nature human behaviour

SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41562-019-0677-4

In the format provided by the authors and unedited.

Homophily and minority-group size explain perception biases in social networks

Eun Lee [1,11*], Fariba Karimi [2,3,11*], Claudia Wagner [2,3], Hang-Hyun Jo [4,5,6], Markus Strohmaier [2,7] and Mirta Galesic [8,9,10]

[1]Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2]Department of Computational Social Science, GESIS, Cologne, Germany. [3]Institute for Web Science and Technologies, University of Koblenz-Landau, Koblenz, Germany. [4]Asia Pacific Center for Theoretical Physics, Pohang, Republic of Korea. [5]Department of Physics, Pohang University of Science and Technology, Pohang, Republic of Korea. [6]Department of Computer Science, Aalto University, Espoo, Finland. [7]Department for Society, Technology and Human Factors & Department of Computer Science, RWTH Aachen University, Aachen, Germany. [8]Santa Fe Institute, Santa Fe, NM, USA. [9]Complexity Science Hub Vienna, Vienna, Austria. [10]Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany. [11]These authors contributed equally: Eun Lee, Fariba Karimi.
*e-mail: eunfeel@email.unc.edu; fariba.karimi@gesis.org

Supplementary Information

Supplementary Tables

Supplementary Table 1. Survey questions about different attributes. Texts of questions asking about respondents' own attributes in our surveys in Germany, South Korea, and the United States, and the national surveys that were the original sources of questions and true population frequencies (see Supplementary Table 2). In addition to questions about their own attributes listed in the table, participants answered questions about the frequency of each attribute in their personal networks and in the general population of their country: "When asked… [question text here] … what percentage of [your social contacts/of adults living in [country]] would answer [for questions 6 and 10, answer '(b)'; for all other questions, answer '(a)']?"

Characteristic	Question text	Source of question text and true population frequencies		
		Germany	S. Korea	U.S.
1. Not having money for food	Have there been times in the past 12 months when you did not have enough moneyto buy food that you or your family needed? (a)Yes–(b)No	Gallup World Poll (2017)	Gallup World Poll (2017)	Gallup World Poll (2017)
2. Donating to charity	In the past month, have you donated money to a charity? (a)Yes–(b)No	Gallup World Poll (2018)	Gallup World Poll (2017)	Gallup World Poll (2017)
3. Experiencing theft	Within the past 12 months,have you had money or property stolen from you or another household member? (a)Yes–(b)No	Gallup World Poll (2017)	–	Gallup World Poll (2017)
4. Belief in a god	Do you believe in god or a supreme being? (a)Yes–(b)No	Ipsos (2010)	Ipsos (2010)	Ipsos (2010)
5. God and morality	Which one of these comes closer to your opinion? (a) It is not necessary to believe in Godin order to be moral and have good values.– (b) It is necessary	Pew (2011)	–	Pew (2011)

	to believe in God in order to be moral and have good values.			
6. Worship attendance	Have you attended a place of worship or a religious service within the past 7 days? (a)Yes–(b)No	Pew (2010)	Pew (2013)	Pew (2012)
7. Religion importance	Is religion an important part of your daily life? (a)Yes–(b)No	Gallup World Poll (2015)	Gallup World Poll (2015)	Gallup World Poll (2015)
8. Smoking	These days, are you smoking any tobacco product at least once a day?(Tobacco smoking includes cigarettes, cigars, pipes, and any other form of smoked tobacco). (a)Yes–(b)No	World Health Organization (2010)	National Nutrition Survey (2016)	World Health Organization (2010)
9. Military force	Do you agree that it is sometimes necessary to use military force to maintain orderin the world? (a)Yes–(b)No	Pew (2011)	–	Pew (2011)
10. Homosexuality acceptance	Which one of these comes closer to your opinion? (a) Homosexuality is a way of life that should be accepted by society. –(b) Homosexuality is a way of life that should not be accepted by society.	Pew (2013)	Pew (2013)	Pew (2013)

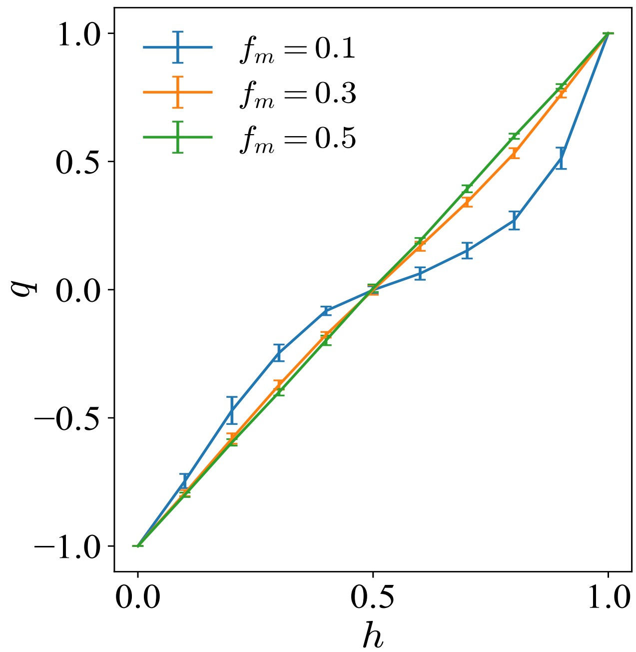
Supplementary Table 2. True population frequencies for each attribute. Minority- and majority-group sizes for each attribute in the general populations of Germany, South Korea and the United States, obtained from national surveys listed in Supplementary Table 1. Response options are the same as answers (a) and (b) in Supplementary Table 1 except for Belief in a god, for which the national survey offered several options: definitely believe in god(s) or supreme being(s) [categorized as (a) in our analyses] - sometimes I believe but sometimes I don't - not sure - don't believe [categorized together as (b)]. The percentages for Experiencing theft, God and morality, Religion importance, Military force, and Homosexuality acceptance do not sum to 100 in all countries because those surveys allowed participants to not answer all questions.

Attribute	Germany (%)		S. Korea (%)		U.S. (%)	
	(a)	(b)	(a)	(b)	(a)	(b)
1. Not having money for food	6	94	17	83	19	81
2. Donating to charity	44	56	40	60	61	39
3. Experiencing theft	10	89	–	–	14	86
4. Belief in a god	32	68	25	75	75	25
5. God and morality	66	33	–	–	46	53
6. Worship attendance	10	90	30	70	41	57
7. Religion importance	39	60	42	57	66	34
8. Smoking	22	78	24	76	15	85
9. Military force	50	50	–	–	75	23
10. Homosexuality acceptance	87	11	39	59	60	33

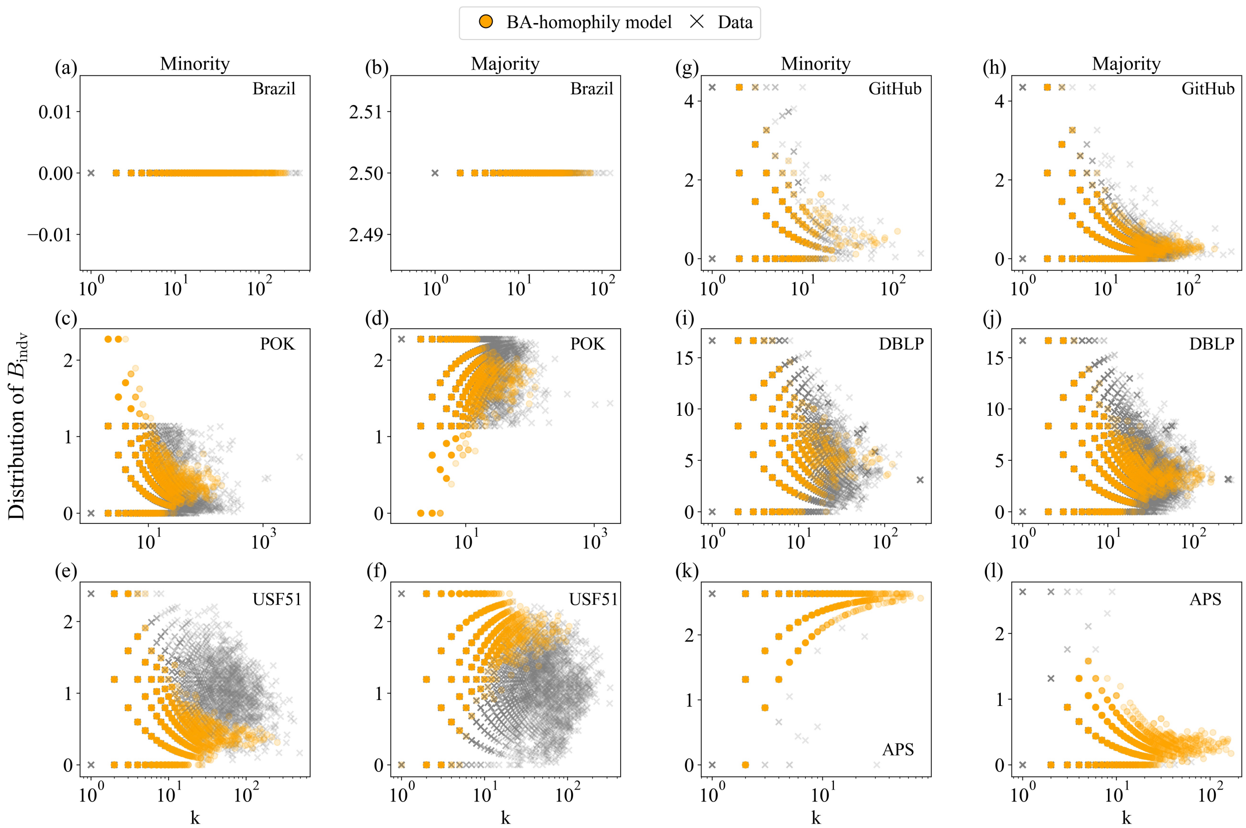
Supplementary Table 3. Additional details for Figure 4 (in the main text): Possible bias in empirical networks with standard deviations. First column shows an empirically observed possible bias in six empirical networks. Second column shows a predicted possible bias by BA-homophily model with symmetric homophily. Third column displays a predicted possible bias by BA-homophily model with asymmetric homophily. The possible biases are rounded to the second digit after the decimal point, and a real number in a parenthesis shows a standard deviation of a predicted value. For the prediction, we used the network size, N=2000, for 20 ensembles.

	Empirically measured possible bias		Possible bias with symmetric homophily		Possible bias with asymmetric homophily	
	Minority	Majority	Minority	Majority	Minority	Majority
Brazil	0.0	2.53	0.0($\pm$0.0)	2.5($\pm$0.0)	0.0($\pm$0.0)	2.5($\pm$0.0)
POK	0.28	2.02	0.33($\pm$0.02)	1.86($\pm$0.02)	0.42($\pm$0.02)	1.86($\pm$0.03)
USF51	1.01	1.03	0.95($\pm$0.06)	1.09($\pm$0.05)	0.96($\pm$0.07)	1.07($\pm$0.06)
GitHub	2.04	0.82	0.98($\pm$0.28)	0.87($\pm$0.12)	1.96($\pm$0.4)	0.85($\pm$0.14)
DBLP	1.28	0.81	1.14($\pm$0.11)	0.82($\pm$0.08)	1.23($\pm$0.1)	0.83($\pm$0.04)
APS	2.43	0.15	2.37($\pm$0.02)	0.15($\pm$0.01)	2.33($\pm$0.03)	0.07($\pm$0.01)

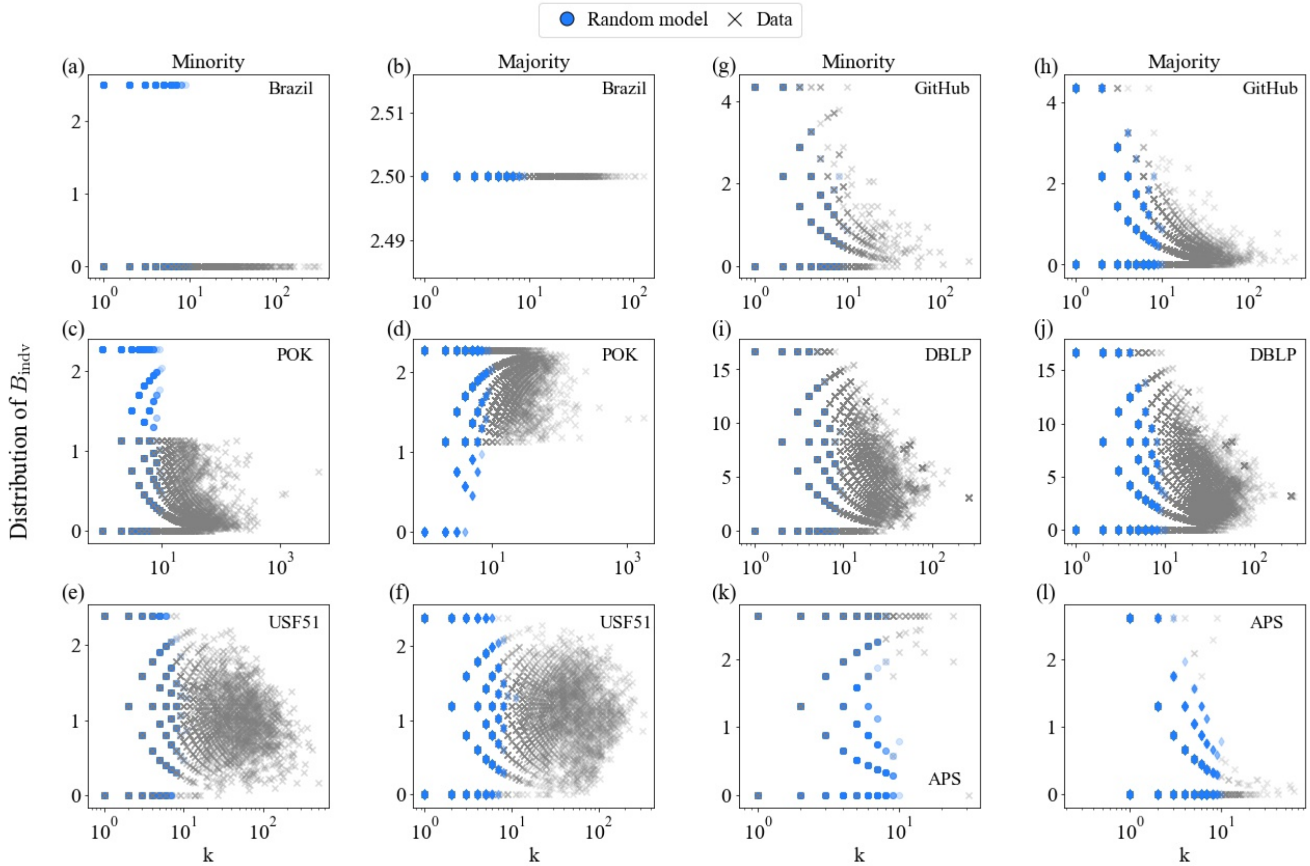
Supplementary Figures



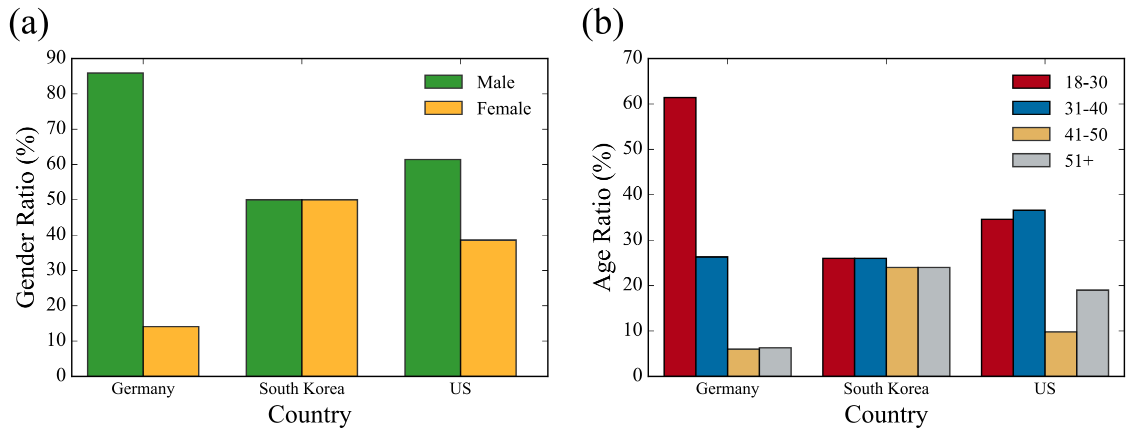
Supplementary Figure 1. Relationship between Newman's assortativity (q) and symmetric homophily (h) in the BA-homophily model, for different sizes of minority group (f_m). The symmetric homophily is equivalent to Newman's assortativity measure (q)[1]. The latter corresponds directly to the homophily parameter in the BA-homophily model after adjusting for scale in a relation $q = 2h - 1$. In the BA-homophily model, $h = 0$ means complete heterophily ($q = -1$), $h = 0.5$ indicates no relationship between network structure and attributes ($q = 0$), and $h = 1$ indicates complete homophily ($q = 1$). The relationship deviates slightly from the linear function for small minority-group sizes but is independent of minority sizes when $h \approx 0.5$.



Supplementary Figure 2. Distribution of possible individual-level perception biases (B_{indv}) as a function of individual's degree (k). Each row represents one empirical network. The left two columns show the results for minority and majority members of the heterophilic empirical networks [Brazilian sexual contact network (Brazil), Swedish online dating network (POK), Facebook network of an university (USF51)] and the right two columns for the minority and majority members of the homophilic empirical networks [GitHub developers' network (GitHub), DBLP developers' network (DBLP), American Physical Society network (APS)]. The gray crosses represent the possible perception bias of each individual estimated from the empirical network, and orange circles show the perception bias predicted by the BA-homophily model. The simulation results are aggregated over 50 iterations and the network size is $N = 2,000$. The x axis is shown in log scale.



Supplementary Figure 3. Distribution of possible individual-level perception biases ($B_{\mathbf{indv}}$) as a function of individual's degree (k) compared with random model. Each row represents one empirical network. The left two columns show the results for minority and majority members of the heterophilic empirical networks [Brazilian sexual contact network (Brazil), Swedish online dating network (POK), Facebook network of an university (USF51)] and the right two columns for the minority and majority members of the homophilic empirical networks [GitHub developers' network (GitHub), DBLP developers' network (DBLP), American Physical Society network (APS)]. The gray crosses represent the possible perception bias of each individual estimated from the empirical network, and green diamonds show the perception bias predicted by a random model. Here, a random model is a modified BA-homophily model without preferential attachment. To construct a random network with homophily (or heterophily), we first select two random nodes. Then, we check the group identity of the nodes and connect them with the corresponding homophily (or heterophily) between their group memberships. We repeat the process until all nodes do not have any links to connect. The simulation results are aggregated over 50 iterations and the network size is $N = 2,000$. The *x* axis is shown in log scale.



Supplementary Figure 4. Gender and age distribution of a survey study. Panel (a) shows the ratio of gender in each country: 85.9% of the German participants, 50.0% of the Korean participants, and 61.4% of the U.S. participants were male. Panel (b) displays the age distribution. The age distribution in Germany was 18–30 years: 61.4%, 31–40 years: 26.3%, 41–50 years: 6.0%, 50+ years: 6.3%; in South Korea it was 18–30 years: 26.0%, 31–40 years: 26.0%, 41–50 years: 24.0%, 50+ years: 24.0%; and in the United States it was 18–30 years: 34.6%, 31–40 years: 36.6%, 41–50 years: 9.8%, 50+ years: 19.0%.

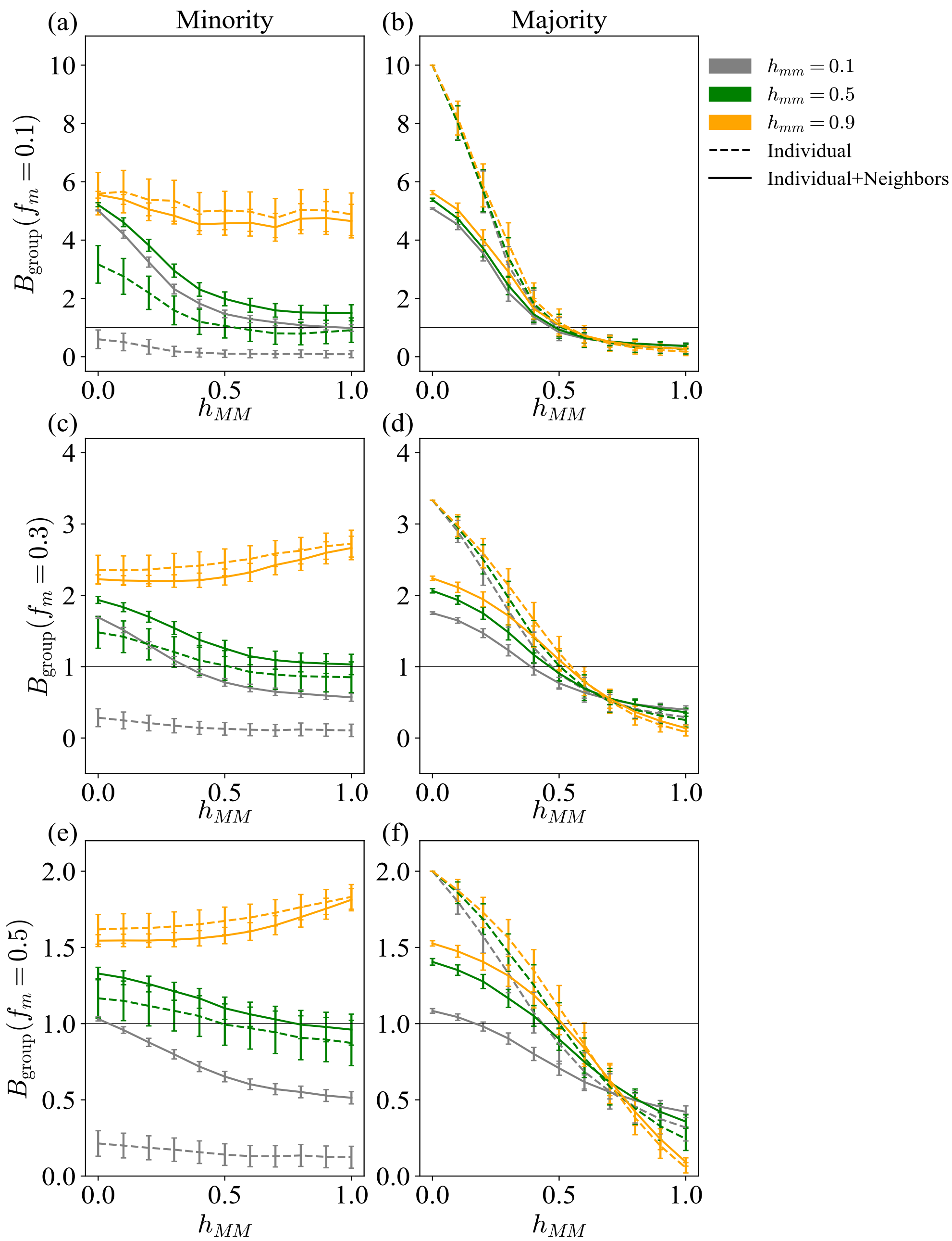
Supplementary Result 1

Group-level perception bias after aggregation of individual with neighbors' perceptions, assuming asymmetric homophily

Here, we investigate to what extent and under what structural conditions individuals can reduce their perception bias for the prevalence of an attribute in a population by asking their friends about their perceptions and integrating those perceptions with their own when homophily is asymmetric. We build on DeGroot's weighted belief formalization by aggregating an individual's perception (ego) with the averaged perceptions of the individual's direct neighbors[2]. In this formalization, individual i with a perception bias of $B_{\text{indv},i}$ asks each neighbor j about her perception bias of that attribute frequency $B_{\text{indv},j}$. The final perception bias of individual i is the weighted average of her and her neighbors' bias: $\hat{B}_{\text{indv},i} = \frac{B_{\text{indv},i} + (\sum_{j \in \Lambda_i} B_{\text{indv},j})/k_i}{2}$, where Λ_i is the set of i's neighbors and $k_i = |\Lambda_i|$.

In the main text we assumed symmetric homophily, while here we assume asymmetric homophily in the BA-homophily model. Supplementary Figure 5 shows a comparison of the group perception (B_{group}) of individuals who are in (a) the minority and (b) the majority, with the bias of their perceptions aggregated with those of their neighbors. The minority-group size is fixed to 0.1 in the top row, 0.3 in the middle row, and 0.5 in the bottom row. The homophily for the minority group h_{mm} is fixed to 0.1,0.5,0.9 (depicted by lines of different colors), while homophily for the majority group h_{MM} ranges from 0 to 1 along the horizontal axis.

Results for asymmetric homophily (Supplementary Figure 5) are generally similar to those for symmetric homophily (Fig. 5 in main text): accounting for the opinion of neighbors can decrease perception bias when networks are heterogeneous. For the majority, aggregating their own perceptions with those of their neighbors leads to a robust improvement in perception accuracy, as shown in Supplementary Figure 5b, d, and f. For the minority, accounting for 1-hop neighbors also helps decrease the bias, though less than for the majority. However, when minority group is small and homophily is highly asymmetric ($h_{mm} = 0.5$ and $h_{MM} < 0.5$), accounting for neighbors can indeed increase the bias see Supplementary Figure 5a).



Supplementary Figure 5. Estimates of possible group-level perception biases (dashed lines) compared to possible biases of the weighted average of perceptions of individuals and their neighbors (solid lines), assuming asymmetric homophily. The minority-group size is fixed at 0.1 in the top row, 0.3 in the middle row, and 0.5 in the bottom row. The homophily for the minority h_{mm}(depicted by lines of different colors) is fixed, while homophily for the majority h_{MM} ranges from 0 to 1 along the horizontal axis.

Supplementary Method 1

Growth rate (C) in BA-homophily model

On the basis of the derivation provided by Karimi et al.[3], we can derive the exact degree dynamics of the BA-homophily model. Let us assume $K_m(t)$ and $K_M(t)$ as the sum of the degrees of each group m and M, respectively. With the number of links a node has, the total number of links at each time step will be $K(t) = K_m(t) + K_M(t) = 2\lambda t$. One can also describe the evolution of each group's degree as

$$\begin{cases} K_m(t+\Delta t) = K_m(t) + \lambda(f_m(1 + \frac{h_{mm}K_m(t)}{h_{mm}K_m(t)+h_{mM}K_M(t)}) + f_M \frac{h_{Mm}K_m(t)}{h_{MM}K_M(t)+h_{Mm}K_m(t)})\Delta t, \\ K_M(t+\Delta t) = K_M(t) + \lambda(f_M(1 + \frac{h_{MM}K_M(t)}{h_{MM}K_M(t)+h_{Mm}K_m(t)}) + f_m \frac{h_{mM}K_M(t)}{h_{mm}K_m(t)+h_{mM}K_M(t)})\Delta t. \end{cases}$$

(Supplementary Equation 1)

Here, one can separate the amount of increase of the degree for each group with the limit $\Delta t \to 0$,

$$\begin{cases} \frac{dK_m}{dt} = \lambda(f_m(1 + \frac{h_{mm}K_m(t)}{h_{mm}K_m(t)+h_{mM}K_M(t)}) + f_M \frac{h_{Mm}K_m(t)}{h_{MM}K_M(t)+h_{Mm}K_m(t)}), \\ \frac{dK_M}{dt} = \lambda(f_M(1 + \frac{h_{MM}K_M(t)}{h_{MM}K_M(t)+h_{Mm}K_m(t)}) + f_m \frac{h_{mM}K_M(t)}{h_{mm}K_m(t)+h_{mM}K_M(t)}). \end{cases}$$

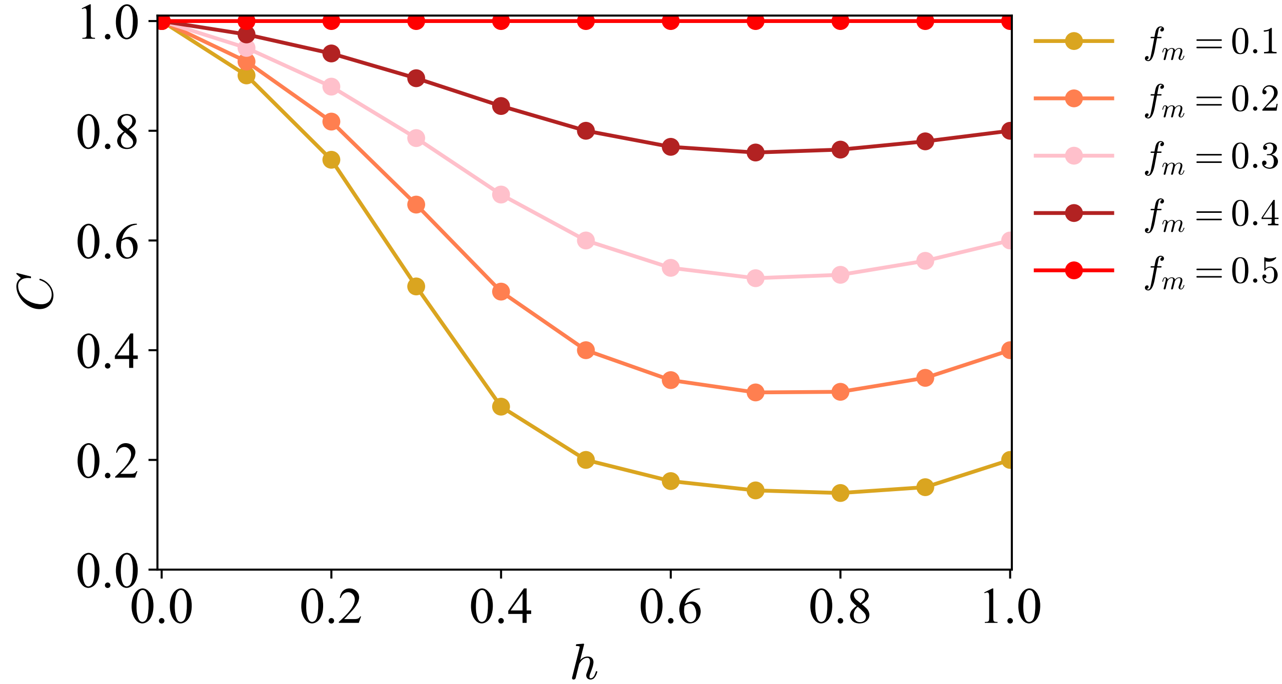
(Supplementary Equation 2)

We can assume that $K_m(t)$ $[K_M(t)]$ grows as a linear function of time. Given this assumption, we can write that $K_m(t) = C\lambda t$ $[K_M(t) = (2-C)\lambda t]$ based on $K(t) = 2\lambda t$.

$$\begin{cases} \frac{dK_m}{dt} = C\lambda = \lambda(f_m(1 + \frac{h_{mm}Ct}{h_{mm}Ct+h_{mM}(2-C)t}) + f_M \frac{h_{Mm}Ct}{h_{MM}(2-C)t+h_{Mm}Ct}), \\ \frac{dK_M}{dt} = (2-C)\lambda = \lambda(f_M(1 + \frac{h_{MM}(2-C)t}{h_{MM}(2-C)t+h_{Mm}Ct}) + f_m \frac{h_{mM}(2-C)t}{h_{mm}Ct+h_{mM}(2-C)t}). \end{cases}$$

(Supplementary Equation 3)

Then, we can derive the relation of C with group sizes and edge density in a group (Eq. 12) from Supplementary Equation 2 by using p_{mm} (Eq. 9) and p_{mM} (Eq. 10).



Supplementary Figure 6. The relation between the minority's degree growth rate ($\boldsymbol{C}$) and the symmetric homophily ($\boldsymbol{h}$). As the minority group's size f_m decreases, the degree growth rate of the minority drastically decreases with increasing symmetric homophily h ($h_{mm} = h_{MM}$). C is a function of h and f_m as described in Eq. 12 (in main text).

Supplementary References

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

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

3. Karimi, F., Génois, M., Wagner, C., Singer, P. & Strohmaier, M. Homophily influences ranking of minorities in social networks. *Sci. Rep.* **8**, (2018).