WHEN ERGMS LEAD TO BIASED SAMPLES: REPLY TO KRETSCHMER ET AL.[1]

In Smith et al. (2016; hereafter STMM), we found evidence for our theory that ethnic homophilous friendship choices relate differently to the ethnic composition of school classes for members of the ethnic majority and minority groups. Kretschmer, Gereke, Winter, and Zhang (2023; hereafter KGWZ) argue for a methodological correction of our work and show that our conclusion is driven by nonconverging exponential random graph models (ERGMs). They conclude that there is no evidence for differential effects of classroom ethnic composition on native and immigrant friendship segregation once they examine a much smaller subsample of classes in which the ERGMs do converge.

It has been seven years since the original STMM publication, so we have had time to reconsider how the research question can be best explored and answered. While we agree with KGWZ's methodological critique, we believe the continued use of ERGMs and fine-grained ethnic categories jettisons so much of the sample that it undermines the ability to adequately explore the research question and test the theory. Instead, we now prefer an alternative approach that preserves more of the sample. In this response to KGWZ, we give an overview of the original work and KGWZ's critique followed by a description of our extended, new study, where we analyze the relationship between ethnic homophily and the ethnic composition in the classroom with different measures of majority and minority groups and a different method (multilevel dyadic regression). Our findings show evidence for our initial theoretical expectations and empirical findings reported in STMM.

ORIGINAL WORK AND CRITIQUE

In our original work, STMM, we examined the relationship between the ethnic composition of classrooms and students' tendency to befriend peers of the same ethnic group, also known as ethnic homophily. We developed a theoretical framework that distinguished between (1) the ethnic diversity of

[1] We thank Alessandra Rister Portinari Maranca for inspecting the MCMC diagnostics plots for ERGM convergence. Direct correspondence to Sanne Smith, Graduate School of Education, 485 Lasuen Hall, Stanford University, Stanford, CA 94305. Email: sannesmith@stanford.edu

 Published by The University of Chicago Press. https://doi.org/10.1086/727858

a classroom and (2) which ethnic group constitutes the numerical majority in the classroom, and we looked separately at native and immigrant homophily. We argued that different concepts of classroom ethnic composition relate independently and differently to the ingroup friendship tendencies of students of the majority group and minority groups.

Using exponential random graph models (ERGMs) on 517 classes for natives' and 262 classes for immigrants' homophily, we conducted a meta-analysis of the resulting class-specific homophily coefficients and showed that immigrant students' tendency to have same-ethnic friends was strongest in moderately diverse classes (i.e., where most immigrant students have several same-ethnic class peers). Also, our analysis provided no evidence that immigrant homophily depends on how much native students befriend each other. We concluded from these findings that immigrant students do not so much experience feelings of ethnic threat from natives but are merely better able to satisfy same-ethnic preferences in classes where they find several same-ethnic peers.

Compared to immigrant homophily, native homophily was found to be relatively low and weakly associated with the total degree of ethnic diversity in the class. Instead, native homophily increased considerably when the out-group was unified; that is, when immigrants befriended each other more and when immigrant diversity was low. As such, the findings suggested that native homophily is triggered by feelings of ethnic threat.

KGWZ's main critique of our work lies in identifying estimation problems in the ERGMs. Their correction of these problems was triggered by two major differences in their analytical strategy. First, KGWZ were able to intercept nonconverging models because the most recent ERGM package (ver. 3.10-4) does not return coefficients and standard errors for nonestimable parameters. This is a clear indication that the models did not converge. We did not catch these nonconverging models because the ERGM package used at the time (ver. 3.1-0) returned reasonably sized but arbitrary coefficients with standard errors close to zero.[2] Second, KGWZ show that we could have identified these nonconverging models by visually checking the MCMC trace plots. Instead of manually inspecting the MCMC trace plots, however, we assessed model convergence by various rule-based methods.[3]

Once KGWZ exclude all classes in which the ERGM model does not converge, they find in line with STMM that native homophily is strongest in the most diverse classes, but they find no evidence that immigrant

[2] SEs were within the -5 and 5 range. Coefficients were excluded from the original analysis if they fell outside that range.

[3] We assessed model convergence by evaluating the change in the log likelihood between the last iterations of the fitting algorithm and examining the correlation between coefficients with varying numbers of iterations, lengths of the MCMC burn-in, and MCMC sample sizes.

homophily depends on the classroom's ethnic composition. Their conclusion is therefore that (1) researchers should always inspect the MCMC goodness-of-fit plots and (2) there is no evidence for differential effects of classroom ethnic composition on native and immigrant friendship segregation when examining the research question with the ERGM approach.

OPPORTUNITIES FOR PROGRESS

We would like to thank KGWZ for their thorough replication and methodological correction of our work. At the time of our publication, we sought a principled means of assessing fit and convergence for ERGMs as we needed to evaluate hundreds of ERGM model results. It was with reproducibility in mind that we decided not to rely on manually inspecting the goodness-of-fit plots and instead focused on quantified model fit information. KGWZ show that our method did not suffice, and we agree with KGWZ that MCMC trace plots should be inspected until a more reliable and efficient method of model convergence is developed for multi-ERGM efforts.

The main challenge shown in both KGWZ and STMM is that ERGMs are notoriously difficult to estimate. When models do not converge, there are three options: estimate the model for a subset of the data where the model converges (dropping most of the sample), leave out (theoretically relevant) predictors, or use a different model (Martin 2018). KGWZ criticized the evidence based on ERGMs including nonestimable coefficients and opted for the first option when presented with nonconvergence. Their approach, however, is subject to other statistical and interpretative issues when it comes to answering the research question. Our theory posited that students from the majority group respond differently to ethnic threat compared to students from the minority group, and because of strict inclusion criteria, KGWZ test this theory for a limited number of classes only: 36% of the classes for native homophily and 17% of the classes for immigrant homophily, respectively.

To some extent, limiting the sample is necessary. We also excluded more than half of the classes for these reasons in STMM. Several classes are not suitable to examine because their sociometric data are unreliable (low student response, too many invalid nominations) or because the school class lacks diversity (it is impossible to study if a student prefers same-ethnic friends when they have no same-ethnic peers to choose from in the first place). Other methods than ERGMs would require similar selections. The use of ERGMs and the associated convergence issues, however, requires an even further restriction of the data. For example, figure 1 shows boxplots of the classes that are included and excluded in KGWZ's analysis. The figure shows that classes included in KGWZ's analysis for native and immigrant homophily are more ethnically diverse and have lower proportions of native students. Most differences are significant and especially pronounced in the analysis of immigrant

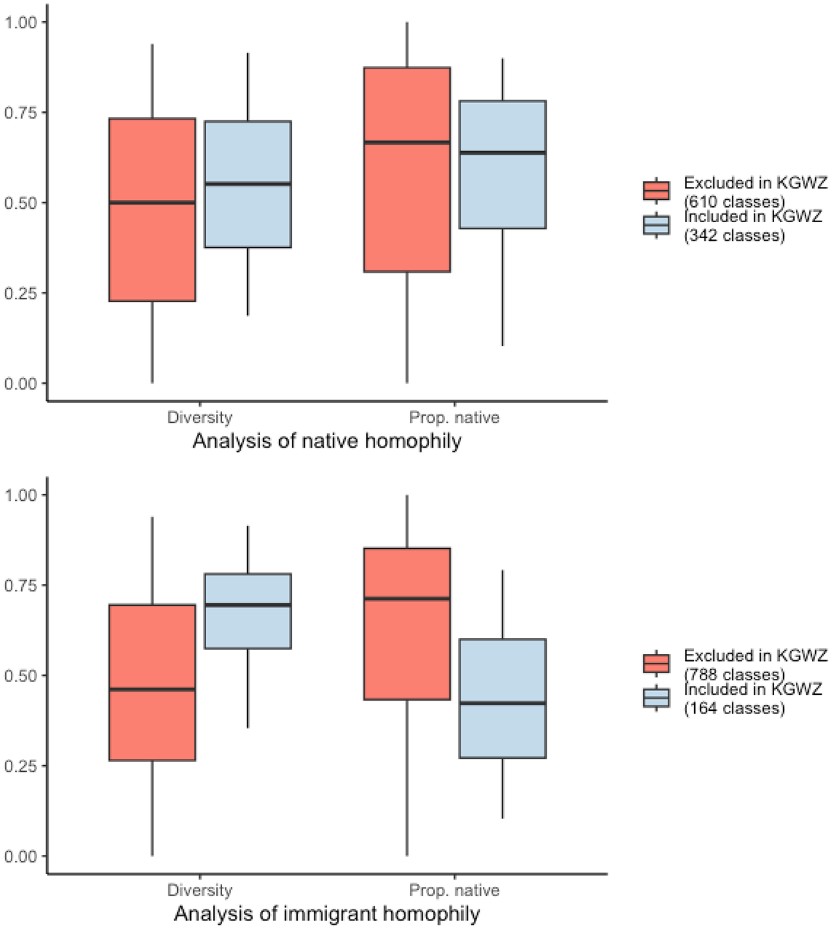

FIG. 1.—Included and excluded classes in KGWZs analysis for estimating native (top) and immigrant homophily (bottom) and their distributions on the diversity and proportion natives variables

homophily.[4] Excluding classes in which ERGMs do not converge leads to a restricted range issue: the effects of the classroom composition on ethnic homophily are only estimated on a narrow range of the predictor variables. Having such a selective sample is problematic for statistical inference to the population of interest, as the correlation between classroom composition and ethnic homophily remains unknown in the less diverse classrooms.

[4] For native homophily, diversity: $t = -4.58$, df $= 893.1$, $P < .001$. For native homophily, proportion of natives: $t = -.45$, df $= 921.47$, $P = .656$. For immigrant homophily, diversity: $t = -15.02$, df $= 427.95$, $P < .001$. For native homophily, proportion of natives: $t = 10.30$, df $= 352.88$, $P < .001$.

Prior research has shown that range restrictions underestimate the correlation between variables of interest (Hunter and Schmidt 1990; Henriksson and Wolming 1998), and this issue may explain why KGWZ do not find a significant relationship between the classroom composition variables and immigrant homophily. There are plenty of school classes in the population of interest where students are exposed to low and more moderate levels of ethnic diversity, and many of these classes are not represented in the ERGM analyses. In addition, we noticed that the sample that remains after strict convergence criteria is not only significantly more ethnically diverse, but also significantly lower SES and includes a relatively high proportion of classrooms in Germany and Sweden and far lower proportions of Dutch and English classes (see online app. A).

All in all, the issue is that ERGMs cannot be applied to most of the data, which begs the question of whether ERGMs are suitable models to explore the research question. KGWZ also acknowledge this in their discussion section by stating that they do “not claim to offer the last word on these substantive questions. As STMM point out, there are good theoretical reasons to expect threat and competition mechanisms to be particularly salient for the majority group, while playing a smaller role in shaping the friendship choices of ethnic minorities” (p. X). Rather than using only the data that fit the ERGM model, we no longer believe that ERGMs are the right approach for this research question and data set. Instead, we think that the data should be analyzed with a method that is more compatible with the research question and characteristics of the data set (Martin 2018). In other words, we realize that choosing the most sophisticated model over the retention of a more representative sample poses a serious threat to validity.

In the next section, we detail a different approach where we propose a redefinition of ethnic same-ethnic friendship and a reconsideration of ERGMs. This approach allows us to retain more of the sample and provides more generalizable results.

A DIFFERENT APPROACH

A Redefinition of Same-Ethnic Friendship

The main issue of nonconvergence in the ERGM models is that immigrant homophily as measured by national origin groups cannot be meaningfully estimated in many networks containing insufficient numbers of immigrant students from the same ethnic group. The analysis in STMM (and the replication of KGWZ) distinguished 158 national origin groups. We chose this approach because research at the time showed the importance of recognizing smaller ethnic groups within larger racial groups (Wimmer and Lewis 2010). The disadvantage, however, of distinguishing 158 groups is that students with an immigrant background often do not have a sufficient number of peers with the same national origin country in class. This leads to nonconvergence in ERGMs and drops more than half of the classes in the sample. As such, we think there is value in collapsing detailed national origins into broader groups of ethnicity.[5]

There are also substantive reasons to deviate from the detailed national origin measure in addition to the technical concerns. Several studies have shown that the social identities of adolescents are more complex than national origin. For example, studies using the same data set show that friendship boundaries are delineated by religious and cultural characteristics (Simsek, Van Tubergen, and Fleischmann 2022; Smith, Maas, and Van Tubergen 2014). Research in the United States shows consistent evidence of racial homophily (Moody 2001) net of more fine-grained measures of ethnic homophily (Wimmer and Lewis 2010). As such, it is unlikely that European students only adhere to national origin groups when finding friends. For example, take a school class that includes two students with an immigrant background, but in which the two students have different national origins (e.g., a Turkish and a Moroccan background). While they do not share the same national origin, they are likely to share a religious identity (i.e., they are both Muslim) and might seek out each other's friendship as such. While students might be crossing national origin boundaries by befriending each other, they are still selecting into social groups that are homogenous in terms of group identification based on one or more cultural expressions and traits such as geographic location, customs, history, language, and religion.

As such, our approach speaks to the issue of dropping many classes and ending up with a selective sample of school classes in which ethnic homophily can be estimated in ERGMs, based on the assumption that immigrants from different national origins are not alike and that only national origin matters in friendship selection. We instead want to explore a measure of ingroup friendship that boosts statistical power and is possibly more closely aligned with (European) student experiences of ingroup friendships. We construct three measures of ethnicity with varying degrees of detail.

National origin groups ($N = 158$).—We start with the most detailed measure of national origin groups. The complete data set consists of 158 national origin groups. This measure is the same as in STMM and KGWZ.

Cluster groups ($N = 13$).—Our second measure of ethnicity is based on the work of Ronen and Shenkar (2013), who have provided an overview of

[5] In STMM, we examined whether our results were driven by the detailed measurement of ethnicity by national origin. We repeated the analysis by collapsing the 158 national groups into larger categories. We distinguished between natives and the most important four immigrant groups within the four countries. All other students with an immigrant background were coded to the continent their parental birth countries belong to (see STMM, appendix). We then reanalyzed the data with ERGMs without the manual check of the goodness-of-fit plots, so the results of that sensitivity analyses are likely biased as well.

clusters of countries that are culturally similar in terms of geography, religion, language, and socioeconomic characteristics. Their work synthesizes data from 10 empirical studies categorizing countries into broader cultures. Using the cluster solutions from these prior 10 studies, Ronen and Shenkar (2013) performed a hierarchical clustering analysis, which resulted in the identification of 70 countries into 11 global clusters (Arabic, Anglo, Nordic, Germanic, Latin America, Near East, Latin Europe, African, Far East, and Confucian). We recoded our national origin groups into natives and immigrants into these 11 clusters (79% of our data) and assigned the remaining countries into clusters based on their regional and religious similarity (see app. B online for a full overview). A disadvantage of this method is that Ronen and Shenkar's synthesis is carried out among adults and focuses on work-related values and attitudes. While we expect measurement bias as we apply their clustering of countries to our adolescent data set, we are unaware of a more appropriate source to cluster countries and consider the large number of countries represented in Ronan and Shenkar's cluster solution (70) an advantage that outweighs the disadvantage.

Crude groups (N = 3).—Our last measure of same-ethnic friendship is based on a deliberately crude measure where we consider students from the host country (natives), students from foreign Western countries, and students from non-Western foreign countries. There is no clear definition of Western versus non-Western countries, but they are typically defined as countries within Europe, North America, and Oceania.[6] While a disadvantage of this measure is the considerable variation within the groups, an advantage is that it allows us to include more classes in our analysis.

A Reconsideration of ERGMs

We initially chose ERGMs to answer our research question because we wanted to estimate ethnic homophily; that is, the preference to befriend peers of the same ethnic group above and beyond other explanations for same-ethnic friendship. We aimed to approximate this sociopsychological preference by controlling for important predictors of same-ethnic friendship and ERGMs seemed to be appropriate models as they allowed us to estimate a same-ethnicity parameter while controlling for network properties. Several studies

[6] In the Netherlands, we also consider students with Indonesian and Japanese national origins as originating from Western countries in line with Statistics Netherlands, but it should be noted that Statistics Netherlands recently announced that it is moving away from this definition and will only distinguish between persons with a migration background (yes/no) or larger regional groups. Because our aim is a measure with very few immigrant groups, we do not follow this latest development and count children with an Indonesion and Japanese origin as Western immigrants. See Statistics Netherlands (2023*a*, 2023*b*).

demonstrate that homophily is overestimated without taking network structures like transitivity and reciprocity into account (Moody 2001; Wimmer and Lewis, 2010). In other words: By running ERGMs, we aimed to avoid a measurement validity issue where part of our measure of same-ethnic friendship would be transitivity instead.

While it is true ERGMs are superior to simpler methods in estimating structural effects, we have mentioned before that running ERGMs comes with the major disadvantage of losing a considerable proportion of the sample in our study. The issue of nonconvergence of ERGMs has also been detailed in Martin (2018, p. 276), where the use of ERGMs has been described as a form of "ritualism" in which ERGMs are considered the "default" or "the right way of approaching network data" even when studies' focal interest is in nonstructural parameters (e.g., homophily).

There has been increasing evidence that simpler methods like dyadic models produce similar results to methods that model the entire network (Lindgren 2010; Ragan et al. 2019; Kim et al. 2022). As such, it is valuable to repeat the STMM analyses with a dyadic model instead of a network model.[7] Whereas we believed at the time of writing STMM seven years ago that ERGMs would alleviate measurement bias of ethnic homophily, we have updated our opinion that bias introduced by sample selection does more harm to the credibility of the results. As such, we run two sets of analyses: ERGMs and dyadic multilevel models (MLMs). For each set of analyses, we will analyze the three different measures of ethnicity (national origin groups, cluster groups, and crude groups).

ERGMs.—We run the same model as STMM and KGWZ using the most recent ERGM R package (4.4.0 (2023-01-26)). We inspect each MCMC density and trace plot individually to determine proper convergence for each class. Using these stricter convergence criteria, we only include converged classes in our meta-analysis.

MLMs.—Our data concerns the dyads of all students within classes. Each dyad reflects whether or not a student (called *ego*) nominated the other student (called *alter*) as one of their five best friends in class. The ego and alter unit of analysis are perfectly nested within classes, but the dyad-level unit of analysis does not hierarchically fit into egos and alters. Dyads are cross-nested in egos and alters (fig. 2). As data of this nature violates the independency assumption (Snijders and Bosker 2011), we conduct an MLM analysis. Ideally, we would carry out a cross-classified analysis, but a model with dyads cross-classified in egos and alters returns singular with a random intercept of exactly zero for egos. This indicates that such a model is not the right model for our data, so we simplify our model by removing the random

[7] KGWZ suggest using multilevel network models, but we refrain from these as we expect similar convergence issues.

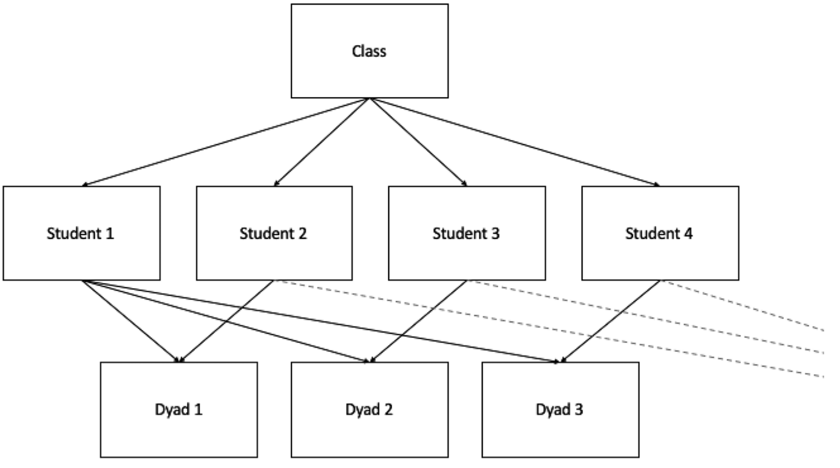

FIG. 2.—The nested structure of the data

intercept parameter for egos. Results with the random intercept and without the random intercept for egos are similar (differences occur at the third or fourth decimal of the coefficient).

Our models include measures of same-group friendship for majority and minority groups and measures of diversity (inverse Herfindahl index),[8] the proportion of natives in class, same-ethnic density, and immigrant diversity (inverse Herfindahl index without taking native students into account). In addition, we include similar measures as in STMM; that is, we control for mutual ties, the number of shared friends (an approximation of transitivity), class size, and gender, socioeconomic status, and gender homophily. More information on the measurement of these variables can be found in STMM.

A final advantage of the MLMs is that we can apply the survey weights to the analysis.[9] As the CILS4EU data set oversampled schools with high immigrant proportions and encountered nonresponse, being able to weigh for the probability that a particular case is included in the study is valuable. We use ego's "house weight" for each dyad. The house weight takes into account the probability of a school and a class being included in the sampling frame and is adjusted for nonresponse on the school, class, and student level. Finally, the house weight rescales the data to the original sample size to avoid deflated standard errors (CILS4EU 2016).

[8] Note that we calculate the inverse Herfindahl using the national origin measure of ethnicity. We checked if the results are sensitive to a measure of the inverse Herfindahl using the cluster ethnicity measure and found that the results are robust.

[9] The results differ slightly between models with and without survey weights, but not to the extent that our conclusions would change.

RESULTS

Native and Immigrant Homophily

Our results show that the MLMs lead to similar estimates of ethnic homophily as ERGMs. Figure 3 shows the native and immigrant homophily coefficient expressed relative to the size of the intercept (full models are found in online app. C, model 1).[10] For each ethnicity measure, figure 3 shows no bias for native homophily and only a slight overestimation for immigrant homophily. This goes against the idea that MLMs would overestimate ethnic homophily (Moody 2001). In addition, figure 3 shows that the coefficient for immigrant homophily decreases with less granular measures. This is expected as these measures are less specific.

The Relationship between Homophily and the Classroom Composition

Tables 1 and 2 show the results of the ERGMs and MLMs for natives and immigrants, respectively. The tables allow a comparison of the results between type of method (ERGM and MLM) and type of ethnicity measure (national origin, cluster, and crude). We included the original STMM results and KGWZ results for comparison. For simplicity, we report the sign of the coefficient (− = negative and + = positive effect, respectively) for significant results. Nonsignificant coefficients are reported as "NS." (The full results including the exact coefficients and standard errors can be found in the online appendixes C and D.)

A first observation to make is that the MLMs are based on a larger sample than the ERGMs and coefficients are more often significant compared to the ERGMs due to more statistical power. For the MLMs, we only excluded classes with unreliable network data.[11] For the ERGMS, a substantial amount of additional classes needed to be excluded because of nonconvergence.

More specifically, table 1 shows significant positive effects of diversity on homophily for native students: the tendency to befriend native students over immigrant students is larger in classes that are more diverse across all types of ethnicity measures and types of models. The original STMM analysis showed evidence for a positive quadratic effect of diversity on native homophily, which is corroborated in the MLMs for all three measures of ethnicity. KGWZ's ERGMs and our additional ERGMs with cruder measures of ethnicity show insignificant quadratic effects. The proportion

[10] The intercept in an ERGM is the "edges" coefficient. We take the relative size to the intercept coefficient, as it is problematic to directly compare the log odds between different models (Mood 2010).

[11] At least 75% of the students participated in the network survey; no more than 10% of the nominations are invalid, no more than two students in the class have never nominated anyone else in the network-related items, and no more than two students in class have never been nominated in any of the network-related items.

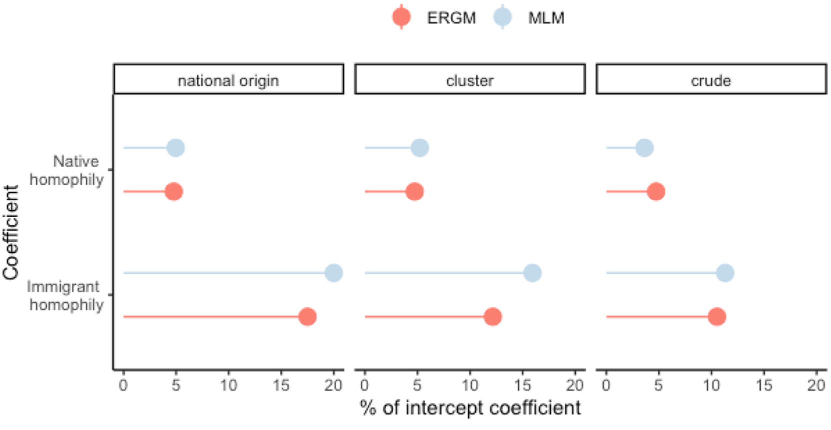

FIG. 3.—Native and immigrant homophily coefficients expressed as a percentage of the intercept coefficients.

TABLE 1
OVERVIEW OF CLASS COMPOSITION EFFECTS ON THE RELATIONSHIP BETWEEN NATIVE HOMOPHILY AND THE LIKELIHOOD OF A FRIENDSHIP TIE

	NATIVE STUDENTS				
	N_{class}	Diversity	Diversity^2	Prop Natives	Prop Natives^2
ERGMS:					
STMM 2016 Original groups (N_{groups} = 158)	523	+	+	−	NS
KGWZ 2023 Original groups (N_{groups} = 158)	342	+	NS	−	NS
STMM 2023 Cluster groups (N_{groups} = 13)	414	+	NS	−	NS
STMM 2023 Crude groups (N_{groups} = 3)	416	+	+	−	NS
MLMs:					
Original groups (N_{groups} = 158)	752	+	+	−	+
Cluster groups (N_{groups} = 13)	752	+	+	−	+
Crude groups (N_{groups} = 3)	752	+	+	−	+

NOTE.—Prop Natives = proportion natives; + = statistically significant positive coefficient ($P < .05$); − = statistically significant negative coefficient ($P < .05$).

TABLE 2
OVERVIEW OF CLASS COMPOSITION EFFECTS ON THE RELATIONSHIP BETWEEN IMMIGRANT HOMOPHILY AND THE LIKELIHOOD OF A FRIENDSHIP TIE

	IMMIGRANT STUDENTS				
	N_{class}	Diversity	Diversity^2	Prop natives	Prop natives^2
ERGMs:					
STMM 2016 National origin groups (N_{groups} = 158)	262	NS	−	NS	−
KGWZ 2023 National origin groups (N_{groups} = 158)	164	NS	NS	+	NS
STMM 2023 Cluster groups (N_{groups} = 13)	213	NS	NS	+	NS
STMM 2023 Crude groups (N_{groups} = 3)	262	NS	NS	+	−
MLMs:					
National origin groups (N_{groups} = 158)	752	NS	NS	NS	NS
Clusters groups (N_{groups} = 13)	752	NS	−	NS	NS
Crude groups (N_{groups} = 3)	752	+	−	NS	−

NOTE.—Prop Natives = proportion natives; + = statistically significant positive coefficient ($P < .05$); − = statistically significant negative coefficient ($P < .05$).

of native effects on native homophily are similar to the diversity effects: the ERGMs show evidence for linear effects, whereas the MLMs suggest nonlinear effects. All in all, this suggests that failing to reject the null hypothesis for the quadratic effect of diversity and the proportion of natives is due to the restricted range issue.[12]

The ERGMs in table 2 do not exhibit significant evidence that immigrant homophily is related to the diversity of the classroom, except for the original STMM ERGM analysis. For the MLMs, the results depend on how we measure ethnicity. The analysis with the most detailed measure (national origin) shows insignificant effects of diversity and diversity. When we move to more broad ethnicity measures, we find a significant negative quadratic coefficient between diversity and homophily for the cluster and crude ethnicity measure. The main effect of diversity is not significant for the cluster ethnicity measure and significant and positive for the crude ethnicity

[12] We examined the MLMs restricting the sample to the classes that converged for the ERGMs and found that the quadratic effects were no longer significant (results not shown). In other words, the difference between the ERGMs and the MLMs can be explained by the sample selection differences.

measure. All in all, the findings suggest that immigrant homophily is unrelated to the classroom's level of ethnic diversity, except in the most diverse classes. There we find that immigrants tend to have lower levels of homophily. In addition, this effect is only picked up in approaches that examine full range of diversity (i.e., the MLMs) and broader ethnicity measures (i.e., the cluster and crude measures). The results for the proportion of native predictors corroborate this: the effects of the proportion of natives are insignificant for the national origin and cluster ethnicity measures and only the quadratic effect is significantly negative in the analysis using the crude ethnicity measure.

In addition to examining the diversity and proportion of native effects, we also repeated the analyses for the same-ethnic friendship density and immigrant diversity effects (table 3; full results are found in online app. D). We argued in our original article that these two effects provide additional

TABLE 3
Overview of the Same-Ethnic Immigrant or Native Friendship Density and Immigrant Diversity Effects on Native and Immigrant Homophily[a]

		Same-Ethnic Friendship Density		Immigrant Diversity	
	N_{class}[b]	Native	Immigrant	Native	Immigrant
ERGMs:					
STMM 2016					
National origin groups (N_{groups} = 158)	523/262	+	NS	−	NS
KGWZ 2023					
National origin groups (N_{groups} = 158)	342/164	NS	NS	NS	+
STMM 2023					
Cluster groups (N_{groups} = 13)	414/213	NS	NS	−	NS
STMM 2023					
Crude groups (N_{groups} = 3)	416/262	NS	NS	−	NS
MLMs:					
National origin groups (N_{groups} = 158)	752	+	+	−	NS
Clusters groups (N_{groups} = 13)	752	+	+	−	−
Crude groups (N_{groups} = 3)	752	+	+	−	−

Note.—Prop Natives = proportion natives; + = statistically significant positive coefficient ($P < .05$); − = statistically significant negative coefficient ($P < .05$).

[a] Same-ethnic friendship density refers to same-ethnic *immigrant* density for natives and same-ethnic *native* friendship density for immigrants.

[b] The first number in this column refers to the number of classes with converging ERGMs estimating native homophily. The second reflects the number of classes with converging ERGMs estimating immigrant homophily.

support for the theory that native homphily is more likely driven by feelings of ethnic threat compared to immigrant homophily.

The original ERGMS showed evidence for the hypothesis that native homophily is stronger in classes where the density of immigrant friendship networks is higher, whereas immigrant homophily is unrelated to native friendship density. KGWZ found only null results in their replication. Our additional ERGM and MLM analyses with different ethnicity measures also do not show clear support for the hypothesis.

For native students, the ERGMs for the cluster and crude measure lack evidence that native homophily is related to immigrant friendship density. The MLMs show significant and positive effects: native homophily tends to be stronger in classes where immigrants have denser friendship networks (net of the total diversity in classrooms).

For immigrant students, the ERGMs results display insignificant effects of same-ethnic native friendship density. The MLMs, however, show *positive* effects. This provides support for the idea that both natives and immigrants feel more threatened by dense groups of friendships among outgroup members, which is against our theory.

Finally, we hypothesized in our original article that ethnic homophily among both immigrants and natives decreases with immigrant diversity while controlling for the share of natives. STMM were only able to reject the null hypothesis for native homophily and KGWZ found null results for both native and immigrant homophily. For native homophily, our ERGMs and the MLM corroborate this hypothesis as we find significant negative effects. For immigrant homophily, we only find evidence for the immigrant diversity hypothesis in the MLMs for the cluster and crude ethnicity measure. This means that finding evidence for the immigrant diversity hypothesis depends on the method and the type of ethnicity measure.

CONCLUSION

The goal of this response was to reexamine our theory concerning the differential effects of the classroom composition on ethnic homophily for native and immigrant students. STMM showed evidence for this theory, but KGWZ showed that these results were based on ERGMs that did not properly converge. Their replication based on converged ERGMs showed no evidence of differential effects. In this response, we agree with KGWZ's methodological correction of our work, but we argue (in line with KGWZ's discussion of their results) that their approach of discarding classes in which the ERGM does not fit limits the ability to answer the original question. We showed that the sample in which ERGMs converge (which KGWZ use) is small and biased. These two features could explain the null results KGWZ report and were given further investigation in this reply.

All in all, the original article, KGWZ's comment, and this response show the difficulty of analyzing ethnic homophily in school classes. The goal of scientific research is to produce generalizable knowledge about the population of interest, and it is therefore important to minimize threats to validity. Threats to validity can show up anywhere in the research process, ranging from failing to achieve an unbiased sample to using inappropriate statistical methods to analyze the data. In this case, we have been balancing these particular threats to validity: ERGMs can arguably better estimate the homophily parameters, but they can only do so in a biased sample. MLMs can be estimated on a less biased sample, but these models are not as sophisticated as ERGMs in taking into account social network dependencies in the data. Similarly, detailed ethnicity measures do justice to the fact that large ethnic groups consist of smaller and more specific national origin groups with distinct histories and experiences. On the other hand, examining national origin groups results in a loss of statistical power, and it is reasonable to explore broad ethnicity measures. Whereas European children of immigrants may identify with their national origin group, children likely resort to broader, pan-national origin groups in small classrooms where they do not expect to meet someone exactly like them.

For future research, it has become clear that goodness-of-fit plots should be individually inspected for ERGMs until a better alternative becomes available. KGWZ point out that most researchers will agree about convergence for many classes (extreme cases of misfit), but there are also borderline cases where researchers will reasonably make different decisions about whether or not the model converged. It is straightforward to be transparent about convergence with one network as researchers can show the goodness of fit in their manuscript, but it is difficult to scale this to hundreds of networks.[13] We want to emphasize that we believe ERGMs are valuable models for many research questions, but they have important disadvantages for this particular study: the school classes are too small to model constrained friendship choices for fine-grained ethnic groups. Our results showed as expected that the sample size of converged ERGMs is smaller with more detailed ethnicity measures (i.e., more groups).

In terms of the substantive research question, our response shows evidence that there are differential effects of the class composition for native and immigrant homophily, but the signal is modest. Future research could explore different research designs to weigh in on the inconsistent findings. For example, Abascal, Xu, and Baldassarri (2021) use an experimental research design to show that various racial groups perceive diversity differently.

[13] The multilevel network model is promising in this endeavor but was beyond the scope of this paper, as the package multisiena was not compatible with the latest version of R at the time of this writing.

White Americans associate diversity with heterogeneity (i.e., evenly distributed racial groups in a particular setting, akin to what we call "diversity") more strongly than Black, Asian, and Latino Americans. Black, Asian, and Latino Americans perceive diversity more strongly with their group's representation instead. In addition, these associations are related to attitudes about immigration. These findings show that majority and minority groups experience the compositions of contexts differently and that these perceptions relate to how people feel about other groups in their communities. In sum, while more research is necessary, we conclude that there is still truth to threat theory.

Sanne Smith
Stanford University

Frank van Tubergen and Ineke Maas
Utrecht University

Daniel A. McFarland
Stanford University

REFERENCES

Abascal, Maria, Janet Xu, and Delia Baldassarri. 2021. "People Use Both Heterogeneity and Minority Representation to Evaluate Diversity." *Science Advances* 7 (11): eabf2507.

CILS4EU. 2016. Children of Immigrants Longitudinal Survey in Four European Countries. Technical Report. Wave 1 – 2010/2011, v1.2.0. Mannheim: Mannheim University.

Henriksson, Widar, and Simon Wolming. 1998. "Academic Performance in Four Study Programmes: A Comparison of Students Admitted on the Basis of GPA and SweSAT Scores, With and Without Credits for Work Experience." *Scandinavian Journal of Educational Research* 42, no. 2 (1998): 135–50.

Hunter, John. E., and Frank L. Schmidt. 1990. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. London: Sage Publications.

Kim, Lanu, Sanne Smith, Linus Dahlander, and Daniel A. McFarland. 2022. "Networking a Career: Individual Adaptation in the Network Ecology of Faculty." *Social Networks*. https://doi.org/10.1016/j.socnet.2022.04.002.

Kretschmer, David, Johanna Gereke, Fabian Winter, and Nan Zhang. 2023. "No Differential Effects of Classroom Ethnic Composition on Native and Immigrant Friendship Segregation: Comment on Smith et al. 2016." *American Journal of Sociology*, vol. 129. In this issue.

Lindgren, Karl-Oskar. 2010. "Dyadic Regression in the Presence of Heteroscedasticity—an Assessment of Alternative Approaches." *Social Networks* 32 (4): 279–89.

Martin, John Levi. 2018. *Thinking through Statistics*. Chicago: University of Chicago Press.

Mood, Carina. 2010. "Logistic regression: Why We Cannot Do What We Think We Can Do, and What We Can Do about It." *European Sociological Review* 26 (1): 67–82.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107 (3): 679–716.

Ragan, Daniel T., D. Wayne Osgood, Nayan G. Ramirez, James Moody, and Scott D. Gest. 2019. "A Comparison of Peer Influence Estimates from SIENA Stochastic Actor-based Models and from Conventional Regression Approaches." *Sociological Methods and Research* 51 (1): 357–95.

Ronen, Simcha, and Oded Shenkar. 2013. "Mapping World Cultures: Cluster Formation, Sources and Implications." *Journal of International Business Studies* 44:867–97.

Simsek, Müge, Frank van Tubergen, and Fenella Fleischmann. 2022. "Religion and Intergroup Boundaries: Positive and Negative Ties among Youth in Ethnically and Religiously Diverse School Classes in Western Europe." *Review of Religious Research* 64 (1): 1–33.

Smith, Sanne, Ineke Maas, and Frank van Tubergen. 2014. "Ethnic Ingroup Friendships in Schools: Testing the By-Product Hypothesis in England, Germany, the Netherlands and Sweden." *Social Networks* 39:33–45.

Smith, Sanne, Frank van Tubergen, Ineke Maas, and Daniel A. McFarland. 2016. "Ethnic Composition, and Friendship Segregation: Differential Effects for Adolescent Natives and Immigrants." *American Journal of Sociology* 121:1223–72.

Snijders, Tom A. B., and Roel J. Bosker. 2011. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage Publications.

Statistics Netherlands. 2023*a*. "New Classification of Population by Origin." May 7. https://www.cbs.nl/en-gb/longread/statistische-trends/2022/new-classification-of-population-by-origin

Statistics Netherlands. 2023*b*. "Person with a Western Migration Background." May 7. https://www.cbs.nl/en-gb/our-services/methods/definitions/person-with-a-western-migration-background.

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116 (2): 583–642.