min agreement	const baseline	monotonicity	empty class tolerance	class symmetry
h_{edge}	✓	✓	✓	✗	✓	✓	✓
h_{node}	n/a	✓	✓	✗	✓	✓	✓
h_{class}	n/a	✓	✗	✗	✗	✗	✓
h_{adj}	✓	✓	✗	✓	✗	✓	✓
h^{α}_{unb}	✓	✓	✓	✓	✓	✓	✓
h_{unb}	✓	✓	✓*	✓	✓*	✓	✓

The proof of this theorem follows from the proof of Theorem 5.1. Note that h_{unb} is differentiable, while h^{α}_{unb} with $\alpha > 0$ is not differentiable at the points where $\sum_i \sqrt{c_{ii}} = 1$.

Thus, unbiased homophily h_{unb} has all the required properties in all cases excluding the case when there is at most one non-zero element on the diagonal (when minimal agreement and monotonicity do not hold). This case rarely appears in real datasets. So, we argue that the simplicity of h_{unb} (in comparison to h^{α}_{unb}) outweighs the fact that h_{unb} has slightly weaker theoretical properties. That is why we recommend using unbiased homophily h_{unb} as a homophily measure in practice. See Table 2 for the list of homophily measures and their properties.

Computing unbiased homophily. The formula (3) is easy to interpret, but there is also an equivalent formula for h_{unb} that is more convenient for computing as it requires only the number of intra-class edges for each class instead of summation over the pairs of classes:

$$h_{unb}(C) = \frac{\left(\sum_i \sqrt{c_{ii}}\right)^2 - 1}{\left(\sum_i \sqrt{c_{ii}}\right)^2 + 1 - 2\sum_i c_{ii}}. \tag{4}$$

We show the equivalence of (3) and (4) in Appendix G.

Unbiased homophily of weighted graphs. For simplicity of the presentation, all the desirable properties were formulated for unweighted graphs. However, they can be easily rewritten for graphs with positive weights. To do so, in Definition 1 of the class adjacency matrix, instead of the number of edges we use the total weight of the corresponding edges. In Definition 3 of the normalized class adjacency matrix, we replace $|E|$ by the total weight of all edges. Note that the sets of normalized class adjacency matrices in the unweighted and weighted cases are the same (up to the transition from rational to real coefficients by continuity in the unweighted case). Thus, for the weighted case, we can define all properties (see Appendix C) in the same way as in the unweighted case. The measures h^{α}_{unb} and h_{unb} for the weighted case have the same properties as for the unweighted case. So, we can use them to measure homophily of weighted graphs.

6 Empirical Comparison

In this section, we empirically compare unbiased homophily with existing homophily measures. First, we consider synthetic examples that demonstrate unwanted behavior of the existing measures, while the results of unbiased homophily are as desired. Then, we compare the level of homophily for different measures on real graph datasets.

6.1 Synthetic Examples

In this section, we compare the behavior of homophily measures for simple synthetic datasets.

Example 1. The first graph G_1 is generated by the Erdős–Rényi model with $p = 0.5$, it has 2 classes and the class size balance is $90 : 10$.

The second graph G_2 is generated by the stochastic block model with intra-class edge probability 0.3 and inter-class probability 0.2, with 2 classes and class size balance $50:50$.

Clearly, G_1 is neutral w.r.t. homophily since graph edges are independent of node labels. In contrast, G_2 exhibits some homophily since intra-class edges have a higher probability. Thus, we expect a good homophily measure to take a higher value on G_2. The results for different homophily measures are listed in Table 3. As we see, edge and node homophily exhibit undesirable behavior for this example.

Table 3: ER with class imbalance vs SBM

	G_1, ER		G_2, SBM
h_{edge}	0.82	>	0.6
h_{node}	0.82	>	0.6
h_{class}	0	<	0.05
h_{adj}	0	<	0.2
h_{unb}	0	<	0.38

Example 2. Both graphs G_1 and G_2 are complete graphs on 6 nodes, but G_1 has one node in each of 6 classes, while G_2 has two nodes in each of 3 classes.

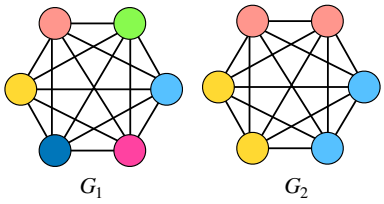


Figure 1: Visualization of G_1 and G_2 in Example 2

Table 4: Homophily values for Example 2

	G_1		G_2
h_{edge}	0	<	0.2
h_{node}	0	<	0.2
h_{class}	0	=	0
h_{adj}	-0.2	=	-0.2
h_{unb}	-1	<	-0.33

Clearly, G_1 is fully heterophilic, while G_2 exhibits some homophily, so we expect a good homophily measure to take a higher value on G_2. The results for different homophily measures are listed in Table 4. As we see, class and adjusted homophily exhibit undesirable behavior.

We see that only unbiased homophily behaves as desired for both of these simple examples. This superiority of unbiased homophily follows from the fact that it satisfies all the desirable properties.

6.2 Homophily of Real Datasets

We measure edge, node, class, adjusted, and unbiased homophily for various real datasets that are often used in graph machine learning literature. The description of the datasets can be found in Appendix J. The results are shown in Table 5.

As expected, node and edge homophily often take higher values, especially on datasets with two classes (that are imbalanced). For instance, for the minesweeper dataset, edge and node homophily are equal to 0.68, while all the other homophily measures indicate that the homophily level is close to zero (below 0.02 for all the other measures).

In turn, h_{class}, h_{adj}, and h_{unb} mostly agree with each other on how homophilic a dataset is. The possible reason for this agreement is that all three measures have been motivated by the constant baseline property. Importantly, h_{adj} and h_{unb} always agree on whether the homophily value is positive or negative since they do satisfy constant baseline.

One can see that the proposed unbiased homophily usually takes higher values than h_{adj} when homophily is positive. For instance, for the amazon-computers dataset, $h_{unb} = 0.91$ and $h_{adj} = 0.68$. Another interesting observation is that for the most heterophilic dataset roman-empire, $h_{adj} = -0.05$ while $h_{unb} = -0.49$, indicating that this dataset is significantly more heterophilic than reported before. This dataset is expected to be quite heterophilic: here the nodes are words, the labels are their syntactic roles, and two nodes are connected if either these words follow each other in the text, or they are connected in the dependency tree of the sentence. This dataset has 18 classes but less than 5% homophilic edges. Adjusted homophily cannot predict the low score for this dataset since h_{adj} does not satisfy minimal agreement. In contrast, h_{unb} is able to predict a low score.

The difference between edge and unbiased homophily can be illustrated on the pubmed and coauthors-cs datasets. For these datasets, h_{edge} is nearly the same (0.80 and 0.81), but h_{unb} differs significantly: 0.77 for pubmed and 0.96 for coauthors-cs. This difference arises because pubmed has 3 classes, while coauthors-cs has 15, with relatively balanced class distributions. Unbiased homophily accounts

Table 5: Homophily of various datasets

Dataset	n	$\|E\|$	m	h_{edge}	h_{node}	h_{class}	h_{adj}	h_{unb}
lastfm-asia	7 624	27 806	18	0.8739	0.8332	0.7656	0.8562	0.9746
coauthor-cs	18 333	81 894	15	0.8081	0.8320	0.7547	0.7845	0.9582
coauthor-physics	34 493	247 962	5	0.9314	0.9153	0.8474	0.8724	0.9480
amazon-photo	7 650	119 081	8	0.8272	0.8493	0.7722	0.7850	0.9329
cora	2 708	5 278	7	0.8100	0.8252	0.7657	0.7711	0.9200
amazon-computers	13 752	245 861	10	0.7772	0.8017	0.7002	0.6823	0.9055
facebook	22 470	170 823	4	0.8853	0.8834	0.8195	0.8206	0.9033
citeseer	3 327	4 552	6	0.7355	0.7166	0.6267	0.6707	0.8520
pubmed	19 717	44 324	3	0.8024	0.7924	0.6641	0.6860	0.7653
github	37 700	289 003	2	0.8453	0.8011	0.3778	0.3778	0.4969
amazon-ratings	24 492	93 050	5	0.3804	0.3757	0.1266	0.1402	0.2769
flickr	89 250	449 878	7	0.3195	0.3221	0.0698	0.0941	0.2264
twitch-de	9 498	153 138	2	0.6322	0.5958	0.1394	0.1394	0.1566
twitch-pt	1 912	31 299	2	0.5708	0.5949	0.1196	0.1069	0.1104
tolokers	11 758	519 000	2	0.5945	0.6344	0.1801	0.0926	0.1016
questions	48 921	153 540	2	0.8396	0.8980	0.0790	0.0207	0.0583
deezer-europe	28 281	92 752	2	0.5251	0.5299	0.0304	0.0304	0.0310
minesweeper	10 000	39 402	2	0.6828	0.6829	0.0094	0.0094	0.0145
actor	7 600	26 659	5	0.2167	0.2199	0.0064	0.0028	0.0053
genius	421 961	922 868	2	0.5932	0.5087	0.0229	-0.0527	-0.0705
roman-empire	22 662	32 927	18	0.0469	0.0460	0.0208	-0.0468	-0.4913

for the fact that in coauthors-cs, having 80% homophilous edges reflects much stronger homophily than in pubmed, due to different class distribution.

7 Homophily for Directed Graphs

Above, we have only discussed homophily for undirected graphs. The natural next step would be to extend our analysis to directed graphs. The study of edge-wise homophily measures in the directed case is important since these measures have a one-to-one correspondence with classification evaluation measures [14]. Thus, if we find a good homophily measure for directed graphs, we also get a good classification evaluation measure that would allow us to compare classification performance for datasets with different numbers of classes.

In Appendix I, we reformulate the desirable properties of homophily measures for directed graphs. Unfortunately, it turns out that in the directed case the properties do contradict each other, which we prove in Appendix I. Thus, no measure satisfying all of them can be constructed. We suggest steps towards modification of these properties to avoid contradictions in the last subsection of Appendix I.

8 Conclusion

In this paper, we address the problem of how to choose a reliable homophily measure. We show via a series of examples that previously used measures are flawed and construct a new one — unbiased homophily — that behaves as desired in all the considered scenarios. We theoretically prove that the proposed measure satisfies the desirable properties introduced in previous studies and some additional ones. Thus, we solve the open problem of whether there exists a homophily measure satisfying all these properties. The proposed measure can be applied to any undirected graphs, and we recommend using it to estimate and compare homophily levels of various graphs in future works. We believe that the proposed measure will become a useful tool for researchers and practitioners.

In contrast, for directed graphs, we prove that the desirable properties contradict each other, thus no homophily measure can have all the desirable properties. This opens a direction for future research — constructing a list of non-contradicting properties and finding the best measure in terms of these properties. Another interesting direction for future studies is to theoretically analyze homophily measures for higher-order relationship networks, such as hypergraphs or simplicial complex networks [15].

References

[1] Sitao Luan, Chenqing Hua, Qincheng Lu, Liheng Ma, Lirong Wu, Xinyu Wang, Minkai Xu, Xiao-Wen Chang, Doina Precup, Rex Ying, et al. The heterophilic graph learning handbook: Benchmarks, models, theoretical analysis, applications and challenges. *arXiv preprint arXiv:2407.09618*, 2024. 1

[2] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems*, 33:7793–7804, 2020. 1, 2

[3] Yao Ma, Xiaorui Liu, Neil Shah, and Jiliang Tang. Is homophily a necessity for graph neural networks? In *International Conference on Learning Representations*, 2022.

[4] Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Revisiting heterophily for graph neural networks. *Advances in neural information processing systems*, 35:1362–1375, 2022. 3

[5] Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of gnns under heterophily: Are we really making progress? *arXiv preprint arXiv:2302.11640*, 2023. 1, 12, 22

[6] Oleg Platonov, Denis Kuznedelev, Artem Babenko, and Liudmila Prokhorenkova. Characterizing graph datasets for node classification: Homophily-heterophily dichotomy and beyond. In *Advances in Neural Information Processing Systems*, volume 36, pages 523–548. Curran Associates, Inc., 2023. 1, 2, 3, 5

[7] Sami Abu-El-Haija, Bryan Perozzi, Amol Kapoor, Nazanin Alipourfard, Kristina Lerman, Hrayr Harutyunyan, Greg Ver Steeg, and Aram Galstyan. Mixhop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In *international conference on machine learning*, pages 21–29. PMLR, 2019. 2

[8] Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020. 2, 22

[9] Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser Nam Lim. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. *Advances in Neural Information Processing Systems*, 34:20887–20902, 2021. 2

[10] Mark E.J. Newman. Mixing patterns in networks. *Physical review E*, 67(2), 2003. 2, 12

[11] Yilun Zheng, Sitao Luan, and Lihui Chen. What is missing in homophily? disentangling graph homophily for graph neural networks. *arXiv preprint arXiv:2406.18854*, 2024. 3

[12] Sitao Luan, Chenqing Hua, Minkai Xu, Qincheng Lu, Jiaqi Zhu, Xiao-Wen Chang, Jie Fu, Jure Leskovec, and Doina Precup. When do graph neural networks help with node classification? investigating the homophily principle on node distinguishability. *Advances in Neural Information Processing Systems*, 36, 2024. 3

[13] Kaspar Riesen and Horst Bunke. Graph database repository for graph based pattern recognition and machine learning. In *Structural, Syntactic, and Statistical Pattern Recognition*, 2008. 3, 13

[14] Martijn Gösgens, Anton Zhiyanov, Aleksey Tikhonov, and Liudmila Prokhorenkova. Good classification measures and how to find them. In *Advances in Neural Information Processing Systems*, volume 34, pages 17136–17147. Curran Associates, Inc., 2021. 9

[15] Arnab Sarker, Natalie Northrup, and Ali Jadbabaie. Higher-order homophily on simplicial complexes. *Proceedings of the National Academy of Sciences*, 121(12), 2024. 9

[16] Fariba Karimi and Marcos Oliveira. On the inadequacy of nominal assortativity for assessing homophily in networks. *Scientific Reports*, 13(1):21053, 2023. 12

[17] C. Lee Giles, Kurt D. Bollacker, and Steve Lawrence. CiteSeer: An automatic citation indexing system. In *Proceedings of the third ACM conference on Digital libraries*, pages 89–98, 1998. 22

[18] Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3:127–163, 2000.

[19] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

[20] Galileo Namata, Ben London, Lise Getoor, Bert Huang, and U Edu. Query-driven active surveying for collective classification. In *10th international workshop on mining and learning with graphs*, volume 8, page 1, 2012.

[21] Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *International conference on machine learning*, pages 40–48. PMLR, 2016. 22

[22] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. In *Relational Representation Learning Workshop*, 2018. 22

[23] Benedek Rozemberczki and Rik Sarkar. Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1325–1334, 2020. 22

[24] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2), 2021. 22

[25] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816, 2009. 22

[26] Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. GraphSAINT: Graph sampling based inductive learning method. In *International Conference on Learning Representations*, 2020. 22

[27] Derek Lim and Austin R. Benson. Expertise and dynamics within crowdsourced musical knowledge curation: A case study of the genius platform. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 373–384, 2021. 22

A Graph Assortativity Measures

In network science literature, there is a concept related to homophily that is called *assortativity*. Assortativity is a general concept that is often applied to *node degrees* and the level of assortativity is typically measured by *assortativity coefficient* [10]. When applied to node labels, the assortativity coefficient reduces to *adjusted homophily* [5].

In terms of the class adjacency matrix C, the assortativity coefficient (for undirected graphs) can be written as:

$$\mathrm{AC} = \frac{\sum_i c_{ii} - \sum_i \left(\sum_j c_{ij} \right)^2}{1 - \sum_i \left(\sum_j c_{ij} \right)^2}, \tag{5}$$

which is equivalent to (1). Thus, the assortativity coefficient (sometimes referred to as *nominal assortativity*) has maximal agreement and constant baseline, but it does not have minimal agreement and monotonicity.

Recently, Karimi and Oliveira [16] proposed a new measure which is called *adjusted nominal assortativity* that is designed to better handle cases when classes are size imbalanced. The formula of adjusted nominal assortativity can be obtained from (5) by dividing each coefficient c_{ij} by $f_i f_j$, where f_i is the fraction of nodes in class i. That is, the formula of adjusted nominal assortativity for undirected graphs is:

$$\mathrm{AC}_{adj} = \frac{\sum_i \frac{c_{ii}}{f_i f_i} - \sum_i \left(\sum_j \frac{c_{ij}}{f_i f_j} \right)^2}{1 - \sum_i \left(\sum_j \frac{c_{ij}}{f_i f_j} \right)^2}.$$

However, it turns out that scaling the coefficients of nominal assortativity by class sizes prevents the obtained measure from satisfying the desirable properties, as we prove below.

Maximal agreement. For fully homophilic graphs, the value of adjusted nominal assortativity is $\sum_i \frac{c_{ii}}{f_i f_i}$. Clearly, by fixing all c_{ii} and changing the number of nodes in classes, we can change the value of $\sum_i \frac{c_{ii}}{f_i f_i}$ (given that at least two of c_{ii} are non-zero).

Minimal agreement. Assume that all classes have the same number of nodes, that is, $f_i = \frac{1}{m}$. Then, adjusted nominal assortativity for a fully heterophilic graph becomes:

$$\frac{-\sum_i \left(\sum_j \frac{c_{ij}}{f_i f_j} \right)^2}{1 - \sum_i \left(\sum_j \frac{c_{ij}}{f_i f_j} \right)^2} = \frac{-\sum_i \left(\sum_j c_{ij} \right)^2 m^4}{1 - \sum_i \left(\sum_j c_{ij} \right)^2 m^4}.$$

Since for different fully heterophilic graphs the value of $\sum_i \left(\sum_j c_{ij} \right)^2$ can be different, the value of adjusted nominal assortativity can also be different.

Constant baseline. Assume that all classes have the same number of nodes, that is, $f_i = \frac{1}{m}$. Then, adjusted nominal assortativity becomes:

$$\frac{\sum_i c_{ii} - \sum_i \left(\sum_j c_{ij} \right)^2}{1 - \sum_i \left(\sum_j c_{ij} \right)^2} = \frac{\sum_i c_{ii}\, m^2 - \sum_i \left(\sum_j c_{ij} \right)^2 m^4}{1 - \sum_i \left(\sum_j c_{ij} \right)^2 m^4}.$$

Let $C = \mathrm{rand}(C)$ (see the definition of $\mathrm{rand}(C)$ in Appendix C for the details), that is, $c_{ii} = \left(\sum_j c_{ij}\right)^2$ for all i. Then,

$$\frac{\sum_i c_{ii}m^2 - \sum_i \left(\sum_j c_{ij}\right)^2 m^4}{1 - \sum_i \left(\sum_j c_{ij}\right)^2 m^4} = \frac{\sum_i c_{ii}m^2 - \sum_i c_{ii}m^4}{1 - \sum_i c_{ii}m^4} = \frac{\sum_i c_{ii}(m^2 - m^4)}{1 - \sum_i c_{ii}m^4}.$$

Since for different $\mathrm{rand}(C)$ the value of $\sum_i c_{ii}$ can be different, the value of adjusted nominal assortativity can also be different.

B Inconsistency of Homophily Measures

In this section, we report the results of additional experiments that are analogous to those in Section 3. As before, we sample 1000 pairs of graphs and for each pair of homophily measures, we count the percentage of pairs where the measures agree on what graph is more homophilic.

PROTEINS and MUTAG [13]. The results for these datasets are reported in Tables 6 and 7. The most notable is the inconsistency between class homophily and other measures in Table 7.

Table 6: Agreement of homophily measures on the PROTEINS dataset

	h_{edge}	h_{node}	h_{class}	h_{adj}
h_{edge}	-	96%	76%	81%
h_{node}	96%	-	76%	81%
h_{class}	76%	76%	-	92%
h_{adj}	81%	81%	92%	-

Table 7: Agreement of homophily measures on the MUTAG dataset

	h_{edge}	h_{node}	h_{class}	h_{adj}
h_{edge}	-	97%	30%	89%
h_{node}	97%	-	30%	88%
h_{class}	30%	30%	-	41%
h_{adj}	89%	88%	41%	-

Synthetic data. We generate synthetic graphs as follows. We sample the number of classes m uniformly at random from $[2, 10]$. Each graph has 100 nodes, nodes are numbered from 1 to 100. We sample $m - 1$ different numbers $a_1, \ldots, a_{m-1}$ from $[1, 99]$ and assign class i to nodes in $(a_{i-1}, a_i]$ (assuming $a_0 = 1$ and $a_m = 100$). For every pair of classes $i \leq j$ we sample p_{ij} uniformly at random from $[0, 1]$. For every pair of nodes in the graph we draw an edge between them with probability p_{ij} where i and j are classes of these nodes. As a result, we obtain graphs with different community structures and with varying numbers of classes and class size balance. The consistency of the homophily measures on this dataset is shown in Table 8.

Table 8: Agreement of homophily measures on the synthetic dataset

	h_{edge}	h_{node}	h_{class}	h_{adj}
h_{edge}	-	97%	67%	69%
h_{node}	97%	-	67%	68%
h_{class}	67%	67%	-	79%
h_{adj}	69%	68%	79%	-

Table 9: Agreement of homophily measures on the real datasets from Table 5

	h_{edge}	h_{node}	h_{class}	h_{adj}
h_{edge}	-	94%	83%	81%
h_{node}	94%	-	85%	82%
h_{class}	83%	85%	-	92%
h_{adj}	81%	82%	92%	-

Node classification datasets. We also conducted a similar measurement on the datasets from Table 5, i.e., we consider all pairs of datasets and check whether two homophily measures agree on which dataset is more homophilic. The results are shown in Table 9. Note that the obtained values are (in most cases) higher than in our previous experiments. This difference can be explained by the fact that the datasets in Table 5 are very different from each other and, in particular, may have very different homophily levels. When the difference in homophily is significant, it is easy for two measures to agree on which dataset is more homophilic. In contrast, in molecular datasets and synthetic data used for Table 8, all graphs are of the same nature and thus are more similar and we

expect more inconsistencies. In general, we expect such inconsistency results to highly depend on a dataset. However, note that in all the cases there are many examples of situations when any two measures are inconsistent.

C Formal Desirable Properties for Scale-Invariant Homophily Measures

Let us formally define a list of desirable properties a good homophily measure h is expected to satisfy. For convenience of further analysis, we formulate these properties only for edge-wise scale-invariant measures. Also for convenience, we split monotonicity property into two properties: homo-monotonicity and hetero-monotonicity.

Denote by $M_{\mathbb{Q}}$ the set of all matrices which are symmetric, non-negative, with *rational* elements that sum to 1 (and with at least two non-zero elements). Clearly, every element of $M_{\mathbb{Q}}$ corresponds to the normalized class adjacency matrix of some graph, and all matrices corresponding to graphs belong to $M_{\mathbb{Q}}$. Therefore, a homophily measure h can be considered as a function on $M_{\mathbb{Q}}$.

Continuity. A homophily measure h is a continuous function.

This property ensures that small changes in the graph structure (i.e., in the normalized class adjacency matrix) result in small changes in the homophily measure, which is a natural requirement. Note that all measures listed in Section 2 are continuous. To show the importance of this property, in Appendix D we give an example of a homophily measure that has all desirable properties excluding continuity but demonstrates unwanted behavior, thus proving that continuity must be required.

Denote by $M_{\mathbb{R}}$ the set of all symmetric non-negative matrices with *real* elements that sum to 1 (and with at least two non-zero elements). Since $M_{\mathbb{Q}}$ is a dense subset of $M_{\mathbb{R}}$, any continuous function h on $M_{\mathbb{Q}}$ can be uniquely extended to a continuous function on $M_{\mathbb{R}}$. Hence, if a homophily measure h is continuous, it can be uniquely extended to a continuous function on $M_{\mathbb{R}}$. Thus, we further assume that continuous homophily measures are defined on $M_{\mathbb{R}}$, i.e., C can be any element of $M_{\mathbb{R}}$.

Maximal agreement. There exists $R_{max} \in \mathbb{R}$ such that $h(C) \leq R_{max}$ and equality holds iff $\sum\limits_{i=1}^{m} c_{ii} = 1$. In other words, if all edges of a graph are homophilic, then $h(C) = R_{max}$, and if a graph has at least one heterophilic edge, then $h(C) < R_{max}$.

Minimal agreement. There exists $R_{min} \in \mathbb{R}$ such that $h(C) \geq R_{min}$ and equality holds iff $\sum\limits_{i=1}^{m} c_{ii} = 0$. In other words, if all edges of a graph are heterophilic, then $h(C) = R_{min}$, and if a graph has at least one homophilic edge, then $h(C) > R_{min}$.

Constant baseline. This property ensures that a homophily measure is not biased towards graphs with small or large number of classes or particular class size distributions. To achieve this, the property requires that if a graph structure is independent of labels, then homophily should be equal to some constant R_{base}.

To formalize this property, for every C_G, we construct a new normalized class adjacency matrix $\mathrm{rand}(C_G)$, corresponding to a graph whose structure is independent of labels, while class degrees distribution is the same as for C_G. After that, we require $\forall C_G : h(\mathrm{rand}(C_G)) = R_{base}$.

Definition 4. For a normalized class adjacency matrix C_G, let $a_i := \sum\limits_{k=1}^{m} c_{ik} = \frac{D_i}{2|E|}$. Then, we define $\mathrm{rand}(C_G)$ as: $\mathrm{rand}(C_G)_{ij} := a_i a_j$.

Let us motivate this definition. Suppose we fix the degrees and labels of all nodes in G and redraw edges between the nodes uniformly at random. That is, every half-edge has an equal probability to form an edge with any other half-edge. It is easy to see that the expected fraction of homophilic edges in class i is $\frac{D_i}{2|E|} \cdot \frac{D_i}{2|E|} = a_i^2$ (up to a negligible term that is usually ignored). Similarly, the expected fraction of heterophilic edges between classes i and j is $2 \cdot \frac{D_i}{2|E|} \cdot \frac{D_j}{2|E|} = 2a_i a_j$. Thus, the elements of $\mathrm{rand}(C_G)$ correspond to expected fractions of edges of each type under the procedure when we randomly redraw all edges in a graph G. Note that column-wise and row-wise sums of $\mathrm{rand}(C_G)$ remain the same as in C_G.

We see that the structure of $\mathrm{rand}(C_G)$ is independent of labels. Thus, it is natural to say that $\mathrm{rand}(C_G)$ is neither homophilic nor heterophilic for any C_G. This motivates the following definition.

Definition 5 (Constant baseline)**.** A homophily measure h has constant baseline if there exists $R_{base} \in \mathbb{R}$ such that for any C_G we have $h(\mathrm{rand}(C_G)) = R_{base}$.

Homo-monotonicity. This property requires that adding homophilic edges increases the value of the measure. Let us reformulate this requirement in terms of the normalized class adjacency matrix C. If we add a homophilic edge within a class i, then in Definition 3, the denominator $|E|$ gets replaced by $|E| + 1$, the numerator of c_{ii} increases by 1, and the numerators of all other elements stay the same. Thus, when we add a homophilic edge within class i, all elements of C are multiplied by $\frac{|E|}{|E|+1} = 1 - \frac{1}{|E|+1}$, and after that c_{ii} is increased by $\frac{1}{|E|+1}$. Due to scale invariance, we can make $|E|$ arbitrarily large and thus make $\frac{1}{|E|+1}$ arbitrarily close to zero. Using the fact that we can add several homophilic edges one by one and the continuity property, we can essentially replace $\frac{1}{|E|+1}$ by any number from $(0, 1)$, from which the following definition follows.

Definition 6 (Homo-monotonicity)**.** Suppose C is not fully homophilic, i.e., $\sum\limits_{i=1}^{m} c_{ii} < 1$. Then, for any i and any $0 < \epsilon < 1$ we have $h((1-\epsilon)C + \epsilon E_{ii}) > h(C)$, where E_{ii} is a matrix unit (it has 1 at the position (i, i) and zeros everywhere else).

In other words, if not all edges of G are homophilic and we add one or several homophilic edges to G, then the resulting graph G' satisfies $h(G') > h(G)$.

Hetero-monotonicity. This property requires that removing heterophilic edges increases the value of the measure. Similarly to the above, when we delete a heterophilic edge between classes i and j, all elements of C are multiplied by $\frac{|E|}{|E|-1} = 1 + \frac{1}{|E|-1}$, and after that c_{ij} and c_{ji} are decreased by $\frac{1}{2(|E|-1)}$.

Definition 7 (Hetero-monotonicity)**.** Suppose C is not fully heterophilic, i.e., $\sum\limits_{i=1}^{m} c_{ii} > 0$. Then, for any $i \neq j$ and any ϵ such that $0 < \epsilon \leq 2(1+\epsilon)c_{ij}$ we have

$$h((1+\epsilon)C - \frac{\epsilon}{2}E_{ij} - \frac{\epsilon}{2}E_{ji}) > h(C),$$

where E_{ij} is a matrix unit (it has 1 at the position i, j and zeros everywhere else).

In other words, if not all edges of G are heterophilic and we remove one or several heterophilic edges from G, then the resulting graph G' satisfies $h(G') > h(G)$.

Empty class tolerance. If we pad a matrix C with a row and a column of zeros, the homophily of the matrix does not change. In other words, if we add to G a new dummy label that is not present in the graph, then the resulting graph G' satisfies $h(G') = h(G)$.

Class symmetry. Given any $m \times m$ matrix and any permutation $\sigma \in \mathbb{S}_m$, we can simultaneously permute the rows and columns by σ. A homophily measure h is called class-symmetric iff for any C and any σ we have $h(\sigma(C)) = h(C)$. In other words, the measure h is invariant to permuting (renaming) the classes. Note that all existing measures have class-symmetry.

D Homophily Measure Without Continuity Property

In this section, we give an example of a homophily measure that has all desirable properties excluding continuity. We show that it exhibits unwanted behavior and this justifies that continuity is indeed a necessary property.

Consider the following homophily measure:

$$h(C) = \sum_i \sqrt{c_{ii}} - 1 \text{ if } \sum_i \sqrt{c_{ii}} \leq 1,$$

$$h(C) = \sum_i c_{ii} \text{ if } \sum_i \sqrt{c_{ii}} > 1.$$

Here is an informal idea of the example. The term $\sum_i \sqrt{c_{ii}}$ has minimal agreement and constant baseline, but it does not have maximal agreement. Also, it has hetero-monotonicity everywhere, but its homo-monotonicity is sometimes violated when $\sum_i \sqrt{c_{ii}} > 1$. So, we cut it at $\sum_i \sqrt{c_{ii}} = 1$ to keep all desirable properties and avoid problems with maximal agreement and homo-monotonicity. After that, for $\sum_i \sqrt{c_{ii}} > 1$ we use $\sum_i c_{ii}$ (i.e., edge homophily), which has all desirable properties but does not have constant baseline. The problem with constant baseline is avoided since all $\mathrm{rand}(C)$ graphs (which constant baseline cares about) satisfy $\sum_i \sqrt{c_{ii}} = 1$.

Thus, the measure $h(C)$ is artificially constructed from two measures one of which is known to have no constant baseline. As a result, when a graph is homophilic ($h(C) > 0$), $h(C)$ is equal to edge homophily which is known to be not a good measure. This is achieved since we have a discontinuity at $h(C) = 0$.

Indeed, let us consider the following normalized class adjacency matrices:

$$L_1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}, \quad L_2 = \begin{bmatrix} \frac{1}{4} + \epsilon & \frac{1}{4} - \epsilon \\ \frac{1}{4} - \epsilon & \frac{1}{4} + \epsilon \end{bmatrix},$$

where $\epsilon > 0$ is a small positive number.

Since for L_1 we have $\sum_i \sqrt{c_{ii}} = \frac{1}{2} + \frac{1}{2} = 1$, we have $h(L_1) = \sum_i \sqrt{c_{ii}} - 1 = 0$. Since for L_2 we have $\sum_i \sqrt{c_{ii}} = \sqrt{\frac{1}{4} + \epsilon} + \sqrt{\frac{1}{4} + \epsilon} > 1$, we have $h(L_2) = \sum_i c_{ii} = \frac{1}{2} + 2\epsilon$. Thus, $h(L_1)$ and $h(L_2)$ differ by at least $\frac{1}{2}$, which is an unwanted behavior, since L_1 and L_2 themself are very close to each other. In particular, for $\epsilon \to 0$, the matrix L_2 tends to $\mathrm{rand}(L_2)$, but $h(L_2)$ does not tend to $R_{base} = 0$, which indicates that the measure is biased.

Now let us briefly discuss why h indeed has all desirable properties except continuity.

Continuity. Not satisfied at the points where $\sum_i \sqrt{c_{ii}} = 1$.

Minimal and maximal agreement. The measure h has minimal and maximal agreement with $R_{min} = -1$ and $R_{max} = 1$.

Homo- and hetero-monotonicity. The term $\sum_i \sqrt{c_{ii}}$ has homo- and hetero-monotonicity when $\sum_i \sqrt{c_{ii}} \leq 1$, as we prove in Appendix E (since this term is proportional to the second term of unbiased homophily). The term $\sum_i c_{ii}$ has homo- and hetero-monotonicity.

Constant baseline. Suppose $C = \mathrm{rand}(C)$. Then, $\sum_i \sqrt{c_{ii}} = \sum_i \sqrt{a_i^2} = \sum_i a_i = 1$. Therefore, $h(C) = \sum_i \sqrt{c_{ii}} = 1$. Thus, constant baseline is satisfied with $R_{base} = 0$.

Empty class tolerance. Satisfied, since both $\sum_i \sqrt{c_{ii}}$ and $1 + \sum_i c_{ii}$ do not change after padding a matrix C with a row and a column of zeros.

Class-symmetry. Satisfied, since both $\sum_i \sqrt{c_{ii}}$ and $1 + \sum_i c_{ii}$ are invariant to permuting (renaming) the classes.

E Proof of Theorem 5.1

In this section, we prove that the measure

$$h^{\alpha}_{unb}(C) := \frac{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} - c_{ij})}{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} + c_{ij})} + \alpha \min\left(\sum_i \sqrt{c_{ii}}, 1\right), \ \alpha > 0$$

has all the desirable properties listed in Appendix C, and its values $R_{max}, R_{base}, R_{min}$ are equal to $1+\alpha, \alpha, -1$, respectively.

Maximal agreement. Suppose $\sum\limits_{i=1}^{m} c_{ii} = 1$ (and therefore all $c_{ij} = 0$). Then the first term is equal to 1. Since $0 \le c_{ii} \le 1$, we have $\sqrt{c_{ii}} \ge c_{ii}$, thus $\sum\limits_i \sqrt{c_{ii}} \ge \sum\limits_i c_{ii} = 1$ and $\min(\sum\limits_i \sqrt{c_{ii}}, 1) = 1$. Therefore, the second term is equal to α, and homophily is equal to $1+\alpha$.

If $\sum\limits_{i=1}^{m} c_{ii} < 1$ (and therefore at least one $c_{ij} > 0$), then the first term is less than 1 and the second term is less than or equal to α. Thus, maximal agreement is satisfied with $R_{max} = 1+\alpha$.

Minimal agreement. If $\sum\limits_{i=1}^{m} c_{ii} = 0$ (and therefore at least one $c_{ij} > 0$), then the first term is equal to -1 and the second term is equal to 0. If $\sum\limits_{i=1}^{m} c_{ii} > 0$, then the first term is greater than or equal to -1 and the second term is greater than 0. Thus, minimal agreement is satisfied with $R_{min} = -1$.

Constant baseline. Suppose $C = \mathrm{rand}(C)$. Then, the first term is equal to 0 since the numerator of the first term is equal to

$$\sum_{i<j}(\sqrt{c_{ii}c_{jj}} - c_{ij}) = \sum_{i<j}(\sqrt{a_i^2 a_j^2} - a_i a_j) = \sum_{i<j}(0) = 0.$$

The second term is equal to α since

$$\sum_i \sqrt{c_{ii}} = \sum_i \sqrt{a_i^2} = \sum_i a_i = 1.$$

Thus, constant baseline is satisfied with $R_{base} = 0 + \alpha = \alpha$.

Homo-monotonicity. W.l.o.g. we can assume that we increase c_{11}. That is, all elements of C are multiplied by $(1-\epsilon)$ and after that the element $(1-\epsilon)c_{11}$ is increased by ϵ. We need to prove that after that the homophily measure has increased, given that C has at least one non-zero non-diagonal element.

Let us see what happens to the first term. Multiplying all elements by $(1-\epsilon)$ does not change the first term, since both numerator and denominator are multiplied by $(1-\epsilon)$. The subsequent addition of ϵ to $(1-\epsilon)c_{11}$ increases both numerator and denominator by δ, where $\delta = 0$ if $c_{22} = \dots = c_{nn} = 0$, and $\delta > 0$ if at least one of $c_{22}, \dots, c_{nn}$ is not zero.

Let us prove that if $\delta > 0$, then the first term has increased. Denote the numerator by N, the denominator by D:

$$\frac{N+\delta}{D+\delta} > \frac{N}{D} \ \Leftrightarrow \ (N+\delta)D > N(D+\delta) \ \Leftrightarrow \ D\delta > N\delta \ \Leftrightarrow \ D > N,$$

where we used the facts that D, $D+\delta$, and δ are greater than 0. The last inequality $D > N$ holds for all C with at least one non-zero non-diagonal element. Thus, we proved that if $\delta > 0$, then the first term has increased. Clearly, if $\delta = 0$, the first term does not change.

Let us see what happens to the second term. Instead of adding ϵ, we calculate the derivative w.r.t. ϵ:

$$\frac{\partial}{\partial \epsilon}\left(\sqrt{(1-\epsilon)c_{11}+\epsilon}+\sqrt{(1-\epsilon)c_{22}}+\ldots+\sqrt{(1-\epsilon)c_{nn}}\right)=$$
$$=\frac{1}{2}\left(\frac{1-c_{11}}{\sqrt{c_{11}}}-\frac{c_{22}}{\sqrt{c_{22}}}-\ldots-\frac{c_{22}}{\sqrt{c_{nn}}}\right)=\frac{1}{2}\left(\frac{1}{\sqrt{c_{11}}}-\sum_i\sqrt{c_{ii}}\right).$$

As long as $1 \geq \sum_i \sqrt{c_{ii}}$, this derivative is positive, since $\frac{1}{\sqrt{c_{11}}} > 1 \geq \sum_i \sqrt{c_{ii}}$. Thus, the second term has a positive derivative with respect to ϵ when $\sum_i \sqrt{c_{ii}} < 1$ and zero derivative when $\sum_i \sqrt{c_{ii}} > 1$.

Now let us combine our results for the first and second terms to prove that homo-monotonicity holds. If $\delta > 0$, then the first term is increasing and the second term is non-decreasing, thus homo-monotonicity holds. If $\delta = 0$, then the only possibly nonzero diagonal element is c_{11} and $c_{11} < 1$, thus $\sum_i \sqrt{c_{ii}} = \sqrt{c_{11}} < 1$. Therefore, the first term stays constant and the second term is increasing, thus homo-monotonicity holds. This concludes the proof that h satisfies the homo-monotonicity property.

Hetero-monotonicity. W.l.o.g. we can assume that c_{12} decreases. That is, all the elements of C are multiplied by $(1+\epsilon)$ and after that the element $(1+\epsilon)c_{12}$ is decreased by ϵ. We need to prove that after that the homophily measure has increased, given that C has at least one non-zero diagonal element.

Let us see what happens to the first term. Multiplying all the elements by $(1+\epsilon)$ does not change the first term, since both numerator and denominator are multiplied by $(1+\epsilon)$. The subsequent subtraction of ϵ from $(1+\epsilon)c_{12}$ increases the numerator and decreases the denominator by positive $\delta = \epsilon > 0$. Denote the numerator by N, the denominator by D:

$$\frac{N+\delta}{D-\delta} > \frac{N}{D} \quad \Leftrightarrow \quad (N+\delta)D > N(D-\delta) \quad \Leftrightarrow \quad D\delta > -N\delta \quad \Leftrightarrow \quad D > -N,$$

where we use the facts that $D, D-\delta$, and δ are greater than 0. The last inequality $D > -N$ holds for all C with at least two non-zero diagonal elements. Thus, we proved that if C has at least two non-zero diagonal elements, then the first term has increased. Clearly, if C has exactly one non-zero diagonal element, then the first term stays equal to -1 and thus does not change.

Let us see what happens to the second term. The expression $\sum_i \sqrt{c_{ii}}$ will be multiplied by $\sqrt{1+\epsilon}$. Therefore, the second term will increase if $\sum_i \sqrt{c_{ii}} < 1$ or stay constant if $\sum_i \sqrt{c_{ii}} \geq 1$. Clearly, if C has exactly one non-zero diagonal element, then $\sum_i \sqrt{c_{ii}} < 1$, thus the second term increases.

Now let us combine our results for the first and second terms to prove that hetero-monotonicity holds. If C has exactly one non-zero diagonal element, then the first term stays constant and the second term increases. If C has at least two non-zero diagonal elements, then the first term increases and the second term non-decreases. This concludes the proof that h satisfies the hetero-monotonicity property.

Empty class tolerance. Satisfied, since both first and second terms do not change after padding a matrix C with a row and a column of zeros.

Class-symmetry. Satisfied, since both first and second terms are invariant to permuting (renaming) the classes.

Continuity. Satisfied, since both first and second terms are continuous.

F Ajusted Homophily Non-Monotonicity Example

Adjusted homophily does not have minimal agreement and it does not have monotonicity when the value of adjusted homophily is low. Let us give an example of unwanted behavior corresponding to the lack of monotonicity.

Consider the following class adjacency matrix. We delete heterophilc edge between classes 1 and 2 and see how adjusted homophily changes.

$$\begin{bmatrix} 0 & \cancel{1} & 0 & 0 \\ \cancel{1} & 0 & 0 & 0 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 8 & 2 \end{bmatrix}.$$

Before deleting the edge the adjusted homophily was equal to $\frac{\frac{2}{11} - \frac{1+1+100+100}{22^2}}{1 - \frac{1+1+100+100}{22^2}} \approx -0.404$. After deleting the edge the adjusted homophily is equal to $\frac{\frac{2}{10} - \frac{100+100}{20^2}}{1 - \frac{100+100}{20^2}} = -0.6$. Thus the adjusted homophily has decreased which contradicts hetero-monotonicity. In contract, the unbiased homophily h_{unb} has increased from $\frac{2-8-1}{2+8+1} \approx -0.636$ to $\frac{2-8}{2+8} = -0.6$. For h^{α}_{unb} we have an increase from $-0.636 + 0.603\alpha$ to $-0.6 + 0.632\alpha$.

G Equivalence of Two Formulas for h_{unb}

Here is a calculation which proves that formulas (3) and (4) for h_{unb} are equivalent:

$$h_{unb}(C) = \frac{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} - c_{ij})}{\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} + c_{ij})} = \frac{2\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} - c_{ij})}{2\sum\limits_{i<j}(\sqrt{c_{ii}c_{jj}} + c_{ij})} = \frac{\sum\limits_{i\neq j}(\sqrt{c_{ii}c_{jj}} - c_{ij})}{\sum\limits_{i\neq j}(\sqrt{c_{ii}c_{jj}} + c_{ij})} =$$

$$= \frac{\left(\sum\limits_i \sqrt{c_{ii}}\right)^2 - \sum\limits_i c_{ii} - \sum\limits_{i\neq j} c_{ij}}{\left(\sum\limits_i \sqrt{c_{ii}}\right)^2 - \sum\limits_i c_{ii} + \sum\limits_{i\neq j} c_{ij}} = \frac{\left(\sum\limits_i \sqrt{c_{ii}}\right)^2 - 1}{\left(\sum\limits_i \sqrt{c_{ii}}\right)^2 + 1 - 2\sum\limits_i c_{ii}}. \quad (6)$$

H Comparing Unbiased and Adjusted Homophily

In this section, we additionally illustrate the difference between unbiased and adjusted homophily with a simple synthetic example.

Here, for a fixed number of classes m and a chosen value of h_{unb}, we construct a normalized class adjacency matrix (assuming that homophilic/heterophilic edges are evenly distributed between classes/class pairs), and calculate h_{adj} based on it. Thus, we can see how adjusted homophily changes if we fix unbiased homophily but change the number of classes. We make the following observations.

Table 10: Adjusted homophily for different values of h_{unb} and m

m, h_{unb}	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
2	-1.000	-0.800	-0.600	-0.400	-0.200	0.000	0.200	0.400	0.600	0.800	1.000
3	-0.500	-0.421	-0.333	-0.235	-0.125	0.000	0.143	0.308	0.500	0.727	1.000
4	-0.333	-0.286	-0.231	-0.167	-0.091	0.000	0.111	0.250	0.429	0.667	1.000
5	-0.250	-0.216	-0.176	-0.129	-0.071	0.000	0.091	0.210	0.375	0.615	1.000
6	-0.200	-0.174	-0.143	-0.105	-0.059	0.000	0.077	0.182	0.333	0.571	1.000
7	-0.167	-0.145	-0.120	-0.089	-0.050	0.000	0.067	0.160	0.300	0.533	1.000
8	-0.143	-0.125	-0.103	-0.077	-0.043	0.000	0.059	0.143	0.273	0.500	1.000
9	-0.125	-0.110	-0.091	-0.068	-0.038	0.000	0.053	0.129	0.250	0.471	1.000
10	-0.111	-0.098	-0.081	-0.061	-0.034	0.000	0.048	0.118	0.231	0.444	1.000

First, when $h_{unb} = 1$, we always have $h_{adj} = 1$. This is expected since this case corresponds to

graphs having only homophilic edges and both measures have maximal agreement. Second, when $h_{unb} = 0$, we always have $h_{adj} = 0$. This is also expected since both measures have constant baseline.

Then, when $h_{unb} = -1$, one can see that h_{adj} increases with the number of classes. Since $h_{unb} = -1$ corresponds to graphs containing only heterophilic edges, this shows the lack of minimal agreement for h_{adj}.

We also see that when $m = 2$, the measures are equal to each other (in our synthetic example). Then, for heterophilic graphs ($h_{unb} < 0$) and fixed h_{unb}, h_{adj} increases with the number of classes, while for homophilic graphs ($h_{unb} > 0$) and fixed h_{unb}, h_{adj} decreases with the number of classes.

I Homophily Measures for Directed Graphs

In this section, we reformulate the desirable properties of a homophily measure for directed graphs. After that we prove that these properties contradict each other. Thus, no measure satisfying all of them can be constructed. In Subsection I.3, we suggest steps towards modification of these properties to avoid contradictions.

I.1 Desirable Properties of Homophily Measures

In this section, a homophily measure is a function h from the set of all directed unweighted graphs (that may include self-loops and multiple edges) with labeled nodes to $\mathbb{R}$.

Definition 8. For a directed graph G, we define $m \times m$ *class adjacency* matrix L_G by $l_{ij} = |\{(u, v) \in E : (y_u, y_v) = (i, j)\}|$.

The *normalized class adjacency matrix* C_G is then defined as $\frac{1}{|E|} L_G$. This matrix can be non-symmetric and it has non-negative elements summing to one. As before, we consider only C_G with at least two non-zero elements.

Below we consider only *edge-wise* and *scale-invariant* homophily measures that are defined exactly as in the undirected case. We define the properties *maximal agreement*, *minimal agreement*, *homo-monotonicity*, *empty class tolerance*, *class-symmetry*, and *continuity* as in the undirected case (see Appendix C). Let us now define *hetero-monotonicity* and *constant baseline*.

Hetero-monotonicity As in the undirected case, if not all edges of G are heterophilic and we remove one or several heterophilic edges from G, then the resulting graph G' should satisfy $h(G') > h(G)$. Formally, let C be not fully heterophilic, i.e., $\sum_{i=1}^{m} c_{ii} > 0$. Then, for any $i \neq j$ and any ϵ such that $0 < \epsilon \leq (1 + \epsilon)c_{ij}$, we have $h((1 + \epsilon)C - \epsilon E_{ij}) > h(C)$.

Constant baseline As in the undirected case, for every C_G, we define the normalized class adjacency matrix $\mathrm{rand}(C_G)$, corresponding to a graph whose structure is independent of labels, while class degrees distribution is the same as for C_G. After that, we require $\forall C_G : h(\mathrm{rand}(C_G)) = R_{base}$.

Definition 9. For a normalized class adjacency matrix C_G, let $a_i := \sum_{k=1}^{m} c_{ik}$ and $b_j := \sum_{k=1}^{m} c_{kj}$. Then, we define $\mathrm{rand}(C_G)$ as: $\mathrm{rand}(C)_{ij} := a_i b_j$.

To motivate this definition, suppose we fix the in-degrees, out-degrees, and labels of all nodes in G and redraw edges between the nodes uniformly at random. That is, every out-half-edge has an equal probability to form an edge with any in-half-edge. Similarly to the undirected case, we observe that the expected fraction of homophilic edges in class i is then $a_i b_i$ and the expected fraction of heterophilic edges from class i to class j is $a_i b_j$. Thus, the elements of $\mathrm{rand}(C_G)$ correspond to expected fractions of edges of each type under the procedure when we randomly redraw all edges in a graph G. Note that column-wise and row-wise sums of $\mathrm{rand}(C_G)$ remain the same as in C_G.

Definition 10 (Constant baseline)**.** A homophily measure h has constant baseline if there exists $R_{base} \in \mathbb{R}$ such that for any C_G we have $h(\mathrm{rand}(C_G)) = R_{base}$.

I.2 Contradiction

It turns out that this set of properties is internally contradictory. The following two propositions hold.

Proposition I.1. *The constant baseline and minimal agreement properties contradict each other.*

Proof. Suppose we have a graph K where some classes have only out-edges and all the other classes have only in-edges. Suppose we also have $K = \mathrm{rand}(K)$, thus constant baseline requires the homophily of this graph to be equal to R_{base}. On the other hand, minimal agreement requires the homophily of this graph to be equal to R_{min} since all edges of this graph are heterophilic. This easily leads to contradiction, as we see in the example below: here constant baseline requires $h(K) = h(L) = R_{base}$ and minimal agreement requires $h(K) = R_{min} < h(L)$, which gives a contradiction.

$$K = \begin{bmatrix} 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Proposition I.2. *Constant baseline and hetero-monotonicity properties contradict each other.*

Proof. Suppose we have graphs $T = \mathrm{rand}(T)$ and $L = \mathrm{rand}(L)$ such that L can be obtained from T by deleting several heterophilc edges. Then, constant baseline requires $h(T) = h(L) = R_{base}$ and hetero-monotonicity requires $h(T) < h(L)$, which gives a contradiction. In example below using the hetero-monotonicity property, we can replace all non-zero elements in the fourth column and third row of T with zeros, obtaining matrix L.

$$T = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & 0 & \cancel{\frac{1}{9}} \\ \frac{1}{9} & \frac{1}{9} & 0 & \cancel{\frac{1}{9}} \\ \cancel{\frac{1}{9}} & \cancel{\frac{1}{9}} & 0 & \cancel{\frac{1}{9}} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Not necessarily edge-wise scalar-invariant measures. We gave the examples above to prove the contradiction between the desirable properties, while the properties were formulated for edge-wise scalar-invariant measures. Let us give a sketch of the proof that the same examples work for any measure which is not necessarily scalar-invariant or even edge-wise.

Indeed, consider our example which shows the contradiction between constant baseline and minimal agreement. Consider two graphs satisfying constant baseline, with corresponding normalized adjacency matrices K and L. Since all edges of the first graph are heterophilic and not all edges of the second graph are heterophilic, the minimal agreement property requires that homophily of the first graph is R_{min} and homophily of the second is greater than R_{min}. Yet, constant baseline requires the homophily of both of them to be equal to R_{base}, thus giving a contradiction.

Consider our example which shows contradiction between constant baseline and hetero-monotonicity. Consider any graph satisfying constant baseline, corresponding to normalized adjacency matrix T. Delete in this graph all edges with at least one end in classes 3 and 4, getting the second graph with normalized adjacency matrix L. Constant baseline require the homophily of both of these graphs to be equal to R_{base}. Yet we have deleted only hereophilic edges, thus hetero-monotonicity requires the homophily of the second graph to be greater than homophily of the first graph, thus giving a contradiction.

I.3 Discussion

As shown above, the list of desirable properties for the directed case is self-contradicting. Hence, it is natural to try to modify these properties to avoid contradictions.

The contradiction between constant baseline and minimal agreement properties is quite severe. This contradiction occurs since there can be cases when C is fully hetero and $C = \mathrm{rand}(C)$. A possible way to resolve this problem is to simply allow h to be undefined (and discontinuous) for all such C.

To resolve the contradiction between constant baseline and hetero-monotonicity, we can rethink the property of monotonicity. Note that our current definition of constant baseline is naturally formulated in terms of normalized class adjacency matrices, while homo- and heter-monotonicity are formulated in terms of edges, and only after that are translated to matrices. Let us formulate a new property that

we call *randomization monotonicity*. First, this property is formulated in terms of normalized class adjacency matrices. Second, it essentially extends the constant baseline from one point to the whole range of the measure: we require some desirable (unbiased) behavior not only for random graphs but also for homophilic/heterophilic structures.

Suppose that for a given graph G, we randomly sample some fraction of its edges and redraw them randomly (while keeping the degree distribution, as before). In terms of the normalized class adjacency matrix, we replace C by $(1-\epsilon)\cdot C + \epsilon \cdot \mathrm{rand}(C)$ for some $\epsilon > 0$. Then, it is natural to require that homophily of the graph gets closer to R_{base}, since a part of its edges become independent of labels. Formally, we require the following.

Randomization monotonicity For any $\epsilon \in (0, 1]$ we have:

$$h(C) < h((1-\epsilon)\cdot C + \epsilon \cdot \mathrm{rand}(C)) \leq R_{base} \text{ if } h(C) < R_{base}\,,$$
$$h(C) > h((1-\epsilon)\cdot C + \epsilon \cdot \mathrm{rand}(C)) \geq R_{base} \text{ if } h(C) > R_{base}\,.$$

We believe that the development and analysis of non-contradicting properties for the directed case is an important subject for future studies.

J Graph Datasets Description

Cora, citeseer, and pubmed [17–21] are three paper citation graph datasets. In cora and citeseer, the labels are paper topics, and in pubmed labels correspond to the type of diabetes addressed in the paper. Coauthor-cs and coauthor-physics [22] are co-authorship graph datasets. Nodes are authors, edge correspond co-authoring a paper, and labels are fields of study. Amazon-computers and amazon-photo [22] are co-purchasing graph datasets. Nodes are products, edge means that two products are frequently bought together, and labels are product categories. Lastfm-asia is a social graph of music streaming site LastFM users living in Asia [23]. Edges correspond to follower relationships, labels are user's nationality. Facebook [24] is a graph dataset, where nodes are official Facebook pages, links correspond to mutual likes, and labels are site categories. Github [24] is a graph dataset, where nodes are GitHub users, edges correspond to follower relationships, and binary label tells that a person is either a web or a machine learning developer. Actor [8, 25] is a graph dataset, where nodes are actors, edges correspond to co-occurrence on the same Wikipedia page, and the labels are based on words from an actor's Wikipedia page. Flickr [26] is a graph dataset of images, where labels are image types. Deezer-europe [23] is a graph of users of the music streaming service Deezer, where labels correpond to user's gender. Twitch-de and twitch-pt [24] are graph datasets of gamers from the streaming service Twitch, where labels correspond to the use of explicit language by the gamers. Genius [27] is large-scale heterophilous graph dataset. Roman-empire, amazon-ratings, minesweeper, tolokers and questions [5] are mid-scale heterophilous datasets.

We transform all graph datasets to undirected graphs and remove self-loops and multiple edges.