Homophily explains perception biases in social networks

Eun Lee1,+, Fariba Karimi2,3,+,*, Hang-Hyun Jo4,5,6, Markus Strohmaier2,7, and Claudia Wagner2,3

1Department of Energy Science, Sungkyunkwan University, Suwon, 16419, Republic of Korea
2Department of Computational Social Science, GESIS, Cologne, 50667, Germany
3Institute for Web Science and Technologies, University of Koblenz-Landau, Koblenz, Germany
4Asia Pacific Center for Theoretical Physics, Pohang 37673, Republic of Korea
5Department of Physics, Pohang University of Science and Technology, Pohang 37673, Republic of Korea
6Department of Computer Science, Aalto University, Espoo FI-00076, Finland
7HumTec Institute, RWTH Aachen University, Germany

+These authors contributed equally to this work.
*Corresponding author. Email: fariba.karimi@gesis.org

ABSTRACT

Individual’s perceptions about the prevalence of attributes in their social networks is commonly skewed by the limited information available to them. Filter bubbles – being exposed to other like-minded people – and majority illusion – overestimation of minorities in social networks – are two examples of how perception biases can manifest. In this paper, we show how homophily and disproportionate group sizes influence the emergence of perception biases in social networks. Using a generative network model with adjustable homophily and group size, we demonstrate analytically and numerically under which conditions and to what extent perception biases can emerge. We compare these theoretical results with empirical investigations of perception biases in six real-world networks with various levels of homophily and group sizes. Our results show (i) that perception biases can emerge in social networks with high homophily or high heterophily and unequal group sizes, (ii) that these effects are highly related to the asymmetric nature of homophily in networks and that (iii) the perception of nodes is not systematically distorted or enhanced by their degree. Finally, we explore under which structural conditions individuals can reduce their perception bias by taking the perception of their direct neighbors into account. These results advance our understanding of the impact of network structure on perception biases and offer a quantitative approach to address this issue in society.

Introduction

In social networks, an individual’s cognitive ability to maintain meaningful social relationships with a large population is fundamentally limited1. Furthermore, people seldom draw representative samples from the overall population since they tend to cluster with others with whom they share common interests or attributes. Connecting to similar others, also known as homophily, is considered one of the most fundamental building blocks of social networks2. These cognitive and structural limits typically constrain the individual’s ability to accurately estimate the prevalence of attributes in social networks (e.g., gender or religion)3. As a result, perceptions of the prevalence of attributes can be potentially inaccurate and biased.

Understanding what factors impact perception biases is of crucial importance since they can have effects on human judgment4,5, the self-esteem of groups6 and social behaviors such as the adoption of behavior and attitudes in social networks7–12. More importantly, a wide perception gap among groups can generate extreme behavior such as violence, in particular among the minority13–16.

In social networks, perception biases can manifest themselves in various ways. A prominent example is the so-called filter bubble, where like-minded people repeatedly interact, reinforcing their own beliefs and opinions17. For example, during the U.S. presidential election researchers observed that Twitter users, including journalists, isolated themselves from other users who supported a different candidate, creating strong insular groups18,19. In such segregated communities, each group is exposed to news that strengthens their own beliefs and opinions, which may lead to a group-level misperception. This phenomenon is known as one of the reasons why many political pundits failed to predict the result of the recent U.S. presidential election showing a judgmental error caused by the perception bias20–22.

The majority illusion is another type of perception bias that occurs in social networks when a minority of people have a high degree of connectivity, such as journalists, politicians or celebrities. The opinion of the minority is systematically
over-represented in the public, creating a majority illusion\textsuperscript{23}. In this case the minority opinion can impact the opinion of the majority, since the majority perceives the minority as majority. For example, a survey recently reported that the majority of Europeans overestimate the Muslim population in their country presumably because of a wide exposure of media coverage on terrorist attacks in recent years\textsuperscript{24}.

Filter bubble and majority illusion are two examples in which the structure of the social network in the presence of disproportionate group sizes can alter the individual’s perception. In fact, many studies have pointed out the importance of group sizes on individual’s perception\textsuperscript{5,25–27} and the effects of degree heterogeneity, degree-degree and degree-attribute correlation on perception biases such as the majority illusion\textsuperscript{23}, friendship paradox\textsuperscript{28,29}, and generalized friendship paradox\textsuperscript{30,31}. However, most previous works have overlooked the impact of attribute-attribute correlations (also known as homophily or heterophily) and disproportionate group sizes on the emergence of perception biases.

In this paper, we show analytically and empirically the extent to which attribute-attribute correlation (heterophily/homophily) and disproportionate group sizes impact how accurate individuals can estimate the prevalence of an attribute in their social network. More specifically, we make three contributions. First, we demonstrate the existence of perception biases in six empirical networks with various levels of homophily and minority sizes. Second, we propose a generative network model with tunable homophily and group sizes that allows us to systematically study the perception bias as a function of homophily to explain the perception biases we observe in empirical networks. Through this model we elucidate under which structural conditions majority perceives a minority as the majority (majority illusion) and a minority perceives itself as a majority (filter bubble). Third, we show how the perception biases can be reduced by aggregating individual’s perception with neighborhood information.

**Quantifying perception biases**

In this paper, we focus on networks where nodes have binary attributes (e.g. black/white, male/female) that are unequally distributed in a population, and this leads to the formation of two unequal sized groups. Given the binary attributes, we define a majority and a minority group by comparing their group size such that the majority group is relatively larger than the minority group. Then, we define the perception bias on two levels: the individual- and the group-level.

To estimate the perception bias of a node on an individual-level, $P_{\text{indv}}$, we assume that node $i$ simply counts how often an attribute appears in its direct neighbourhood to estimate the prevalence of the attribute in the whole population. That means, we assume that nodes perceive the prevalence of an attribute in their direct neighbourhood and only rely on this information to inform their estimate. We focus on the estimation of the minority attribute, since the results for the majority attribute are complementary for binary attributes. Therefore, we measure to what extent the fraction of minority nodes in the local neighbourhood of a node $i$ deviates from the true fraction of minority nodes, $f_M$. For node $i$, we define the individual-level perception bias as follows:

$$
P_{\text{indv}} = \frac{1}{f_M} \sum_{j \in \Lambda_i} x_j / k_i \label{eq:indv}
$$

where $k_i$ is the degree of node $i$, $\Lambda_i$ is a set of neighborhood of node $i$, $x_j$ is the attribute of node $i$’s neighbor $j$ (1 for minority and 0 for majority) and $f_M$ is the true fraction of minority nodes in the entire network.

The group-level perception bias is defined in a similar way as the individual-level perception bias by aggregating the perceptions of all nodes in a group:

$$
P_{\text{group}} = \frac{1}{f_M} \sum_{i \in N_g} \sum_{j \in \Lambda_i} x_j / \sum_{i \in N_g} k_i \label{eq:group}
$$

where $N_g$ is the set of nodes in a group $g$. The intuition behind the group-level perception is that the group is considered as one bulk containing all members of the group. In other words, group-level perception reflects the collective perception of the group. This convention makes analytical calculations easier at the group level by using a mean-field approach and has been used in other studies that explored the generalized friendship paradox in networks\textsuperscript{30,31}.

Unless otherwise stated, we will use $P_{\text{indv}}$ and $P_{\text{group}}$ to address individual- and group-level perception biases. If it is necessary to distinguish between minority and majority, we will add a superscript such as $M$ for the minority and $J$ for the majority. By this definition, perception bias can vary from 0 to 1 $f_M$, i.e., $P_{\text{group}} \in [0, 1/f_M]$ and $P_{\text{indv}} \in [0, 1/f_M]$. A perception bias value below 1 indicates an underestimation of the group size, while values above 1 indicate overestimation. If the value equals 1, no perception bias exists and the group or the individual perfectly perceives the true fraction of minority nodes. Furthermore, the perception gap between groups can be defined simply as an absolute difference between perception biases of the two groups: $|P_{\text{group}} - P_{\text{group}}^M|$.

As an example, Fig. 1 illustrates how we define the perception bias on the individual-level and group-level for homophilic and heterophilic networks. Here, the color of a node depicts its group membership. Orange nodes belong to the minority and
blue nodes are in the majority. The central node \( i \) estimates the size of the minority group based on the fraction of orange nodes in its local neighborhood (dashed circle line). The individual-level perception bias of the majority node \( i \) is \((1/6)/(1/3) = 0.5\) in the homophilic network (Fig. 1(a)), which means that node \( i \) underestimates the size of the minority by the factor 0.5. Consequently, it overestimates the size of its own group. In the heterophilic network (Fig. 1(b)), the perception bias of node \( i \) is \((4/6)/(1/3) = 2\), implying that the node overestimates the size of the minority by a factor of 2. On a group-level, the majority group (blue nodes) perceives the size of the minority as 4/18 in the homophilic network and 10/18 in the heterophilic network, respectively. Therefore, the majority group underestimates the size of the minority group \((4/18)/(1/3) = 0.67\) times in the homophilic network and overestimates it by a factor of \((10/18)/(1/3) = 1.67\) in the heterophilic network.

**Generative network model with tunable homophily and group size**

To gain insights into how perception biases manifest in social networks, we develop a network model that allows us to create scale-free networks with tunable homophily and group sizes. This network model is a variation of the Barabási-Albert model with the addition of homophily parameter \( h \). In this model, the probability that a newly introduced node connects to an existing node \( i \), denoted by \( p_i \), is proportional to the product of the degree \( k \) and homophily parameter \( h \): \( p_i \propto kh \).

Thus, degree and homophily parameter regulate the probability to connect to nodes that share the same attribute value. The value of homophily ranges from 0 to 1. If the homophily is low \((0 \leq h \leq 0.5)\), nodes have a higher tendency to connect to nodes with opposing attribute values. As the homophily parameter increases, the probability of connectivity among nodes with the same attribute increases. In extreme homophily \((h = 1)\) nodes are strictly connecting to other nodes with the same attribute value and therefore two separated communities emerge. One advantage of using this model is that it generates networks with scale-free degree distributions observed in many large-scale social networks. The second advantage is that it generates networks in which the homophily parameter can be symmetric or asymmetric. Given nodes with two attributes, \( a \) and \( b \), in symmetric homophily the tendency to connect to nodes of the same attribute is the same for both groups, \( h_{aa} = h_{bb} = h \), therefore one parameter is sufficient to generate the network. In the asymmetric homophily case, two homophily parameters are needed to regulate the connectivity probabilities for each group separately since \( h_{aa} \neq h_{bb} \).

Figure 2 depicts the perception of the minority size from the perspective of the minority group (Fig. 2(a)) and the perspective of the majority group (Fig. 2(b)) for different minority fractions \((f_M)\) and homophily values \((h)\). The solid lines show the analytical results (see Method section) and the circles are numerical results. The results indicate that in heterophilic networks \((0 \leq h \leq 0.5)\) the minority underestimates their own size and the majority overestimates the size of the minority due to high connectivity of minority to majority (majority illusion). The overestimation increases as the minority size \((f_M)\) decreases. In homophilic networks \((0.5 \leq h \leq 1)\), the minority overestimates their own size and the majority underestimates the size of the minority due to high connectivity among members of the same group (filter bubble). The numerical results are in excellent agreement with the analytical results explained in the Method section.

So far, we observe, analytically and numerically, that homophily and group size differences play a crucial role in shaping perception biases in synthetic networks. Next, we explore six real networks with various levels of homophily and group size differences, quantify the perception bias in these networks and examine to what extent our model can capture the perception bias observed in real-world networks.

**Results**

**Perception biases in real-world networks**

To examine perception biases in real-world networks, we study six empirical networks with various level of homophily and minority sizes. The network characteristics of the data are presented in Table 1. The detailed descriptions of the dataset collections and references can be found in the Method section. These empirical networks have different structural characteristics and show different levels of homophily or heterophily with respect to one specific attribute (see the Supplementary section). We measure the symmetric homophily and the asymmetric homophily in all networks. The symmetric homophily is equivalent to Newman’s assortativity measure \((r)\) that measures the Pearson coefficient of the edges that exist between nodes with different attribute values compared to what we would expect from their degree. The Newman assortativity measure corresponds directly to the homophily parameter in our model when adjusted for the scale. In our model \( h = 0 \) means complete heterophily \((r = -1)\), \( h = 0.5 \) indicates no relationship between structure and attributes \((r = 0)\), and \( h = 1 \) indicates complete homophily \((r = 1)\) (see the Supplementary section).

The value of symmetric homophily for each network is as follows: \( h_{Brazil} = 0 \) in the Brazilian sexual contact network, \( h_{PSSO} = 0.15 \) in the online dating network PussOKram, \( h_{USF51} = 0.47 \) in the Facebook friendship network, \( h_{DBLP} = 0.56 \) in the scientific collaboration network, \( h_{GitHub} = 0.6 \) in the Github mutual follower network and \( h_{APS} = 0.74 \) in the American Physical Society citation network. Homophily is measured between genders (female and male) except in the APS network in which homophily \( h_{APS} \) is measured between different academic fields namely classical statistical mechanics (CSM) and quantum statistical mechanics (QSM).
In reality, the tendency of groups to connect to other groups can be asymmetrical. Given the number of edges that run between nodes of the same group and our model parameters, we can estimate the asymmetric homophily in the empirical networks analytically.\textsuperscript{32} As we will see shortly, it turns out that asymmetric homophily has an important impact on the predictability of perception bias in empirical networks.

We use the measured homophily and minority size in the empirical networks to generate synthetic networks with similar characteristics. This will enable the comparison of perception bias in the empirical networks with perception bias in synthetic networks. This comparison can also be seen as a validation of the network model. If the model explains the biases in the real-world networks well, we can use the model to gain insights about how homophily and group size differences impact individual-level and group-level perception biases.

Figure 3 shows the group-level perception biases in the empirical networks. Here we measure how accurately the minority and majority groups estimate the size of the minority group. The empirical value of perception bias is shown by the cross signs. As expected from the insights gained from the synthetic networks, we observe that in the heterophilic networks the minority group underestimates their own group size and the majority overestimates the size of the minority. Conversely, in the homophilic networks, the minority overestimates their own size and the majority underestimates the size of the minority. The underestimation and overestimation increase when homophily or heterophily is higher. The empirical perception bias in real-world networks is then compared with perception bias in synthetic networks with similar size and a similar level of homophily.

When using symmetric homophily (triangles in fig. 3), our model does well at estimating perception bias for the majority of networks, except Github and APS networks which exhibit a higher level of asymmetry in the homophily among the groups (see Table 1). To investigate whether clustering plays a role in determining the group perception, we modify the synthetic homophilic network model to account for more realistic clustering coefficients similar to the empirical networks. We modify the network model similar to Holme-Kim method in which the connectivity follows the same principle as homophilic network model in addition to closing triads in a stochastic fashion.\textsuperscript{34} When correcting for the number of triangles in the model, our results do not show any improvement compared to the network model without correct clustering (see Supplementary material). We then generate the synthetic model with asymmetric homophily (rectangular shapes in fig. 3) and observe that the model can estimate perception bias very well for all six empirical networks. This suggests the configuration of our model using asymmetric homophily is capable of estimating perception biases of groups accurately.

In addition, we examine the relationship between the individual perception of nodes and their degree. Here, we address the question whether the accuracy of individual estimates of the minority group size changes as a function of the degree of nodes. More specifically, we explore if high or low degree nodes are better at predicting the size of the minority in a social network. Figure 4 shows the average individual perception versus degree $k$. Heterophilic networks are shown in the top row and are ordered with decreasing heterophily. Homophilic networks are shown in the second row and are ordered with increasing homophily. We observe that perception bias is rather stable with respect to degree. This suggests that there is no systematic relationship between the degree of a node and the accuracy of its social perception. We do not find any empirical evidence that high degree nodes are better at estimating the prevalence of attributes in their social networks. It is interesting to note that the perception gap between the two groups (i.e. the distance between the blue and the orange line) increased with increasing levels of heterophily and homophily. Although the relation between the degree and the individual-level perception bias does not follow a simple pattern compared to synthetic networks (see Supplementary material), our model (dashed lines) can predict the trend in the data accurately, in particular when we use asymmetric homophily.

### Social Influence and Perception Bias

Previous studies suggest that the accuracy of individual judgments can be improved when taking individual’s average judgments who are organized in a decentralized network when people make decisions independently or communicate with those who have negatively correlated beliefs.\textsuperscript{35, 36} If people inside the group are not independent and can observe the judgment of their peers, then social influence may lead to group thinking and decrease the accuracy of the group, especially in centralized networks.\textsuperscript{37}

Here, we investigate to what extent and under which structural conditions individuals can improve their estimates of the prevalence of an attribute in a population by asking their friends for advice. We build on DeGroot’s weighted belief formalization by aggregating one’s perception (ego) with the averaged perceptions of the individual’s direct neighbors (1-hop) and explore how the accuracy of individual’s perception changes as a function of homophily.

Figure 5 compares the average ego perceptions from the perspective of (a) minority and (b) majority nodes compared with the weighted averages of the perceptions of their 1-hop neighbors. The minority size is fixed to 0.2. The results show that the estimates of nodes in both groups become more accurate in heterophilic networks when nodes take the estimates of their direct peers into account (green dots are closer to the gray dashed line compared to purple dots in the log-scale). The improvement in the perception is due to the fact that nodes are exposed to neighbors with diverse attributes and therefore accounting for the neighborhood perception results in improving the ego perception. In homophilic networks, ego’s perception does not show a
significant improvement when considering the neighborhood perception since nodes are exposed to neighbors with similar attributes.

This suggests that inside their filter bubbles individual’s perception cannot be enhanced by their peers since they are similar to them and do not add new information to increase the accuracy of individual’s perception. However, in heterophilic networks, individuals benefit from considering their neighbors’ perception since they are not only exposed to other like-minded individuals. While the overall trend is not surprising, our results reveal how the accuracy of these estimates changes as a function of homophily.

**Discussion**

In this paper, we made a first attempt to examine how the structure of a social network driven by homophily can aggravate perception biases of groups and individuals. We focused on the perception of the prevalence of an attribute (e.g. gender) in a social network and assumed that nodes in this network only perceive the prevalence of this attribute based on their direct neighbourhood. We defined perception bias by how accurate individuals and groups estimate the size of the minority in their social network. We showed that depending on the level of homophily, groups overestimate or underestimate the size of the minority group. Therefore, homophily impacts the accuracy of the estimates of individuals in both groups. Concretely, as homophily increases in a social network, nodes of the minority group overestimate their group size, and nodes of the majority group underestimate the size of minorities. In this case minorities are subject to a false consensus since they will overestimate the extent to which their opinions, beliefs, preferences, values, and habits are normal and typical compared to others\(^{39, 40}\). On the contrary, majority may ignore the opinion and values of minorities resulting in segregation among groups. Consequently, these differences in values and norms can cause extreme behavior among minorities\(^{16, 41}\). In heterophilic networks, individuals of the majority overestimate the size of the minority and individuals of the minority underestimate their own size. This leads to situations in which the majority inaccurately infers the prevalence of a certain behavior in the population which may affect the adoption of specific attitudes, opinions and behavior\(^ {42}\).

Using six empirical networks with various levels of homophily and minority sizes, we found that perception biases at the level of groups or individuals strongly depends on the extent of homophily or heterophily in social networks. To explain the underlying structural mechanisms for perception biases in social networks, we compared the observed perception biases in real-world networks with perception biases in synthetic networks with tunable homophily and group sizes. We find that our network model can explain the empirical observations very well, especially when asymmetric homophily is assumed. With that, we conclude that one of the most crucial elements in shaping group perception in real-world social networks is homophily which can be symmetric or asymmetric among groups. This finding highlights the importance of considering homophily and group sizes in modeling realistic social networks with attributes.

Our results revealed that in heterophilic networks where individuals with opposing attributes are more likely to be connected, perception biases can be reduced by aggregating individual’s perception with the perception of their direct neighborhood to inform their own estimate. In homophilic networks, however, these socially informed estimates do not lead to more accurate perception.

In agreement with previous findings that have pointed out the negative consequences of large perception gaps between groups (in particular for minority groups)\(^ {15, 16}\), we have shown analytically and numerically that structural segregation and large group size differences can increase perception gaps. Future studies are needed to validate and examine perception biases as social experiments or lab experiments. We hope that this paper offers insights into how to measure and reduce perception gaps between groups and fuels more work on understanding the impact of network structure on individual’s perception in society.

**Method**

**Analytical derivation for group-level perception bias**

Our network model is a generative growth model of graphs with homophily and preferential attachment\(^ {32}\). Let us refer to the minority as \( M \) and the majority as \( J \). \( K_M(t) \) and \( K_J(t) \) are the total number of degrees for each group of the minority and the majority at time \( t \) respectively. At each time step, one node arrives and connects with \( m \) existing nodes in the network. Therefore, the total degree of the growing network at time \( t \) is:

\[
K(t) = K_M(t) + K_J(t) = 2mt,
\]

(3)
The overall evolution of the degree for each group, $K_M$ and $K_J$, in the limit of $\Delta t \to 0$ can be written as:

\[
\begin{align*}
\frac{dK_M}{dt} &= m \left( f_M \left( 1 + \frac{hK_M(t)}{hK_M(t) + h'K_J(t)} \right) + f_J \frac{h'K_M(t)}{hK_J(t) + h'K_M(t)} \right) \\
\frac{dK_J}{dt} &= m \left( f_J \left( 1 + \frac{hK_J(t)}{hK_J(t) + h'M(t)} \right) + f_M \frac{h'M(t)}{hK_M(t) + h'K_J(t)} \right),
\end{align*}
\]

where $h$ is the homophily parameter regulating connectivity among nodes of the same group and $h' = 1 - h$ is the complementary homophily among nodes of different groups. In this case the model is only dependent on one $h$ parameter meaning that homophily among members of the same group is symmetric. $f_M$ represents the fraction of the minority and $f_J = 1 - f_M$ is the fraction of the majority. Since $K_M(t)$ and $K_J(t)$ are linear functions of time they can be written as:

\[
\begin{align*}
K_M(t) &= Cmt \\
K_J(t) &= (2 - C)mt.
\end{align*}
\]

Plugging Eq. 5 into Eq. 4, we get

\[
\frac{dK_M}{dt} = m \left( f_M \left( 1 + \frac{hCmt}{hCmt + h'(2mt - Cmt)} \right) + f_J \frac{h'Cmt}{h(2mt - Cmt) + h'Cmt} \right),
\]

where $C$ is the growing rate of the degree of minority $C \in [0, 2]$. $C$ is a third-order polynomial function and can be solved numerically (see the Supplementary material). For the minority group, the overall evolution of degree for a node ($k_M$) is:

\[
\frac{dk_M}{dt} = \frac{k_M}{\tau} \left( \frac{f_M h}{hC + h'(2 - C)} + \frac{f_J h'}{h'C + h'(2 - C)} \right).
\]

Similarly, for the majority group the overall evolution of degree for a node ($k_J$) is:

\[
\frac{dk_J}{dt} = \frac{k_J}{\tau} \left( \frac{f_J h}{h'C + h'(2 - C)} + \frac{f_M h'}{h'C + h'(2 - C)} \right).
\]

In Eq. 7, the first term inside the parenthesis displays the average probability of a connection between a minority node to another minority node and the second term depicts to the probability of a connection between a minority and a majority node. Thus the probability of a minority connecting to another minority is $p_{M,M} = f_M h / (hC + h'(2 - C))$. As as result, the group-level perception of minority about their own size, $p_{group}^M$, would be:

\[
p_{group}^M = \frac{1}{f_M} \frac{p_{M,M}}{2p_{M,M}} = \frac{1}{f_M} \frac{2f_M h}{hC + h'(2 - C)}. \tag{9}
\]

The analytical calculation is shown in Fig. 2(a) as solid lines. One can see that the minority reveals larger overestimation in homophily regime ($h > 0.5$) compared to what we would expect from the baseline. The analytical derivation is intuitive. For example, when $f_M = 0.5$ in an extreme homophily such as $h = 1.0$, the degree growth is $1 \cdot C = 1$. Note that even though $p_{group}^M$ does not depend on the group size, the relation between the group size and the group perception is encapsulated in the parameter $C$ (see the Supplementary material). In this case $p_{group}^M = 2$ as can be seen in Fig. 2(a).

Similarly, we can compute the perception bias for the majority group $p_{group}^J$ as follows:

\[
p_{group}^J = \frac{1}{f_M} \frac{p_{M,J}}{p_{M,J}} = \frac{1}{f_M} \left( \frac{C}{2 - C} \frac{f_J h'}{h'C + h'(2 - C)} + \frac{f_M h'}{h'C + h'(2 - C)} \right), \tag{10}
\]

The analytical $p_{group}^J$ are shown in Fig. 2(b) with different minority sizes $f_M$, and match the numerical values very well. Since $p_{group}^M$ is strongly proportional to $h$ and $p_{group}^J$ relies on $h'$, a reversed relation between Fig. 2(a) and (b) can be observed. In the case of complete heterophily ($h = 0$) with $C = 1$ regardless of $f_M$, $p_{group}^M$ and $p_{group}^J$ are reduced to

\[
\begin{align*}
p_{group}^M &= 0 \\
p_{group}^J &= \frac{1}{f_M} \left( \frac{f_J}{2 - C} + \frac{f_M}{2 - C} \right) = \frac{1}{f_M} \left( \frac{1}{2 - C} \right) = \frac{1}{f_M}. \tag{11}
\end{align*}
\]
Relation between individual-level perception bias and degree

Here, we investigate the relation between the degree of nodes and the perception bias. We aim to uncover how the degree connectivity of nodes improves or worsens the perception bias. For this, we define degree-level perception, as an average individual-level perception among nodes with the same degree $k$ (Eq. 1). Figure S5 shows the perception bias as a function of degree for majority and minority nodes. Additionally, we show the perception gaps between groups on the last row.

To gain insights on the relation between the degree and perception bias in Fig. S5, it is useful to consider the process of adding links to the generative network model. The link formation in this network is based on the combination of the preferential attachment mechanism that is governed by degree $k$ and the homophily mechanism $h$. In addition, if the group sizes are unequal that would also lead to differences in the degree growth of the groups. Therefore, three parameters $k, h, f_M$ should be considered to explain the relation between the degree and perception bias.

Let us now consider a homophilic network ($h = 0.9$) with $f_M = 0.1$ as an example (see the red line in Fig. S5(a) and (d)). The overall trend for the perception of the minority shows that their perception becomes worse as their degree increases. In a highly homophilic network, popular minority nodes can mainly attract other minority nodes, since they cannot compete with popular majority nodes that accumulate much higher degree. For the majority in a homophilic network (red line in Fig. S5(d)), the perception improves as degree increases. This is because due to a larger size of the majority, the most popular majority nodes occasionally attract minority nodes. As the group sizes become equal, the dependency between degree and perception vanishes and the perception becomes only dependent on the homophily and group size $P_{adv.} = \frac{1}{2} h$.

Let us now turn to the heterophilic network case ($h = 0.1, f_M = 0.1$, see the blue line in Fig. S5(a) and (d)). For minority, the overall trend shows that as degree increases, the perception of the minority becomes worse. This is due to their smaller size and the nature of heterophilic connection, minority’s degree can grow large and attract more and more majority and thus the underestimation of their own group size becomes worse. Conversely, for the majority, perception bias improves as degree increases. This is because despite the fact that they are more attracted to the minority since the minority size is limited, higher degree majority have higher chance to attract other majority nodes occasionally and thereby their perception estimation improves. Again, as the group sizes become equal, the dependency between degree and perception vanishes and the perception become only dependent on the homophily and group size $P_{adv.} = \frac{1}{2} h$.

These results suggest that there is an interplay between group sizes, homophily, and the degree in the linking process. Considering the perception gap between groups (Fig. S5(g), (h) and (i)), these results emphasis on the importance of adjusting group sizes or reducing the extreme homophily and heterophily to achieve a lower perception gap between groups.

Empirical networks

The first network captures sexual contacts between sex-workers and sex-buyers\(^{43}\). The network consists of 16,730 nodes and 39,044 edges. There are 10,106 sex-workers and 6,624 sex-buyers (minority size $f_i = 0.4$). In this network, no edges among members of the same group exist, and the network is purely heterophilic.

The second network is an online Swedish dating network from PussOKram.com (POK)\(^{44}\). This network contains 29,341 nodes with strong heterophily ($h = 0.15$). Based on the high bipartivity of the network, we are able to infer the group of nodes using max-cut greedy algorithm. The results are in good agreement with the bipartivity reported in \(^{45}\). Since the group definition is arbitrary, we label nodes based on their relative group size as ‘minority’ and ‘majority’. Here, the fraction of the minority in the network is 0.44.

The third network is a Facebook network of a university college in the United States (USF51)\(^{46}\). The network is composed of 6,253 nodes including information about the gender. In this network male students are in the minority occupying 42% of the network and network exhibit a low gender homophily\(^{46}\).

The fourth network depicts scientific collaborations in computer science and is extracted from DBLP\(^{47}\). We used a new method that combines names and images to infer the gender of the scientists with high accuracy\(^{48}\). We use on a 4-years snapshot for the network. After filtering ambiguous names, the resulting network includes 280,200 scientists and 750,601 edges (paper co-authorships) with 63,356 female scientists and 216,844 male scientists. This network shows a moderate level of homophily ($h = 0.56$).

The fifth network is extracted from the collaborative programming environment Github. The network is one snapshot of the community (extracted 04.08.2015) with information about the first name and family name of the programmers. We use name and family names similar to the previous method to infer the gender of the programmers. After removing ambiguous names, the network comprises of 120,338 men and 7,330 women. Here, women belong to the minority group and only represent 5% of the population. The network displays a moderate gender homophily 0.6.

The last network is a scientific citation network of the American Physical Society (APS). Citation networks depict an extent of attention to communities around different scientific fields. We use the PACS identifier to select papers on the same topics. Here, we chose statistical physics, thermodynamics and nonlinear dynamical system sub-fields (PACS = 05). Within a specific sub-field, there are many sub-topics that form communities of various sizes. To make the data comparable with our model,
we choose two sub-topics that are relevant, namely classical statistical mechanics (CSM - 05.20.-y) and quantum statistical mechanics (QSM - 05.30.-d). The resulting network consists of 1,853 scientific papers and 3,627 citation links. Among nodes, the number of the minority is 696 and for the majority is 1,157. Here, the minority group in these two sub-topics is CSM ($f_M = 0.38$). This network shows the highest homophiliy compared to other empirical datasets ($h = 0.74$).

References
1. Dunbar, R. I. M. Neocortex size as a constraint on group size in primates. *J. human evolution* **22**, 469–493 (1992).
2. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: Homophily in social networks. *Annu. review sociology* **27**, 415–444 (2001).
3. Galesic, M., Olsson, H. & Rieskamp, J. Social sampling explains apparent biases in judgments of social environments. *Psychol. Sci.* **23**, 1515–1523 (2012).
4. Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cogn. psychology* **5**, 207–232 (1973).
5. Manis, M., Shedler, J., Jonides, J. & Nelson, T. E. Availability heuristic in judgments of set size and frequency of occurrence. *J. Pers. Soc. Psychol.* **65**, 448 (1993).
6. Loughnan, S. *et al.* Economic inequality is linked to biased self-perception. *Psychol. Sci.* **22**, 1254–1258 (2011).
7. Botvin, G. J., Botvin, E. M., Baker, E., Dusenbury, L. & Goldberg, C. J. The false consensus effect: Predicting adolescents’ tobacco use from normative expectations. *Psychol. Reports* **70**, 171–178 (1992). URL https://doi.org/10.2466/pro.1992.70.1.171.
8. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. *Sci.* (2012). URL http://science.sciencemag.org/content/early/2012/06/20/science.1215842. DOI 10.1126/science.1215842.
9. Centola, D. The spread of behavior in an online social network experiment. *Sci.* **329**, 1194–1197 (2010). DOI 10.1126/science.1185231.
10. Christakis, N. A. & Fowler, J. H. The collective dynamics of smoking in a large social network. *The New Engl. journal medicine* **358**, 2249–2258 (2008). URL http://dx.doi.org/10.1056/nejmsa0706154. DOI 10.1056/nejmsa0706154.
11. Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proc. Natl. Acad. Sci.* **107**, 5334–5338 (2010). URL http://www.pnas.org/content/107/12/5334.abstract. DOI 10.1073/pnas.0913149107. http://www.pnas.org/content/107/12/5334.full.pdf.
12. Salganik, M., Dodds, P. & Watts, D. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Sci.* **311**, 854–856 (2006).
13. Tajfel, H. *Social identity and intergroup relations* (Cambridge University Press, 2010).
14. Kessler, T., Harth, N. S. & Nügler, L. A. Prejudice and extremism: Explanations based on ingroup projection, perspective divergence, and minimal standards. *Soc. Psychol.* **82**, 359 (2002).
15. Smith, H. & Ortiz, D. *Is it just me?: The different consequences of personal and group relative deprivation* (Cambridge University Press, 2002).
16. Bernholz, P. Supreme values as the basis for terror. *Eur. J. Polit. Econ.* **20**, 317–333 (2004).
17. Pariser, E. *The filter bubble: What the Internet is hiding from you* (Penguin UK, 2011).
18. Journalists and Trump voters live in separate online bubbles. https://news.vice.com/story/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows. 2017-12-08.
19. Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A. & Van Bavel, J. J. Emotion shapes the diffusion of moralized content in social networks. *Proc. Natl. Acad. Sci.* **114**, 7313–7318 (2017).
20. How Journalists Failed to Predict Trump’s Victory. https://www.nytimes.com/video/us/politics/100000004758375/how-journalists-failed-to-predict-trumps-victory.html. 2016-11-09.
21. Here’s Why the Media Failed to Predict a Donald Trump Victory. http://fortune.com/2016/11/09/media-trump-failure/. 2016-11-09, filter bubble.
22. How did journalists fail to predict Donald Trump’s election win? [http://www.cbc.ca/news/world/how-did-journalists-fail-to-predict-donald-trump-s-election-win-1.3848591](http://www.cbc.ca/news/world/how-did-journalists-fail-to-predict-donald-trump-s-election-win-1.3848591) 2016-11-12.

23. Lerman, K., Yan, X. & Wu, X.-Z. The “majority illusion” in social networks. *PloS one* **11**, e0147617 (2016).

24. Islam in europe. [http://www.economist.com/blogs/graphicdetail/2015/01/daily-chart-2](http://www.economist.com/blogs/graphicdetail/2015/01/daily-chart-2) (2015-01-07).

25. Kerr, N. L. Illusions of efficacy: The effects of group size on perceived efficacy in social dilemmas. *J. experimental social psychology* **25**, 287–313 (1989).

26. Wilder, D. A. Perception of groups, size of opposition, and social influence. *J. Exp. Soc. Psychol.* **13**, 253–268 (1977).

27. Bartsch, R. A. & Judd, C. M. Majority—minority status and perceived ingroup variability revisited. *Eur. J. Soc. Psychol.* **23**, 471–483 (1993).

28. Feld, S. L. Why your friends have more friends than you do. *Am. J. Sociol.* **96**, 1464–1477 (1991).

29. Wu, X.-Z., Percus, A. G. & Lerman, K. Neighbor-neighbor correlations explain measurement bias in networks. *Sci. Reports* **7** (2017).

30. Eom, Y.-H. & Jo, H.-H. Generalized friendship paradox in complex networks. *Sci. Reports* **4**, 4603 (2014).

31. Jo, H.-H. & Eom, Y.-H. Generalized friendship paradox in network with tunable degree-attribute correlation. *Phys. Rev. E* **90**, 022809 (2014).

32. Karimi, F., Génois, M., Wagner, C., Singer, P. & Strohmaier, M. Visibility of minorities in social networks. *arXiv preprint arXiv:1702.00150* (2017).

33. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).

34. Holme, P. & Kim, B. J. Growing scale-free networks with tunable clustering. *Phys. review E* **65**, 026107 (2002).

35. Page, S. E. *The difference: How the power of diversity creates better groups, firms, schools, and societies* (Princeton University Press, 2008).

36. Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci.* 201615978 (2017).

37. Lorenz, J., Rauhut, H., Schweitzer, F. & Helbing, D. How social influence can undermine the wisdom of crowd effect. *Proc. Natl. Acad. Sci.* **108**, 9020–9025 (2011).

38. DeGroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121 (1974).

39. Ross, L., Greene, D. & House, P. The “false consensus effect”: An egocentric bias in social perception and attribution processes. *J. experimental social psychology* **13**, 279–301 (1977).

40. Krueger, J. & Clement, R. W. The truly false consensus effect: an ineradicable and egocentric bias in social perception. *J. personality social psychology* **67**, 596 (1994).

41. O’gorman, H. J. Pluralistic ignorance and white estimates of white support for racial segregation. *Public Opin. Q.* **39**, 313–330 (1975).

42. Centola, D., Willer, R. & Macy, M. The emperor’s dilemma: A computational model of self-enforcing norms. *Am. J. Sociol.* **110**, 1009–1040 (2005).

43. Rocha, L. E., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol.* **7**, e1001109 (2011).

44. Holme, P., Edling, C. R. & Liljeros, F. Structure and time evolution of an internet dating community. *Soc. Networks* **26**, 155 – 174 (2004). URL [http://www.sciencedirect.com/science/article/pii/S0378873304000085](http://www.sciencedirect.com/science/article/pii/S0378873304000085). DOI [http://dx.doi.org/10.1016/j.socnet.2004.01.007](http://dx.doi.org/10.1016/j.socnet.2004.01.007).

45. Holme, P., Liljeros, F., Edling, C. R. & Kim, B. J. Network bipartivity. *Phys. Rev. E* **68**, 056107 (2003). URL [https://link.aps.org/doi/10.1103/PhysRevE.68.056107](https://link.aps.org/doi/10.1103/PhysRevE.68.056107).

46. Traud, A. L., Mucha, P. J. & Porter, M. A. Social structure of facebook networks. *Phys. A: Stat. Mech. its Appl.* **391**, 4165 – 4180 (2012). URL [http://www.sciencedirect.com/science/article/pii/S0378437111009186](http://www.sciencedirect.com/science/article/pii/S0378437111009186). DOI [http://dx.doi.org/10.1016/j.physa.2011.12.021](http://dx.doi.org/10.1016/j.physa.2011.12.021).

47. Dblp data. [http://dblp.uni-trier.de/](http://dblp.uni-trier.de/). Accessed: 2016-09-30.
48. Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M. & Strohmaier, M. Inferring gender from names on the web: A comparative evaluation of gender detection methods. In Proceedings of the 25th International Conference Companion on World Wide Web, 53–54 (International World Wide Web Conferences Steering Committee, 2016).

49. Alstott, J., Bullmore, E. & Plenz, D. Powerlaw: a python package for analysis of heavy-tailed distributions. PLoS ONE 9, e85777 (2014).
Acknowledgments: We thank Kristi Winters, Julian Kohne and Petter Holme for insightful conversations. Funding: We thank GESIS for funding E. Lee’s research visit. E.L. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT with grant No. NRF- 2017R1A2B2005957. H.-H.J. acknowledges financial support by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (2015R1D1A1A01058958). Author contributions: All authors conceived the project, developed the argument, and wrote the paper. E.L. and F.K. conducted experiments and wrote the code. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All the data and codes implementing the analysis can be found online at https://github.com/frbkrm/NtwPerceptionBias.
Figures and Tables

(a) Homophilic network
(Filter bubble)

(b) Heterophilic network
(Majority illusion)

Fig. 1. Individual-level and group-level perception bias. Nodes belong to one of the two groups (blue and orange); Blue nodes are the majority and orange nodes are the minority. The minority fraction is $1/3$ in both networks ($f_M \approx 0.33$). Subfigure (a) depicts a homophilic network, while subfigure (b) shows a heterophilic network. Perception bias is measured in two ways: an individual-level (grey, dashed circles) and on a group-level (grey, shaded area). On an individual-level, in the homophilic network, the central node $i$ perceives the size of the minority to be $4/6 \approx 67\%$. Therefore, in homophilic networks, node $i$ underestimates the minority size by factor of $0.5$ and in heterophilic network it overestimates the minority size by the factor $2$. On a group-level the majority group perceives the size of the minority as $4/18 \approx 20\%$ in the homophilic network, and $10/18 \approx 56\%$ for the heterophilic network. Therefore, the majority group underestimates the size of the minority group $0.67$ times in the homophilic network and overestimates the minority size $1.67$ times. In other words, depending on the topological structure of the network, perception of individuals and groups about their own and other groups’ sizes is distorted.
Fig. 2. Estimation of minority group size from the perspective of (a) the minority group and (b) the majority group as a function of homophily $h$ and the fraction of the minority $f_M$. Different colors refer to networks with different minority fractions ($f_M$). The analytic results are displayed as solid lines with the same color of $f_M$ in numerical results (circle dots). In the heterophilic networks ($0 \leq h \leq 0.5$), the minority (a) underestimates their own size, and the majority (b) overestimates the size of the minority due to the high connectivity between the majority to the minority (majority illusion). In homophilic networks ($0.5 \leq h \leq 1$), the minority overestimates their own size (filter bubble) and the majority underestimates the size of the minority. The insets show the same information on log scale to make the amount of underestimation and overestimation comparable. As group sizes become more disproportionate, perception bias increases.
| Data   | Nodes | Minority          | Majority          | Symmetric h | Asymmetric h (minority, majority) |
|--------|-------|-------------------|-------------------|-------------|----------------------------------|
| Brazil | 16,730| sex-workers 6,624 (40%) | sex-buyers 10,106 | 0           | 0, 0                             |
| POK    | 29,341| minority 12,868 (44%) | majority 16,473   | 0.15        | 0.2, 0.17                       |
| USF51  | 6,253 | male 2626(42%)     | female 3,627      | 0.47        | 0.48, 0.47                      |
| DBLP   | 280,200| female 63,356(22%) | male 216,844      | 0.56        | 0.57, 0.57                      |
| Github | 127,668| female 7,330 (6%)  | male 120,338      | 0.6         | 0.69, 0.54                      |
| APS    | 1,853 | CSM 695(37%)       | QSM 1.158         | 0.74        | 0.88, 1.0                       |

Table 1. Characteristics of empirical networks. The table shows six real-world networks with two groups. Group sizes, symmetric homophily and asymmetric homophily are reported for each network.
Fig. 3. Group-level perception biases in six empirical social networks. The figure shows how accurately the (a) minority group and (b) majority group estimate the size of the minority in real social networks with different levels of homophily. The symmetric homophily values of the empirical social networks are depicted on the x-axis. Homophily is measured between genders (female and male) except for the APS data where homophily is measured between different academic fields such as classical statistical mechanics (CSM) and quantum statistical mechanics (QSM). Empirical results are compared with our generative network model of homophily with assumption of symmetric homophily (triangle) and asymmetric homophily (square). We use the same scale of homophily on the x-axis for symmetric and asymmetric homophily to make the comparison easy. The perception bias increases as heterophily or homophily increases in networks. The results of the generative network model with asymmetric homophily are in excellent agreement with the empirical results highlighting the importance of considering the asymmetric homophily. The inset shows the perception gap (differences between the $P_{\text{group}}$ of minority and majority) for each network. The perception gap is higher for more extreme heterophilic or homophilic networks. Our synthetic networks are generated with $N = 2000$ nodes and averaged over 20 simulations and the mean and standard deviation are shown.
Fig. 4. Individual-level perception bias ($P_{indv}$) of the minority size as a function of degree ($k$). Each panel represents one empirical network. Top row consists of networks with heterophilic property (a) Brazilian sexual contact network, (b) Online dating network PussOKram (POK), (c) University network of Facebook (USF51) and bottom row consists of networks with homophilic property (d) DBLP scientific collaboration, (e) Github collaboration network, and (f) APS co-authorship network. Perception bias is consistent in all ranges of degree and we don’t find any clear pattern in empirical networks. This suggests that the social perception of nodes is not systematically distorted or enhanced by their degree. The results of the simulation (dashed lines) based on our generative network model with asymmetric homophily predict the pattern observed in the empirical data very well. In heterophilic networks (top panel) the majority there is an overestimation of perception for the majority and underestimation for the minority. The pattern is reversed in homophilic networks (bottom panel). In addition, the gap of perception is higher in more extreme heterophilic or extreme homophilic regions. The perception gap decreases as the level of heterophily or homophily decreases.
Fig. 5. Individual-level perception biases of minority size from the perspective of nodes (purple lines) compared to the weighted average of ego and her 1-hop neighborhood (green lines). The insets show the same results in log-scale for better comparison. The purple lines are calculated from Eq. 1 and averaged over all nodes in the group. The green lines show the weighted average of the ego and the perception of its direct neighborhood (1-hop). The results show that the estimation of minority size in both groups becomes more accurate in heterophilic networks when nodes take the estimates of their direct peers into account (green dots are closer to the gray dashed line compared to purple dots). In homophilic networks nodes do not benefit from aggregating their own estimate with their direct neighborhood due to the similarity and redundancy of the information. Minority fraction is fixed for 0.2 and simulations are done on networks with 2000 nodes and averaged over 50 runs.
Fig. S1. Degree growth rate of the minority \( C \) as a function of the homophily parameter \( h \) for various values of the minority fraction \( f_M \). Here \( f_M \) ranges from 0.1 to 0.5.

\( C \) denotes the degree growth rate of the minority while the degree growth rate of majority is proportional to \( 2 - C \). In this model of network with preferential attachment, homophily, and unequal group sizes, \( C \) can be expressed as a polynomial function \(^{32}\):

\[
\begin{align*}
(h - h')(h' - h)C^3 &+ ((2h - (1 - f_M)h')(h - h') + (2h' - f_M(2h - h'))(h' - h))C^2 \\
+ 2h(2h' - f'(2h - h')) - 2f_Mh'(h' - h) - 2(1 - f_M)(h')^2C \\
- 4f_Mh'h & = 0
\end{align*}
\]

(12)

As shown in Fig. S1, the growth rate of minority is high in the heterophilic regime and it decreases with increasing homophily \( h \). The lowest growth rate is observed when homophily is between 0.7 to 0.9. The smaller the minority fraction \( f_M \), the more the growth rate decreases with increasing homophily.
1.3. Correlation between properties

**Fig. S2.** Degree-degree ($r_{kk}$), degree-attribute ($r_{kz}$), and attribute-attribute ($r_{xx}$) correlations in the generative network model. The correlations are calculated based on the Pearson Correlation in the networks as a function of homophily $h$ and minority size $f_M$. The network size is 2000, and averaged over 20 simulations.

The degree-degree ($r_{kk}$) and group attribute-group attribute ($r_{xx}$) correlation are measured using Newman’s assortativity method that is based on the Pearson correlation coefficient.$^{33}$

Figure S2 shows the correlation between degree-degree, degree-attribute and attribute-attribute in our homophilc network model. Figure S2(a) reveals a strong negative correlation between degrees in heterophilic region. This suggests that large degree nodes (so-called hubs) tend to links to low degree nodes. In homophilic networks, we observe a weak negative correlation representing the connection between nodes with relatively similar degree. As the minority size decreases, the negative correlation increases in the heterophilic region.

The degree-attribute correlation analysis (cf. fig. S2(b)) shows that in the heterophilic region, the correlation between degree and attribute is positive. This indicates that the minority has relatively large degree in that region. However, the correlation between degree and attribute become negative in the homophilic region since the minority has a smaller degree than the majority.

Figure S2(c) displays the attribute-attribute correlations. Except for very small minority size, this correlation has a linear relation to the homophily parameter in the model such that $r_{xx} = 2h - 1$. Consequently, we can translate the symmetric homophily in our model to Newman’s assortativity measure.
2. Characteristics of empirical networks

2.1. The basic network properties of empirical networks

| Data   | Number of links | Avg. degree | Avg. clustering | $r_{kk}$ | $r_{kx}$ | $r_{xx}$ |
|--------|-----------------|-------------|-----------------|----------|----------|----------|
| Brazil | 39,044          | 4.67        | 0.0             | −0.11    | 0.1      | −1.0     |
| POK    | 115,684         | 7.88        | 0.05            | −0.05    | 0.03     | −0.68    |
| USF51  | 126,480         | 40.45       | 0.24            | 0.17     | 0.02     | −0.06    |
| DBLP   | 750,601         | 5.36        | 0.68            | 0.96     | −0.03    | 0.1      |
| Github | 186,360         | 2.92        | 0.12            | 0.08     | −0.01    | 0.068    |
| APS    | 3,627           | 3.91        | 0.35            | 0.15     | −0.09    | 0.87     |

Table 2. Network characteristics of empirical social networks. The number of links, average degree (Avg. degree), average clustering (Avg. clustering), the degree-degree correlation ($r_{kk}$), the degree-attribute correlation ($r_{kx}$), and the attribute-attribute correlation ($r_{xx}$) are represented.

Table 2 shows network characteristics of six real-world networks. The number of links is the number of existing edges between nodes in a network. Average degree is defined as the total sum of degree for nodes in a network divided by the total number of nodes. Clustering is defined as the fraction of existing triangles out of all possible triangles in a network. Degree-degree correlation $r_{kk}$, degree-attribute correlation $r_{kx}$ and attribute-attribute correlation $r_{xx}$ are calculated based on the Pearson correlation suggested by Newman\(^3\). The heterophilic networks (Brazil, POK, USF51) have negative attribute-attribute correlation. The homophilic networks (DBLP, Github, APS) show positive attribute correlation. DBLP network has the highest average clustering, and USF51 has the largest average degree.
2.2. Degree distributions of empirical networks

**Fig. S3. Degree distributions of six empirical networks.** All networks show heavy-tailed degree distribution and in some cases with a natural cut-off. The parameter $\alpha$ shows the exponent of the degree distribution from power-law fitting.

Figure S3 shows the heavy-tailed degree distributions of the six empirical networks. We use the power-law package in networkx\textsuperscript{49} to fit the distribution and report the exponent of the power-law.


Fig. S4. Comparison of group-level perception bias with a model including clustering. The group-level perception bias with clustering is compared with the empirical perception bias \( P_{\text{group}} \) in six empirical networks.

To investigate whether clustering plays a role in determining the group perception, we modify the synthetic homophilic network model to account for higher clustering similar to Holme-Kim algorithm\(^{34}\). In the modified model, the connectivity follows the same principle as homophilic network model in addition to closing triads in a stochastic fashion.

Figure S4 shows the group-level perception in the empirical networks compared with the perception biases that our model produced when using symmetric homophily, asymmetric homophily, and symmetric homophily with clustering. Adding more realistic clustering to the synthetic model does not improve the predictive power of the model. However, when correcting for the asymmetric homophily among groups, the model fits the empirical observations very well.
3. Perception bias in synthetic networks

3.1. Individual-level perception bias as a function of homophily and the minority’s size

Figure S5. Averaged individual-level perception bias as a function of the degree ($k$) and minority size ($f_M$).

First row represents the individual-level perception bias for the minority group, while the second row shows the same information for the majority. Each column corresponds to a fixed minority group size. First column is for $f_M = 0.1$, second one is for $f_M = 0.3$ and last column is for $f_M = 0.5$. The last row shows the difference between the perception biases of minority and majority.

Figure S5 depicts the individual-level perception bias in synthetic network model as a function of degree of the nodes. One can see that the individual perception biases decreases with increasing minority faction $f_M$.

In the homophilic region, the overall trend for the perception of the minority shows that their perception becomes worse as their degree increases. In a highly homophilic network, popular minority nodes can mainly attract other minority nodes, since they cannot compete with popular majority nodes that accumulate much higher degree. For the majority in a homophilic network (e.g. red line in Fig. S5(d)), the perception improves as degree increases. This is because due to a larger size of the majority, the most popular majority nodes occasionally attract minority nodes. As the group sizes become equal, the dependency between degree and perception vanishes and the perception becomes only dependent on the homophily and group size $P_{indv} = \frac{1}{2} h$.

In the heterophilic region, for minority, the overall trend shows that as degree increases, the perception of the minority
becomes worse. This is due to their smaller size and the nature of heterophilic connection, minority’s degree can grow large and attract more and more majority and thus the underestimation of their own group size becomes worse. Conversely, for the majority, perception bias improves as degree increases. This is because despite the fact that they are more attracted to the minority since the minority size is limited, higher degree majority have higher chance to attract other majority nodes occasionally and thereby their perception estimation improves. Again, as the group sizes become equal, the dependency between degree and perception vanishes and the perception become only dependent on the homophily and group size $P_{indv.} = \frac{1}{j_M}$. 