Local Pluralistic Homophily in Networks: A New Measure Based on Overlapping Communities

Fernando Barraza[1,4][0000−0003−0161−2220], Carlos Ramirez[1][0000−0002−4254−5610], and Alejandro Fernández[2,3][0000−0002−7968−6871]

[1] Universidad Javeriana de Cali, Colombia
[2] LIFIA, FI, Universidad Nacional de La Plata, Argentina
[3] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina
[4] Universidad de San Buenaventura Cali, Colombia

Abstract. Pluralistic homophily is an important phenomenon in social network analysis as nodes tend to associate with others that share their same communities. In this work, we present the concept of local pluralistic homophily of a node in a network, along with a method to measure it. It is based on the assortativity index proposed by other authors. We analyze the distribution of local pluralistic homophily in different networks using publicly available datasets. We identify patterns of behavior of the proposed measure that relate to various structural and semantic characteristics of a network. These findings are significant because they help better understand how pluralistic homophily affects communities. Furthermore, our results suggest possible applications of local pluralistic homophily in future research.

Keywords: Local Pluralistic Homophily · Networks · Communities.

1 Introduction

In the field of social network analysis, homophily is known as *assortativity* or *assortative mixing*. Assortative mixing is the tendency of a node to link with nodes similar to it. The similarity between nodes can be defined based on a common characteristic, such as the node degree [1], but other features can also be used. Social networks have shown a high assortativity, contrary to biological and technological networks where a low assortativity is observed (i.e., *disassortativity*) [2]. The study of homophily and communities is essential to understand the structure and dynamics of social networks.

In network science, a community is a group of nodes that have a. higher likelihood of connecting to each other than to nodes from other[3]. To detect these groups, several techniques are used, such as differential equations, random walks, spectral clustering, and modularity maximization. In some cases, the network can be partitioned into disjoint communities, while in others, the structure can be viewed as communities where some nodes belong to two or more of them. In these cases, it's said that the network has overlapping communities. In a study conducted by Yang and Leskovec [4], it was found that overlapping

community areas in networks exhibit a higher density of connections compared to non-overlapping ones. Furthermore, the authors noted a noteworthy pattern where nodes belonging to multiple communities tend to connect with similar nodes based on their community memberships. This phenomenon, termed as *pluralistic homophily*, highlights the tendency of nodes to exhibit preference towards nodes with similar attributes across multiple communities.

In this article, we introduce the concept of local pluralistic homophily as the tendency of a specific node to link with others with the same overlap. Our measures are based on the already-established concept of assortativity and provide a more suitable way to measure the phenomenon of pluralistic homophily. We define an extended measure of homophily based on the similarity of nodes that share community memberships. Our measures reveal how much nodes from a network with overlapping communities prefer to attach to others with similar membership quantities. We validate the measure using six datasets from social networks and collaboration networks. Our results suggest the likely application of local pluralistic homophily in node attribute prediction and other tasks common in the network analytics field.

2   Theoretical Framework

In social networks, individuals often exhibit a strong tendency to associate with others who share similar attributes, such as political beliefs and social status. This phenomenon is referred to as *assortativity*. However, in certain networks, such as dating networks, an opposite pattern can be observed, where the majority of links exist between nodes with dissimilar attributes. This is known as disassortativity. An assortativity metric, as defined in [2], captures the inclination of nodes to form direct links with others who possess similar properties. The properties of nodes can be categorized as enumerative or scalar. Enumerative properties have a finite set of possible values, such as gender, nationality, and profession in the context of a social network composed of individuals. On the other hand, a common example of assortative mixing based on scalar properties is mixing by degree, where nodes with high degrees tend to connect with other highly connected nodes, while nodes with low degrees are primarily linked to nodes with few connections.

Def. Let a graph be an ordered pair $G = (V_G, E_G)$ where V_G is the vertex set whose elements are the vertices denoted by v (or nodes of the graph), and E_G is the edge set whose elements are the edges, denoted by e (or links between vertices of the graph), such that $V(G) \equiv V_G = \{v_1, v_2, \ldots, v_n\}$, and $E(G) \equiv E_G = \{e = (v_x, v_y) \mid v_x, v_y \in V\}$

2.1   Pluralistic homophily

The concept of pluralistic homophily is introduced in [4]. Previous studies on community detection have typically assumed that nodes within the same community share similar properties. However, the authors of this paper argue that

nodes can also be considered similar based on the number of shared memberships they have across multiple communities. This phenomenon is known as pluralistic homophily.

To illustrate this concept, consider a community of students who all play football. In this case, the similarity between two individuals could be based on their shared interest in football. However, if those same individuals also play basketball, we might hypothesize that they share an even stronger similarity than if they only had one shared activity. This is an example of pluralistic homophily, where the similarity between two individuals is based on their shared memberships across multiple communities.

While homophily has received significant attention in research, the concept of pluralistic homophily remains under-explored. To address this research gap, we propose a novel measure of assortativity that specifically considers overlapping community memberships. In our measure, we extend the commonly used correlation coefficient for quantifying homophily by incorporating the degree of overlap in community memberships. This overlap refers to the number of communities to which each node belongs in the network. By introducing this measure, we aim to gain a deeper understanding of the role of pluralistic homophily in community detection and other network analysis tasks.

Based on equation 3 in [1], we define a coefficient denoted by h that measure the pluralistic homophily of a network as,

$$h = \frac{1}{\sigma_q^2} \sum_{xy} m_x m_y (e_{m_x m_y} - q_{m_x} q_{m_y})$$ (1)

where m_x, m_y are the number of membership of the nodes x and y in a link, $e_{m_x m_y}$ is the joint probability for the memberships values x and y, and q is the quantity membership distribution of the network. Note that equation 1 uses the number of memberships of each node as the scalar value to calculate the assortativity as defined in [2]. Under this equation, the assortativity value must be interpreted according to the nearing of the result to zero. If the result is near -1 means that, in general, nodes trends to link with others having a different scalar number. On the contrary, if it nears 1, the nodes tend to link with others with a similar scalar number. A result near zero means that nodes tend to link randomly. To avoid subjectivity in the interpretation of the results on how much h is nearest or farthest to zero, we define a ε-value to establish clear limits in such appreciation. In general, a $|h| < \varepsilon$ means a network non-assortative by node membership, $h < -\varepsilon$ is a disassortativity one, while $h > \varepsilon$ exhibits assortativity mixing, in other words, is pluralistic *homophilically* speaking.

2.2   Local Pluralistic homophily

It is common to find cases where some nodes tend to link differently than most other nodes on the network. This is, despite a network that can exhibit an overall positive or negative assortativity, some individual nodes could have an opposite

tending. For these cases, [5] introduces the concept of local assortativity, making a comparative analysis between the assortativity of a node and the network's assortativity. This distinction has allowed some works [6], [7] to understand the composition of patterns found throughout the network structure, as well as to observe the different assortative mixing throughout the hierarchical levels of the networks. From this perspective, the local assortativity can then be seen as the contribution that each node makes to the network's assortativity or also a group of them (for example, a partition, a community, etc.).

Local assortativity can also be observed when thinking in terms of pluralistic homophily. This means that it is possible that a specific node with a high quantity of memberships links to others with also high memberships, even when the pluralistic homophily of the entire network is near -1. Based on this behavior, the local pluralistic homophily can be defined as the tendency of a specific node to link to other nodes with similar overlapping quantities.

We examine the works about local assortativity reviewed in [8], looking for a reasonable way to compute the local pluralistic homophily. From these works, we based on equation 4.10 on [9] to define the local pluralistic homophily h of a node v as,

$$h_v = \frac{\sum\limits_{i \in N(v)} (m_v - \mu_q)(m_i - \mu_q)}{2M\sigma_q^2} \qquad (2)$$

where i iterates over all the neighbors of node v, m_v is the number of communities to which node v belongs, while m_i is the number of communities to which node i belongs. M is the total number of nodes in the network, and μ_q and σ_q are the mean and standard deviation of the probability distribution of node memberships in the network. The resulting value of h_v represents numerically the trend of a specific node v to link to other nodes with near memberships number. The summation of h_v values for all nodes is congruent with the results obtained using equation 1 for the pluralistic homophily of the entire network.

Figure 1 shows an example of calculating the assortative mixing by the number of memberships of the network nodes, i.e. the pluralistic homophily h of the entire network. Taking ε as 0.1, the value of $h = -0.032$ show a non-assortative mixing due to $|h| < \varepsilon$. This is consistent with the result of h_v for most of the nodes (v_1, v_3, v_4, v_5, v_7). However, a node is disassortative $h_{v_6} < -\varepsilon$, and other assortative $h_{v_2} > \varepsilon$. This behavior can be explained, for instance, by node v_2 being more assortative than the others because it connects more frequently with nodes that have a similar number of memberships. Node v_6, on the other hand, is the most disassortative because it links to nodes with dissimilar numbers of memberships. In the following sections, we will show this behavior in real networks with a different number of nodes, links, and communities.
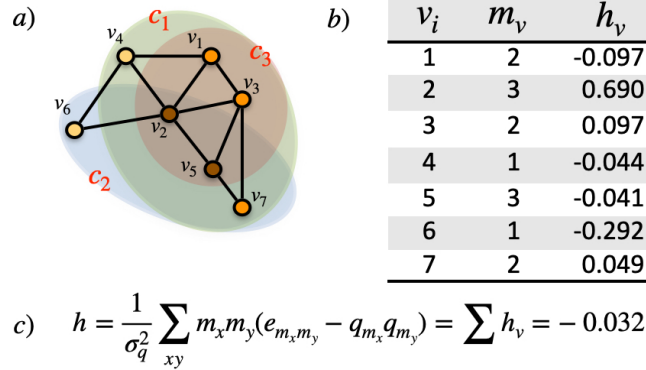
a)

b)

v_i	m_v	h_v
1	2	-0.097
2	3	0.690
3	2	0.097
4	1	-0.044
5	3	-0.041
6	1	-0.292
7	2	0.049

c) $$h = \frac{1}{\sigma_q^2} \sum_{xy} m_x m_y (e_{m_x m_y} - q_{m_x} q_{m_y}) = \sum h_v = -0.032$$

Fig. 1: Calculus of pluralistic homophily of a network and its nodes. a) A network with 7 nodes (filled with different colors to differentiate the number of memberships of each node) and 3 detected communities in shaded colors (c_1:green, c_2:sky blue, c_3:brown). b) A table that calculates the pluralistic homophily h_v for every node v_i in the network with its respective number of memberships m_v. c) Calculus of the pluralistic homophily for the entire network. We can see the result is equal to the sum of all local pluralistic homophily h_v from the table.

3   Methodology

We performed a series of experiments to analyze the behavior of pluralistic homophily in six datasets containing networks and communities related to them. To calculate pluralistic homophily, we used the equation 1 presented above. In addition, we calculated the local pluralistic homophily of each node in the networks using the equation 2. We used these resulting measures to generate complementary cumulative continuous distribution (CCDF) plots that allowed us to analyze what distribution follows these data for each dataset. We also compare the behavior of pluralistic homophily with what we have called simple homophily, which is based on the degree of the nodes. This comparison is relevant as it allows us to understand better the role of pluralistic homophily in network structure and how it differs from other measures of homophily. In particular, we expect to find that pluralistic homophily highlights clustering patterns in communities that are not evident in simple homophily. We also expect evidence of the behavior of local pluralistic homophily versus the node membership and node degree. This is useful to understand how pluralistic homophily behaves in some hubs (nodes with high degrees) and in peripherical nodes (nodes with low degrees).

3.1   Datasets

For our study, we selected six data sets of real-world networks and their respective communities: Stackoverflow (SO), DBLP, Amazon Product, Livejournal,

Youtube, and Orkut. SO is a collaborative network for questions and answers on programming and related topics. The SO network is reconstructed by users as nodes, and the edges are the questions that are asked or answered between them. Communities in this network were generated using the HLC community detection algorithm [10]. These communities reflect, through questions that have been tagged with technology topics, common and correlated interests that are addressed in the interaction between the members of said communities. An overlap between those communities occurs when a user belongs to several communities, given their interest in multiple technologies. The other networks are Livejournal, an online blogging community; Orkut, an online social network; Youtube, a video-sharing website; All of them are considered social networks in which users are represented as nodes, and the edges represent the explicit relationships established between them. These networks allow users to form groups that other members can then join. These groups are the communities, and overlapping happens when a user belongs to more than one group. The two last networks are Amazon and DBLP. In the former, every product is a node, and the edges are formed between them when they are bought together. Product categories define the communities. And latter is DBLP, where authors are the nodes, edges mean coauthoring between them, and each magazine and conference where they publish defines a community. All network datasets[5], except SO, with their respective ground-truth communities, are publicly available at [11]. The choice of these datasets and their source provides high reliability in the data to be used in our experiments as well as the possibility of comparing them with the works of other authors who have also been using them extensively.

Table 1 shows basic characteristics and measures for the datasets. In summary, we include three networks considered collaborative and the other three considered social networks, each with different characteristics in terms of size, structure, and methods used to define their communities. This will help us gain a better understanding of pluralistic homophily behavior across diverse contexts.

The table presents the characteristics of the networks. Generally, we observe that these datasets differ in size, i.e. in the number of all nodes, links, and communities they exhibit. Also, we present the values of pluralistic homophily h and homophily by degree r (also known as degree assortativity) on which we will go into more depth in the next section. According to the size of the networks, SO, DBLP, and Amazon could be considered medium-sized. The rest are large-size. Values of average degree $\langle d \rangle$ are consistent with the size of the networks, except for the Youtube network where there is a low $\langle d \rangle$ compared with the others of similar size. With respect to the properties of the communities, we can observe three kinds of community sizes: Small ones such as DBLP and Youtube, medium as SO and Amazon, and large ones such as Livejournal and Orkut. These community sizes are consistent with the size of the networks, except again for the Youtube network where the community size is more like that of small networks.

---

[5] Dates from datasets are shown on the SNAP web page except for SO which was downloaded from the site https://archive.org/details/stackexchange and communities generated until 2021.

Dataset	Network Properties			Community Properties		Assortativity Measures	
	N	E	$\langle d \rangle$	K	$\langle m \rangle$	h	r
StackOverflow	790.458	1.872.715	4.76	115.969	1.85	0.0332	−0.0381
DBLP	317.080	1.049.866	6.62	13.477	2.27	0.2166	0.2665
Amazon	334.863	925.872	5.53	75.149	6.78	0.4887	−0.0588
LiveJournal	3.997.962	34.681.189	17.35	664.414	1.79	0.2132	0.0451
Youtube	1.134.890	2.987.624	5.26	16.386	0.11	0.0647	−0.0369
Orkut	3.072.441	117.185.083	76.28	6.288.363	34.85	0.2335	0.0158

Table 1: Comparison of network and community properties, as well as assortativity measures, for six different datasets. N: number of nodes, E: number of edges, $\langle d \rangle$: average node degree, K: number of communities, $\langle m \rangle$: average number of community memberships per node, h: pluralistic homophily coefficient, r: degree assortativity coefficient.

The average number of community memberships per node $\langle m \rangle$, a value intrinsically related to the pluralistic homophily seems not to be directly related to network size. For example, while Amazon's network has the second largest $\langle m \rangle$, Livejournal network has one of the smallest, despite the latter having a network size ten times the former. We then have different values of $\langle m \rangle$ independent of the size of the network, which provide us with a solid base in the data sets to explore the behavior of pluralistic homophily in the conducted experiments.

4   Results

We present the results of the local pluralistic homophily analysis conducted on each of the six datasets described earlier. We calculated the entire pluralistic homophily of each network and the local pluralistic homophily for all nodes belonging to any related community. Notably, as we can see in table 1 pluralistic homophily h displays positive values across all datasets, contrasting with the mixing results of assortativity by the degree r that show some negative ones. DBLP, Amazon, Livejournal, and Orkut show pluralistic homophily $h > \varepsilon$, while the other networks, SO and Youtube do not, $|h| < \varepsilon$, having as ε the same value defined for the toy example above. Relating to assortativity by degree r, only the DBLP network shows assortative mixing, the others non-assortative. This is the only network that shows assortative mixing in both h and r as we have shown in the table 1. None shows disassortativity in both measures for the entire network. These very varied results in the correlation between h and v, suggest that overall at the level of the entire network, there may not be a uniform direct relationship between the tendency of nodes to connect with nodes of the same degree with the tendency of those nodes to connect with nodes that share the same communities. Regarding that both the quantities related to the size of the input data sets (N, E and K) and also to the average values ($\langle d \rangle$ and $\langle m \rangle$) influence the value of h and r respectively, we can see that there is no obvious correlation either.

Results at the entire network level contrast with the results at the local level. To visualize the results of how is the tendency of nodes to link with others with the same m we calculate the distribution of local pluralistic homophily. Before interpretation of when a node exhibits pluralistic homophily or not, we set $\varepsilon = 0.1 \cdot 10^{-5}$, a very small value. We must take into account as noted above that $h = \sum h_v$, so given the significant difference in the number of nodes v between the toy example (shown above) and the real networks, it is necessary to apply a scaling factoring to ε of -5 which is related to the size difference between the networks. The rationale behind applying a scaling factor to ε of -5 is related to the difference in network size. As the magnitude of ε values varies with the network size, we need to adjust ε to account for this difference and ensure meaningful comparisons between networks. When scaling up or down the network size by an order of magnitude of 10 (e.g., from a small network to a large network), ε values need to scale proportionally to maintain consistency in measuring pluralistic homophily. Applying a scaling factor of -5 means dividing the original value of ε by 10^5, ensuring that the adjusted value of ε is suitable for the scale of the network under consideration. This scaling procedure allows us to compare the local pluralistic homophily values across different-sized networks effectively, enabling a fair assessment of the tendencies of nodes to exhibit pluralistic homophily. By considering the scaling factor, we account for the size discrepancy and ensure that the interpretation of pluralistic homophily remains consistent across diverse network sizes

We show then, the complementary cumulative probability distribution function (CCDF) of the local pluralistic homophily for each network (see figure 2). The "S" shaped curves present in all CCDF figures indicate a distribution with a high probability of low values, followed by a decreasing probability as values increase, and then a long tail to the right that indicates the presence of extreme values. The initial part of the curve indicates that there are many nodes in the network with low local pluralistic homophily, while the final part of the curve indicates that there are a small number of nodes with high local pluralistic homophily. The abrupt decrease in probability close to zero indicates an inflection point in the distribution, which coincides with the value determined for ε and can be interpreted as a threshold from which local pluralistic homophily begins to increase significantly. The presence of extreme values may be indicative of nodes or groups of nodes in the network (hubs) that have extremely high or low local pluralistic homophily, which may be important for understanding the overlap of communities and their relationship with the structure of the network. Note for instance that, in the case of DBLP and Amazon such a decrease at a certain threshold is not so pronounced, maybe there are additional factors at play that influence the distribution as the node degree. In order to approach a better understanding of these patterns, we present the next figure.

We created scatter plots showing the relationship between a node's degree d and its local pluralistic homophily h_v, as well as between d and the number of communities it belongs to, denoted as m_v. Each dot in the scatter plot represents a node v in the network (see figure 3). We colored the dots according to whether
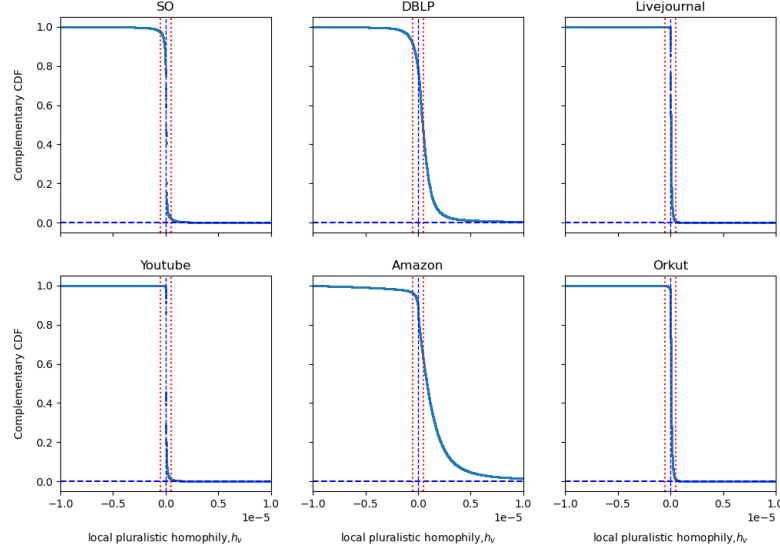
Fig. 2: CCDF local pluralistic homophily for each dataset. The x-axis range in each subplot is deliberately narrow, from $-1 \cdot 10^{-5}$ to $1 \cdot 10^{-5}$, to provide a close-up view of the h_v values at that level. It is important to note that the narrow x-axis range does not necessarily indicate a low tendency of nodes to be linked but rather serves the purpose of zooming in on the figure for better visualization of the localized h_v values. The red vertical dotted lines denote the range value of h_v to be considered as non-assortative as $|\varepsilon| < 0.1 \cdot 10^{-5}$.

the m_v value is low or high. At first glance, the figures show different patterns but looking in detail there are similarities between some of them that can reveal interesting findings. For instance, the dispersion of points looks much larger in some networks as d_v increases. This is most evident at DBLP and Amazon. It can also be noted that this dispersion is marked by the value of h_v taking the zero value on the y-axis as the initial point of said dispersion, which is less accelerated in Livejournal and Orkut, and a little less observable in SO.

We observed that, as the degree of the node d_v increases, the points become more dispersed. However, it is important to note that these observations are based on a sample rather than the entire population. While we did not observe a significant trend in the proportion of low and high values of m_v across all figures, suggesting that the distribution of m_v may not be strongly correlated with the degree of the node, further statistical analysis is needed to establish a conclusive relationship. It is worth noting that in the DBLP dataset, there appears to be a potential trend of more high values of m_v beyond a certain value of d_v, while in the SO dataset, the opposite trend is observed, with lower values of m_v after a

certain point. For now, this is an initial analysis that requires going further to determine the statistical significance of the relationships.
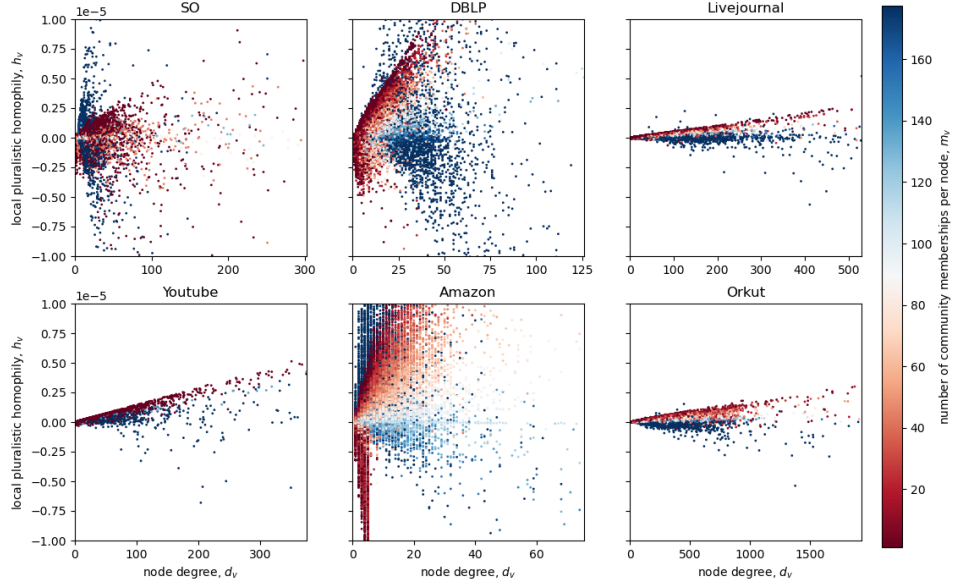

Fig. 3: A scatter plot of node degree d_v versus pluralistic homophily h, with the number of community memberships per node m_v represented by dot colors. A random sample of 100,000 nodes is plotted. In this representation, darker shades of blue indicate higher values of m_v, while brighter shades of red indicate lower values of m_v. Please note that the x-axis limits vary in each figure, with the range set from 0 to $\sigma_{d_v} + 10$ standard deviations of d_v. The sample selection is based on a random sampling approach, ensuring an unbiased representation of nodes in the network.

In our analysis, we aim to examine the relationship between h_v and m_v and identify patterns in the distribution of nodes across the scatter plots. Specifically, we investigate the concentration of nodes with high h_v and low h_v values and how they relate to m_v. We examine the behavior of average m_v values in relation to h_v and how these values are represented in the scatter plots. To better understand this relationship, we analyze different regions of the scatter plots. We observe that, on average, m_v values rise until reaching a certain point, which corresponds to the average values of m_v. This behavior occurs similarly at low values of h_v. However, in this case, an inverse behavior seems to emerge, indicating a different relationship between h_v and m_v. By conducting a thorough

analysis and examining the scatter plots, we aim to gain insights into the specific trends and relationships between h_v, m_v, and the distribution of nodes. This analysis will provide a more comprehensive understanding of the interplay between these variables and their implications within the network. The patterns in the figure show similarities between a group of networks comprising SO, DBLP, and Amazon, versus a second group comprising the Livejournal, YouTube, and Orkut networks. In the first group, the nodes with the highest pluralistic homophily start at the first values of d_v, but as we move along their x-axis, these values tend to decrease. However, at some point on the said axis, the high values of h_v begin to appear rapidly in nodes with low m_v, and then surprisingly rise again until reaching the average values of m_v. This behavior occurs similarly at low values of h_v. However, in this case, it seems an inverse behavior, that is, they start with low values of m_v first and then increase as the degree of the node increases. In the second group of networks, the patterns are different from the networks in the first one. In this case, the high values of pluralistic homophily are in nodes with low m_v while the low ones are in higher proportion in nodes with high values of m_v. This is maintained along the x-axis however with a greater spread of points as d_v increases, forming a kind of "comet" going left to right on such axis. In general, the networks in the first packet exhibit a great dispersion on the pluralistic homophily of the nodes with a relatively low degree d_v, while the second one show that the dispersion along moves all the node degrees. Also, it could be said that the areas with a high density of values of h_v under the ε range indicate that the nodes have a greater diversity in terms of connections with nodes of a different number of community membership, i.e. they link to other nodes indistinctly if share the same number of community memberships. On the other side, nodes with higher or lower values of ε indicate a greater homogeneity in the connections. We follow to discuss these findings according to the type and characteristics of the networks analyzed.

5   Discussion

Our findings reveal interesting patterns in the behavior of local pluralistic homophily values across the analyzed networks. The distribution of these values exhibits a bimodal shape, with two modes representing assortative and non-assortative connections between nodes. However, in the Amazon and DBLP networks, we observe a slightly higher number of nodes connecting with others who share a similar number of community memberships. This suggests that areas with the highest and lowest overlapping of communities exhibit a comparable density of nodes. This behavior may be influenced by the network type, as both Amazon and DBLP networks are informational in nature. In contrast, social networks display different behavior. The most connected nodes, or hubs, in these networks, exhibit high and low levels of overlap, belonging to multiple or few communities. Interestingly, this does not seem to significantly affect the proportion of nodes belonging to many or few communities.

The shape of the local pluralistic homophily distribution provides insights into the presence or absence of specific homophily patterns in the network. In information or collaboration networks, nodes are more likely to connect with others who share their interests or work areas, resulting in a greater similarity in the number of communities they belong to. In social networks, on the other hand, the diversity of interests and connections tends to be higher, explaining the concentration of the cumulative distribution around zero. The greater heterogeneity in the number of community memberships in social networks leads to a wider dispersion of local pluralistic homophily values. Surprisingly, we did not find a clear relationship between network size and the local pluralistic homophily exhibited by its nodes. Instead, our results indicate that this relationship depends more on the network type and its structural characteristics. Different patterns emerge, indicating a varying tendency for nodes to connect with others having the same level of overlap. We also observed distinct patterns between the number of connections and memberships of nodes in specific areas of the networks.

These observations highlight the unique behavior of pluralistic homophily and its potential applications in network analysis. It is evident that the behavior of pluralistic homophily carries significant value in understanding network dynamics and uncovering underlying social phenomena. However, further research is needed to delve deeper into the specific mechanisms driving these observations and explore their implications in different contexts.

6   Conclusions and Future Work

We have presented a measure that serves to estimate the tendency of nodes to connect with others that share a similar level of overlap, that is when they are members in similar numbers of various communities. Our experiments have shown what is the relationship between the said level of overlap with the degree of the nodes in the network. Likewise, despite the fact that networks can exhibit a general behavior in this sense, this is different from what is presented at the level of each node, and this varies depending on where such node is located in the network structure. A measure is then presented which allows us to consistently estimate the pluralistic homophily of an entire network and at the level of each node present in it. We consider this as an important contribution to the discussion of assortativity in general since assortativity based on the overlapping of communities allows for analysis different from the widely explored assortativity based on the degree of nodes. This leads us to think about possible future works, such as the use of the local pluralistic homophily measure in methods for various applications, such as the prediction of the properties of a given node or the prediction of links between nodes in a network. These methods range from the use of distance-based algorithms as a weight in the similarity between nodes and links, as well as the construction of feature representation vectors to be used in machine-learning models to make such predictions.

References

1. Newman, M. E. J. (2002). Assortative Mixing in Networks. Physical Review Letters, 89(20), 1–4. https://doi.org/10.1103/PhysRevLett.89.208701
2. Newman, M. E. J. Mixing Patterns in Networks. Physical Review E 67, no. 2 (February 27, 2003): 026126. https://doi.org/10.1103/PhysRevE.67.026126.
3. Barabási, Albert-Laszlo (2014). Network Science Communities.
4. Yang, J., Leskovec, J. (2012). Structure and Overlaps of Communities in Networks, 12. Retrieved from http://arxiv.org/abs/1205.6228
5. Piraveenan, M., Prokopenko, M., Zomaya, A. Y. (2008). Local assortativeness in scale-free networks. Epl, 84(2), 2008–2011. https://doi.org/10.1209/0295-5075/84/28002
6. Danziger, M. M., Wang, Z., Havlin, S., & Boccaletti, S. (2016). Assortativity and leadership emerge from anti-preferential attachment in heterogeneous networks. Scientific Reports, 6(1), 1-7. https://doi.org/10.1038/srep21297
7. Tan, F. Xia, Y. Zhu, B. (2014) Link Prediction in Complex Networks: A Mutual Information Perspective. PLOS ONE 9(9): e107056. https://doi.org/10.1371/journal.pone.0107056
8. Noldus, R. Mieghem, P. (2015) Assortativity in complex networks. Journal of Complex Networks 3, 507–542 doi:10.1093/comnet/cnv005
9. Thedchanamoorthy, G. (2014). New approaches and their applications in measuring mixing patterns of complex networks. The University of Sydney. https://ses.library.usyd.edu.au/handle/2123/13211
10. Ahn, Y., Bagrow, J. P., Lehmann, S. (n.d.). Link communities reveal multi-scale complexity in networks.
11. Leskovec, J., & Krevl, A. (2014). SNAP Datasets: Stanford Large Network Dataset Collection. Retrieved from http://snap.stanford.edu/data