Department of Methodology

MY499 Dissertation

Interethnic Friendships under Ethnically Segregated Pathways in Education: The Case of Malaysia

Candidate Number:	56074
Word Count:	9,556 words
Supervisor:	Dr Eleanor A. Power
Date:	15 August 2023

I am grateful for the generous support from the LSE Saw Swee Hock Southeast Asia Centre Dissertation Fieldwork Grant and the LSE Department of Methodology Dissertation Funding Grant which enabled the fieldwork for data collection purposes.

Abstract

Ethnic homophily in plural societies leads to the perpetuation of social inequalities and unresolved ethnic tensions. The Contact Hypothesis posits that intergroup contact can reduce prejudice and improve ethnic relations. This paper examines the associations between intergroup contact in national schools, the lack thereof in ethnic-exclusive education institutions, and the extent of homophily in the friendship networks of Malaysian university students. Malaysia offers a unique context to study ethnic relations in a plural society due to ethnic segregation in education pathways. Utilising moderation analysis of the ethnic homophily terms in an egocentric exponential random graph model (ERGM), I find evidence that attending national secondary schools is associated with lower levels of ethnic homophily in university friendship network. I find no evidence to support the claim that attending ethnic-exclusive education institutions amplify ethnic homophily. These findings suggest that friendship formation at the university is affected by experiences in both primary and secondary school, implying that contact effects are enduring.

Keywords: Ethnic Homophily, Contact Hypothesis, University Friendship Networks, Malaysia, Egocentric Exponential Random Graph Models

Table of Content

I. Introduction 4

II. Context in Malaysia 9

III. Research Design 14

IV. Data Collection 20

V. Analysis 25

VI. Conclusion 44

References 48

Appendices 54

List of Tables and Figures

Table 1: Typography of Malaysian Primary and Secondary Education Landscape.....13

Table 2: Predictors of Proportion of Friends of Malay, Chinese, or Indian Ethnicity.....38

Table 3: Ego ERGM Results.....40

Table 4: Results of ERGM based on Simulated Complete Network.....41

Table 5: Results of Moderation Analysis.....43

Figure 1: Summary of Respondent Characteristics 26

Figure 2: Ethnic Diversity of Universities in Sample.....27

Figure 3: Sankey Diagram of One-Step Education Pathways by Ethnicity.....28

Figure 4: Sunburst Diagrams of Education Pathways by Ethnicity.....29

Figure 5: Distribution of Ethnicities of Primary and Secondary Schools in Sample.....31

Figure 6: Degree Distribution of Personal Networks in Sample 32

Figure 7: Degree Distribution by Ethnicity 33

Figure 8: Distribution of EI Index by Ethnicity (dotted line = mean) 34

Figure 9: Proportion of Alter Ethnicity Among All Alters Nominated by Ego Ethnicity.....36

I. Introduction

Social networks are characterised by homophily, where individuals tend to associate with others who share similar attributes, such as socioeconomic background, values, or interests (McPherson, Smith-Lovin and Cook, 2001). This natural inclination has profound impact on the perpetuation of social inequalities, as economic opportunities and resources may be concentrated and consolidated within a homogenous network of similar individuals (DiMaggio and Garip, 2012). These effects are exacerbated in plural societies, where a population divided along ethnic lines coexist within a single political entity (Smith, 1971). Ethnic homophily in interpersonal relationships implies assortativity by ethnicity in the population, and social inequalities manifest as ethnic disparities. Division and disparities then fuel tensions and precipitate ethnic conflict as different groups contend for control over economic resources and political power (Hoetink, 2011). Mediating ethnic relations is a key priority for policymakers in plural societies.

Building a unifying national identity can reduce conflict and maintain harmony between groups (Gaertner *et al*., 2000); however, the process of replacing ethnic identity with an overarching national identity presents a formidable challenge. Tajfel and Turner's (1979) Social Identity Theory posits that identity emerges from the human propensity to categorise oneself and others into groups based on salient attributes such as shared cultural practices and traditions, language, and collective history. Self-consciousness of group membership leads to the identification of in-groups, people whom one perceives share the same social identity, and out-groups, people who do not. The existence of an "other" is necessary as a means for differentiating the in-group in the process of identity development (Triandafyllidou, 1998). Plural societies imply the existence of many "others", as individuals are situated a setting with a variety of cultural practices, language, and even physical characteristics seen frequently

almost every day, thus fuelling the development of ethnic consciousness over a feeling of national identity. Ethnic identity provides belonging and feelings of inclusion which affects social interactions. There is a tendency to prioritise distinctions rather than commonalities with "others" who does not share a common identity. This inclination leads to a preference towards individuals who are like oneself, resulting in the prevalence of homophily in personal connections.

Ethnic differences are an enduring social reality in plural societies, but antagonism need not be. Allport (1954) argues that contact between different social groups under favourable conditions leads to reduced prejudice and increased acceptance of out-group members. Empirical evaluations of the "Contact Hypothesis," as Allport's idea came to be known, focus on the effect of a "contact intervention" on perception towards members of out-groups. Contact interventions vary in nature, ranging from brief, controlled interactions to unscripted engagements over a longer period. Researchers measure a variety of outcomes, including self-reported or socially evaluated levels of attitudes towards the out-group and behavioural outcomes such as observed interactions and friendships among members of different social groups. There is evidence to support the Contact Hypothesis across various contact intervention types and outcomes (Pettigrew and Tropp, 2006; Paluck, Green and Green, 2019). These findings suggest that intentional efforts in encouraging direct contact and shared activities between ethnic groups can lead to improved relations. However, the interventions examined typically involve contact spanning less than a year, and outcomes are measured immediately after the intervention. Generalisability to prolonged and sustained interethnic contact, such as that in plural societies, is limited.

In plural societies, tensions arise precisely due to the inevitability of intergroup contact. Studying the effects of contact in these settings requires the recognition that social interactions occur in different foci of activity, such as in schools, work establishments, voluntary organisations, neighbourhoods, and so on (Feld, 1981). Among these, schools hold the greatest potential to affect one's perception towards out-group members over a prolonged period, given that mandatory education of at least seven to twelve years is implemented in most countries. National schools in plural societies provide a unique point of convergence in the life journeys of individuals from various backgrounds (Selvadurai *et al.*, 2015). They offer unique opportunities for studying sustained contact between different groups in an ethnically diverse social environment, which has been found to be associated with an increase in the formation of interethnic friendships or an increase in the ethnic diversity of friendships (Kokkonen, Esaiasson and Gilljam, 2015). National schools are also mediums for instilling mutual understanding among ethnic groups, which potentially affect the long-term preferences and perceptions of individuals. Studying the individuals who attended public schools in plural societies allows investigation into the durability of the contact effect.

This paper adds to the small number of studies focusing on the contact effect of schooling (Thijs and Verkuyten, 2014; Merlino, Steinhardt and Wren-Lewis, 2019; Burgess and Platt, 2021). I explore previously unexamined questions in the Contact literature by considering the case of Malaysia, a country with great ethnic diversity. Specifically, I investigate the relationship between ethnic diversity in the friendship networks of university students and their education pathway, taking advantage of the plethora of options for both primary and secondary education in Malaysia. With the assumption that national schools, which have greater ethnic diversity, offer greater opportunities for intergroup contact, the Contact Hypothesis implies that students who attended national schools will exhibit more

positive attitudes towards individuals from a different ethnic group, which is associated with a higher probability of holding interethnic friendships and lower levels of ethnic homophily.

I also test whether ethnic segregation in the education system amplifies the effects of ethnic homophily. Segregation within schools or the education system creates an environment where opportunities for meaningful intergroup contact are limited. When students from different ethnic backgrounds are systematically separated into distinct schools, their interactions with peers from other groups become constrained. This separation can reinforce existing homophilic tendencies, and the lack of interethnic interactions can perpetuate in-group biases, which hinder the development of intercultural understanding. The amplification of ethnic homophily in segregated educational settings may hinder the positive intergroup relations and contribute to the persistence of social and cultural divisions.

Using a sample of 203 personal friendship networks of university students with a total of 2,182 friends, I analyse how educational pathway affects with the level of homophily in these networks. I utilise recent advances in social network analysis on the moderation analysis of network statistics in exponential random graph models (Duxbury, 2023). The results support the claim that intergroup contact improves relations between ethnic groups in a plural society. In addition, I find no evidence to suggest that ethnic segregation into different types of schools at the primary and secondary levels. These findings offer some empirical evidence to support the effectiveness of national schools in fostering multiculturalism while informing the political debate in Malaysia on the abolition of vernacular schools, where the proponents of abolition claim that these schools are detrimental to nation-building and encourage ethnic segregation.

The rest of this paper is organised as follows: the next section provides a brief background on Malaysia, providing a brief historical narrative that led to the present demographics and an overview of its unique education system. Section III outlines the hypotheses that are tested in this paper and describes the research design. Section IV describes the data collection method. Section V presents the results of statistical analyses. Finally, Section VI concludes by discussing the implications of the findings and proposes recommendations for future research.

II. Context in Malaysia

Malaysia is a plural society consisting of three major ethnic groups, namely Malay, Chinese, and Indian, along with other smaller ethnic minorities. The Department of Statistics Malaysia (2023) reports that 57.9 percent of the population are ethnic Malays, numbering 17.6 million. Chinese Malaysians make up 22.6 percent of the population with 6.9 million. The 2 million Indian Malaysians make up 6.6 percent of the population. The remainder 12.9 percent consists of indigenous tribes, East Malaysian ethnicities, and others. For the purposes of this study, I focus on interethnic relationships between the three major ethnic groups in West Malaysia, or peninsular Malaysia. Ethnic relations and politics in the East Malaysian states of Sabah and Sarawak are different and warrants a separate discussion due to different ethnic compositions and proportions, an arguably different social environment, and the autonomy that the state government have over the states' education policy.

Each of the three major ethnic groups have distinct cultural traditions and history, and each hold onto their respective identities firmly. The majority group, ethnic Malays, view themselves as the indigenous population of the peninsular Malaysia. The Malay culture, the Malay language, and the Islamic faith are the three defining characteristics of the Malay ethnicity (Omar, 2011). The bulk of Chinese and Indian migration to present-day Malaysia occurred in the 19th century under British colonial rule, which was a period of high demand for labourers in tin mines and sugar and rubber plantations. Chinese and Indian migrants did not assimilate into Malay culture but retained their own separate cultural identities. Many scholars attribute this to the divide-and-rule strategy employed by the colonial administration (Abraham, 1983). Under this policy, British colonisers allowed the traditional Malay ruling class to retain political domination of the indigenous peasant population, while Chinese and Indian migrants were accorded some degree of self-governance. The Chinese migrants primarily originated

from the Guangdong and Fujian provinces of present-day China which enjoyed access to the South China Sea and where maritime trade routes have been established centuries before. Chinese capitalists controlled the entire chain of migration, from recruitment to the provision of common goods necessary to maintaining an overseas Chinese community (Amrith, 2010). On the other hand, Indian foremen (*kongany*) were responsible for the recruitment of their fellow villagers in the present-day South Indian state of Tamil Nadu and for the welfare of the migrant labourers who were indentured to the European plantation owners (Lee, 1989).

Intergroup contact during the colonial era was minimal and separate political parties formed to serve the interests of each ethnic group during the negotiations for Malayan independence during the 1950s (Noor and Leong, 2013). During independence in 1957, the Malaysian Federal Constitution became the foundation of the new Federation. It recognises the special position of ethnic Malays and other indigenous populations in the country. The term Bumiputera (translated as "sons of the soil") is later coined to collectively refer to this special group. The division of the economy along ethnic lines has resulted in economic disparities between ethnic groups even after independence. Ethnic struggle came to define the first decade of Malayan independence, with riots breaking out in 1969. The event is attributed to longstanding inequalities among ethnic groups. In the aftermath, a series of affirmation action policies which encompasses economics, land ownership, education, and other facets of society were implemented under the New Economic Policy (NEP).

One of the educational goals of the NEP is to increase the proportion of Bumiputera students in tertiary education. This is done through special pre-university institutions that were exclusive to Bumiputera students. Lee (2012) estimates that 55 percent of university places were allocated to Bumiputera students for a period of 30 years from the implementation of the

NEP. Over the same period, demand for higher educated workers rose in the quickly industrialising Malaysia. The limited slots at public universities for Chinese and Indian students mean that they had to attend private universities to obtain tertiary education. Lee (2017) argues that enrolment quotas contributed to ethnic segregation through the displacement of non-Bumiputera students from private universities. Despite the initial NEP aim to reduce ethnic inequalities, ethnic segregation at the university level emerged as Bumiputera students tend to attend public universities while non-Bumiputera students attend private universities.

Ethnic segregation is also present in the primary and secondary education system in Malaysia, with different education pathways for each ethnic groups as each tend to choose a distinct type of primary and secondary school to attend. At the primary level, there are Chinese and Tamil vernacular primary schools in additional to national primary schools. The medium of instruction in these schools are Mandarin and Tamil respectively. Raman and Sua (2010) find that almost all Chinese parents prefer to send their children to vernacular school, and for Indian parents that figure exceeds half. There are also religious schools or Islamic primary schools that cater to the needs of the majority Muslim Malay community. At the secondary level, Chinese Independent Schools provide Chinese-medium secondary education to students utilising a curriculum which incorporates Chinese cultural values and that is distinct from the national curriculum. Additionally, several Bumiputera exclusive secondary schools are established following the implementation of the NEP. Despite the plethora of options for secondary education, national secondary school remains the most common choice for Malaysian students due to its affordability and availability in all states of peninsular Malaysia. Table 1 contains a typography of the different types of schools for primary and secondary education in Malaysia.

Ethnicity is a highly salient social classification in Malaysian society today. Almost all aspects of life require the acknowledgement of one's ethnicity: at birth, all individuals are assigned an ethnicity; opening a bank account requires individuals to fill out their ethnicity; applying to public universities requires filling out one's ethnicity. The pervasiveness of ethnicity as a social categorisation influences the social networks of Malaysians. Studies have noted that ethnic homophily characterises the social relationships in Malaysia (Tey, Halimah and Singaravelloo, 2009; Yeoh, 2006). This feature of Malaysian society and ethnic segregation across all levels of the education system provides a unique context for studying the dynamics of interethnic friendship formation in plural societies.

Table 1: Typography of Malaysian Primary and Secondary Education Landscape

Primary Schools	
Sekolah Kebangsaan (SK)	Translated as "National Schools," these are national primary schools. There is a total of 5,948 such schools throughout Malaysia, which is one school for roughly 1,345 households.
Sekolah Kebangsaan Jenis Cina (SJKC)	These are Chinese vernacular primary schools, where Mandarin is taught in addition to English and Malay. Science and mathematics are taught in Mandarin and English. There is a total of 1,305 SJKCs in Malaysia.
Sekolah Kebangsaan Jenis Tamil (SJKT)	These are Tamil vernacular primary schools, where the Tamil language is taught in addition to English and Malay. Science and mathematics are taught in English and the Tamil language. There is a total of 528 SJKTs in Malaysia.
Sekolah Agama	Translated as "Religious Schools," these are schools that incorporate Islamic teachings in the curriculum. They are established and administered by state religious authorities rather than the Ministry of Education. At the primary level, these schools run parallel to the national school system, and many Muslims attend national school in the morning and Sekolah Agama in the afternoons or evenings.
Secondary Schools	
Sekolah Menengah Kebangsaan (SMK)	Translated as "National Secondary School," these are the national schools where most Malaysians attend for secondary education. There is a total of 1,989 SMKs throughout Malaysia.
Chinese Independent School (CIS)	These are secondary education institutions that use Mandarin as the medium of language. They do not receive fundings from the Malaysian government and are self-funded or sustained by donations by the Chinese community (Siah *et al.*, 2015). Many were established by Chinese migrants during the 19th and 20th centuries. There is a total of 60 CIS today.
Sekolah Menengah Kebangsaan Agama (SMKA)	Translated as "Religious Secondary School," these are established and administered by state religious authorities rather than the Ministry of Education, like Islamic schools at the primary level. There is a total of 61 SMKAs in Malaysia.
Maktab Rendah Sains MARA (MRSM)	Translated as "MARA Science Junior College," these are special boarding schools managed by Majlis Amanah Rakyat (MRSM, or translated as "People's Trust Council"), which is a government agency formed for the purposes of Bumiputera empowerment. MRSM implements a 90 percent Bumiputera enrolment quota, and its aim is to accelerate the acquisition of science and technology subjects among the Bumiputera population.
Sekolah Berasrama Penuh (SBP)	Translated as "Full Boarding Schools," these are selective secondary schools modelled after British boarding schools. SBPs were established through the Second Malaysia Plan (1975 - 1980) with the aim of providing quality science and technology education to Malaysian students with high potential.

III. Research Design

This paper explores how schooling affects ethnic homophily by testing the following hypotheses:

Hypothesis 1: Students who attended national schools, which are assumed to have greater ethnic diversity, are more likely to have a diverse friendship group at university.

Hypothesis 2: Students who attended schools that are exclusive to certain ethnicities, whether due to language or religious barriers or ethnic quotas, are less likely to have a diverse friendship group at university.

I utilise a self-administered survey consisting of 203 Malaysian university students as respondents. The survey collects information on the respondents (egos) and their friends (alters). The resulting set of personal network data, or egocentric network data, allows me to explore how attributes of each ego, such as their reported ethnicity and education pathway, affect the composition of their friendship network. Sampling egocentric network data offers several advantages over complete networks. Firstly, it is significantly less costly to implement as complete networks require interviewing all members of the selected social setting to capture relationships between all the members. Secondly, it mitigates concerns over the protection of the anonymity of individuals involved in the study as individuals in a complete network are more likely to be identifiable. Finally, probabilistic sampling methods can be incorporated into egocentric sampling of network data, which ensures that the relevant groups of interest are adequately sampled (Perry, Pescosolido and Borgatti, 2018). On the other hand, complete networks are suitable when the position of ego in the population network, or the roles of

individuals in the network. Since this is not the primary focus of this study, egocentrically sampled network data is adequate.

This paper approaches ethnic homophily in a way that deviates from the conventional framework of distinguishing between baseline and inbreeding homophily. Baseline homophily is obtained from a network model of random assortment. This measures homophily which is simply due to the chance of meeting diverse groups in the network. The observed network is then compared to this baseline network model, and the difference between the two is referred to as inbreeding homophily. This measure of homophily is interpreted as a preference for similar individuals that causes them to associate more often than expected given their relative numbers in the pool of possible relationships (Marsden, 1988). Calculating baseline and inbreeding homophily requires complete network data and knowledge of the proportions of the constituent groups in the network.

In contrast, this paper approaches homophily from an individual-centric perspective. Homophily is conceptualised as the inclination of an individual to form friendships with others of the same race. An increased likelihood of forming friendships with those sharing the same ethnicity indicates higher levels of ethnic homophily, while a decrease in the likelihood of forming friendships with those of the same ethnicity indicates lower levels of ethnic homophily. In egocentrically sampled network data, each personal network can be thought of as a single complete network, for which a measure of ethnic homophily can be obtained. The hypotheses of this paper can then be operationalised as I explore how individual attributes are associated with the level of homophily of each individual network. An individual exhibiting a higher likelihood of forming friendships within the same ethnicity is inherently prone to cultivating a larger circle of same-ethnicity friends.

The hypotheses stated above can be categorised as a "theory of networks" (Borgatti and Halgin, 2011), where how attributes of an individual, in this case, ethnicity and education pathway, are said to affect the individual's social network. Attending national school is expected to increase the level of ethnic homophily in an individual's personal friendship network at university. An underlying assumption is that an individual's experience before university affects friendship formation patterns at the university level. The hypotheses are derived from the Contact Hypothesis. National schools have a greater level of ethnic diversity than other types of schools, thus offering students with enhanced levels of contact with students of other ethnicities. This reduces prejudice and encourages positive perceptions of out-groups, which leads to greater likelihood of having a more diverse friendship network at university than students who did not attend national school, implying lower levels of ethnic homophily. On the other hand, schools that are exclusive to certain ethnicities, such as the vernacular schools and the Bumiputera exclusive institutions, mean lower levels of contact with members of the out-group as these students are excluded from enrolment through two distinct mechanisms. The first is exclusion based on language or religious barriers. Vernacular schools, religious schools, and Chinese Independent Schools are examples of this type of exclusion. The second means of exclusion is based on ethnic quotas in Bumiputera exclusive institutions.

Social network analysis offers a range of tools to analyse network data. It encompasses methods to describe network patterns and analyse network processes. For inferential analysis, network researchers utilise exponential random graph models (ERGM), which models the probability of observing a particular network as a function of a vector of network statistics (Hunter *et al.*, 2008). A simplified way of expressing an ERGM is:

$$\Pr(Y = y) = \frac{\exp(\theta^{\top}\delta(x, y))}{c(\theta)} \qquad (Eq. 1)$$

where $\delta(x, y)$ is a mapping function that computes sufficient statistics encompassing the change in the probability of observe the network in response to the change in an exogenous attribute, x, of a network actor while leaving the rest of the network unchanged. $\theta^{\top}$ is the coefficient of the network statistic, and $c(\theta)$ is a normalising constant to ensure that probabilities range between the values zero and one. While ERGMs are originally used to model complete networks, Krivitsky and Morris (2017) developed a suite of statistical tools that allow ERGMs to be implemented using egocentric network data.

The key network statistic of interest is ethnic homophily, which can be interpreted as the increase in the log odds of observing the given network in response to two actors being of the same ethnicity as opposed to being of different ethnicities, holding on the rest of the network fixed. The ethnic homophily ERGM term itself does not answer the hypotheses of this paper. Rather, it produces a means for measuring the level of ethnic homophily in the network. I conduct moderation analysis of the ethnic homophily ERGM term, where I analyse the change in the homophily ERGM term in response to a change in another actor attribute, such as the type of school the individual attended for primary or secondary education. Essentially, moderation analysis of the ethnic homophily ERGM term allows me to answer the question of how going to a particular type of school moderates the effect of homophily.

In order to perform the moderation analysis, the network statistics must be interpreted in terms of their average marginal effects (AME) to overcome the problem of scaling in nonlinear probability models (Duxbury, 2023). Omitted variables, model misspecification, and measurement error can contribute to violation of the assumption on the error variance, leading to biased coefficient estimate for interaction terms. Although coefficients are biased due to the problem of scaling, the predictions from ERGM are not affected. This property enables

moderation analysis by interpreting ERGM coefficients in terms of marginal effects. The marginal effect can be calculated at each dyad by taking the partial derivative for the coefficient:

$$\text{Marginal Effect at } i = \theta \frac{\partial p_{ij}}{\partial x} \qquad (Eq. 2)$$

This quantity is the increase in the probability of observing a tie, p_{ij}, in response to a one-unit increase in network attribute x at observation i. To incorporate information on marginal effects at all observations, Duxbury (2023) recommends averaging over all observations to obtain the AME:

$$AME_x = \theta \frac{1}{n} \sum \frac{dp_{ij}}{dx} \qquad (Eq. 3)$$

AME_θ is the average change in p_{ij} in response to a one-unit increase in x.

Moderation analysis involves the estimation of the change in the AME, ΔAME, of the ethnic homophily term when a moderating variable is increased by one, or in the case of a binary variable, when the variable is one instead of zero.

$$\Delta AME = AME_x^{g=1} - AME_x^{g=0} \qquad (Eq. 4)$$

With x being the ethnic homophily term, $AME_x^{g=1}$ gives the average change in the probability of a tie when two individuals are of the same ethnicity when the moderating variable g is equal to one.

$$AME_x^{g=1} = \theta \frac{1}{n} \sum \frac{dp_{ij}^{g=1}}{dx^{g=1}} \qquad (Eq. 5)$$

On the other hand, $AME_x^{g=0}$ is the average change in the probability of a tie when two individuals are of the same ethnicity when the moderating variable g is equal to zero.

$$AME_x^{g=0} = \theta \frac{1}{n} \sum \frac{dp_{ij}^{g=0}}{dx^{g=0}} \qquad (Eq. 6)$$

Asymptotic standard errors are obtained with the Taylor linearisation method, also called the delta method (Raykov and Marcoulides, 2004).

Since the hypotheses concern how schooling affects the strength of ethnic homophily, g is a vector of binary variables for each type of school. I perform hypothesis testing on the key variable of interest – the change in AME. The null hypothesis H_0 is that there are no moderation effects, or no change in the AME. In other words, schooling does not change the effect of ethnic homophily. This is tested against the alternative H_1 that attending national schools, which have a diverse mix of ethnicities, negatively moderates the effect of ethnic homophily, and the other alternative H_2 that attending a school that is exclusive to certain ethnicities positively moderates the effect of ethnic homophily.

$$H_0: \Delta AME_\theta = 0$$

$$H_1: \Delta AME_\theta < 0$$

$$H_2: \Delta AME_\theta > 0$$

IV. Data Collection

Between 15 May 2023 and 29 May 2023, I conducted fieldwork in five universities in the Greater Kuala Lumpur area to collect data on the personal friendship networks of Malaysian university students. I focus on the friendship networks of university students because schools are avenues where youths are socialised into their ethnic cultures (Collins and Annett, 1975), and the effects of socialisation is likely more prominent upon finishing school. This is, to the best of my knowledge, the first large-scale collection of the friendship network data of Malaysian university students. Due to the absence of publicly available datasets on friendship networks in Malaysia, and previous studies have also relied on self-administered surveys to study interethnic friendships (Yeoh, 2006; Tey *et al*., 2009). The previous study focused only on Economics students at *Universiti Malaya* (UM), Malaysia's premier public university. To capture more variation in the education pathways, I utilise a broader sampling scope which includes two public universities, namely UM and *Univesiti Kebangsaan Malaysia* (UKM, translated as "National University of Malaysia"), and three private universities. The inclusion of private universities allows for a wider variety of education pathways in the sample since most students in public universities must have attended national secondary school to be eligible to enter public universities. Including private universities also allows higher variation in ethnicities, since Chinese and Indian Malaysians are more likely to attend private universities (Lee, 2017).

The primary tool for data collection is Network Canvas, a recently developed tool that allows researchers to develop custom surveys to capture network data and deploy them in the field (Birkett *et al.*, 2021). Network Canvas includes a suite of applications. I use the Architect application to develop the survey protocol three months before commencing fieldwork. The final survey design is decided after rounds of development, pilot testing with Malaysian

respondents, and incorporating feedback from pilot testers. The Interviewer application provides a clean and easily accessible interface for the survey respondents to engage with the survey as it visualises their friendship network. I install the Interviewer application on both my laptop and tablet, so that I may conduct two surveys simultaneously while on the field. The full survey in the Interviewer interface is included in Appendix 1. Finally, I utilise the Server application to securely store responses from both my laptop and tablet and to export them in comma-delimited format for further analysis.

The first section of the survey is the name generator phase in network data collection. At this stage, respondents are presented with a series of fifteen prompts which aid them in recalling the names of their friends. Since friendship network data is collected by allowing respondents to self-report a list of friends, there could be variations in how different individuals perceive friendship. To overcome this challenge, I deconstruct the notion of friendship into distinct prompts based on functional specificity, or by the functional roles that social connections often possess (Cutrona and Russell, 1990). The fifteen prompts in the survey are constructed based on situation-specific queries, with questions from four categories: work, personal life, leisure, and social support. These situation-specific prompts align with the part-whole hierarchical structure of friendships (Munck, 2020). Appendix 2 includes the full list of the fifteen questions in the name generator section of the survey.

The second section is the name interpreter stage, where I obtain further information on each friend that the respondent nominated. I ask the respondent about the gender and ethnicity of each friend, as well as how close the friend is to the respondent. The first two alter attributes are elicited from the point of view of the respondent, which may not necessarily correspond to the actual gender or ethnicity of the named friend. For example, a respondent may think that a

friend identifies as Malay when in reality the friend would identify as another ethnicity. The likelihood of misclassification is low due to the saliency of ethnic classifications in the Malaysian context. The presence of misclassification is also inconsequential to the research design since inferences are made on each respondent's friendship formation patterns, which are influenced by each respondent's perception of their friends' attributes, rather than the actual attributes of their friends.

The final section gathers ego-level data, which are information about the respondent, including their ethnicity, primary and secondary school type, university institution, gender, and year of birth. Respondents select from six different options for ethnicity, including the three major ethnicities in Malaysia, an option for East Malaysian, an option for being mixed, and an option for others. There are six options for primary school, including *Sekolah Kebangsaan* (translated as "National Primary School"), the two vernacular primary schools, Islamic primary school, and two additional options for international schools and others respectively. There are seven options for secondary school, including *Sekolah Menengah Kebangsaan* (SMK, translated as "National Secondary School"), Chinese Independent School (CIS), three Bumiputera-exclusive institutions (namely MRSM, SBP, and Religious Secondary School in Table 1), and two additional options for international schools and others respectively.

I utilise a non-probabilistic, quota-based sampling method when deciding which students to approach to enter the survey. Carrying my laptop and tablet, both with the Network Canvas Interviewer app installed, and a folder containing consent forms, I roamed between a few identified hotspots at the university campuses where students tend to spend their free time, such as discussion spaces, the cafeteria, and study spaces around walkways. I have set a quota of 200 students, of which half should be from public universities and half from private. During

the first few days of data collection at each university, I approach students in these hotspots randomly asking if they would like to participate in the survey. Once a student expresses interest, I provide them with the information sheet. When they agree to take part, I ask them to sign the consent form and pass them my laptop or tablet where they can interact with the Interviewer interface. I leave the respondent to complete the survey by following the instructions on the Interviewer application screen, while I make myself available in case the respondents have any questions. The decision to leave the respondent to complete the survey on their own is to minimise the risk of social desirability bias by eliminating a present interviewer during the survey administration (Kreuter, Presser and Tourangeau, 2008). I excluded students who have been in the university for less than a year from the survey as these students are less likely to have a crystallised friendship network. During the final day at each university, I observe the ethnic composition of respondents that I have collected so far. If any of the three major Malaysian ethnic groups seem to be underrepresented in the sample, I prioritise asking students of that ethnic group. Also, if any gender group seem to be underrepresented in the sample, I prioritise students of that gender.

A potential challenge to subsequent analysis is the lack of variation in alter ethnicity, as core networks tend to be homogenous (Cornwell, 2009). However, a sample of at least 25 alters for each respondent is sufficient to generate stable estimates of network measures such as those pertaining to its composition (McCarty *et al.*, 2007). Additionally, the task required of the respondents, to introspect and recall names from their network of friends in response to open-ended prompts, is one that is cognitively taxing (McCarty *et al.*, 2001). Hence, there is the risk of satisficing, where the cognitive burden leads respondents to exert minimal effort to provide accurate or thoughtful responses. To mitigate this challenge, I strategically place the name generator phase, which requires the most cognitive effect, in the first section of the survey.

Satisficing occurs when the utility from exerting cognitive is low (Matzat and Snijders, 2010). As an incentive to encourage maximum participation from the respondent, I provide a monetary compensation of RM10 (approximately £2) which is given to the respondents upon completing the survey.

V. Analysis

After removing duplicate alter names, the final sample consists of 203 egos and 2,182 alters; in other words, a total of 203 individuals responded to the survey and they collectively named 2,182 friends in total. Figure 1 contains a summary of ego characteristics. I collapse three secondary school categories, "SBP," "MRSM," and "Agama," into a new category called "Bumiputera exclusive institutions." These three types of secondary schools are typically only attended by Malays. In the sample, all students who reported that they went to SBP also reported identifying as Malay. Sekolah Menengah Kebangsaan Agama (coded "Agama") are Islamic religious schools. All students who reported that they went to Agama also reported identifying as Malay, except for one who identified as being East Malaysian. MRSM implements a 90 percent Bumiputera quota. All students who reported that they went to MRSM also reported identifying as Malay, excluding one Indian and one East Malaysian.

Figure 1: Summary of Respondent Characteristics

Several attributes are directly affected by the non-probabilistic, quota-based sampling method, including ethnicity, university, and gender. Towards the end of my fieldwork at each university, I ensure that I have met the quota for the number of students at each university. I also aim for gender balance and for the distribution of ethnicities in the sample to reflect the demographic distribution of the university. The respondents' year of study and year of birth are also partly influenced by sampling, since I only sample students who have been in university for more than one year, which means that most respondents are in their first year of university, and many are born between 2000 and 2003.

The largest ethnic group in the sample is Chinese, a result which could be influenced by the positionality of the researcher in the data collection process. It is possible that I find it

easier to approach ethnic Chinese students, and therefore I have tried to be as random as possible in my selection of respondents. The large group of Chinese respondents could also be because Chinese students are more likely to say yes to taking my survey because I am a Chinese Malaysian. However, in the distribution of ethnicities at each university in Figure 2, the proportions of ethnicities seem to reflect the demographic proportions at private and public universities, where Malays tend to attend public universities while Chinese and Indian students tend to attend private universities. This suggests that the risk of sampling bias arising from the data collection fieldwork is minimal.

Figure 2: Ethnic Diversity of Universities in Sample

Note: UM and UKM are public universities, while Sunway, Monash, and UNM are private universities.

One of the key assumptions made about education pathways seems to hold in the sample: there is indeed ethnic segregation into different levels of education in the sample. Figure 3 contains a Sankey diagram that illustrates the one-step education pathways. Ethnicity plays a large role in determining what type of school one goes to for primary education. Most Chinese go to SJKC, most Malays go to SK, and SJKT is exclusively attended by ethnic Indians. At the secondary level, CIS is almost exclusively attended by students who went to SJKC. To illustrate all the unique education pathways taken by each ethnic group, I generate a series of Sunburst diagrams in Figure 4.

Figure 3: Sankey Diagram of One-Step Education Pathways by Ethnicity

Figure 4: Sunburst Diagrams of Education Pathways by Ethnicity

Malays and Chinese, the two largest ethnic groups represented in the sample, have the most distinct education pathways. Most Chinese in the sample go to SJKC for primary school, while most Malays go to SK. For Malays, the typical pathways after SK are going to SMK or going to a Malay-exclusive institution. For Chinese students who went to SJKC, most proceed to SMK, while the rest go to either international schools or Chinese Independent Schools. None of the Chinese students following these pathways attend public universities. Half of the Indians in the sample went to SK, and the same students then proceeded to SMK. A significant portion of Indians go to SJKT, and from there proceed to SMK. In the sample, SMK, or national secondary school, seems to be a converging point for all ethnic groups, which suggests that SMK might have greater ethnic diversity than the other schools.

Figure 5 illustrates the ethnic diversity of all the different primary and secondary schools. Vernacular schools and Malay-exclusive schools are the least diverse, while SMK has the most diverse sample of students, conforming to expectations. Malays are overrepresented in SK, which might be because Chinese students attend SJKC and many Indian attend SJKT. Both primary and secondary international schools and others are also relatively diverse, but they make up a far smaller proportion of the full sample.

Figure 5: Distribution of Ethnicities of Primary and Secondary Schools in Sample

Moving past ego-level characteristics, I now present and comment on some descriptive statistics at the network level, starting with the degree distribution. The degree of a personal network is the number of friends (alters) the respondent (ego) nominates. Figure 6 plots the degree distribution for all 203 egos on a histogram. The degree distribution approximates a normal distribution but is slightly right-skewed by outliers who nominate many friends. On average, each respondent nominates 10.8 friends, while the median degree is 9. Figure 7 plots the degree distribution by ethnicity.

Figure 6: Degree Distribution of Personal Networks in Sample

Figure 7: Degree Distribution by Ethnicity

Using the alter data, it is possible to construct several measures of homophily for each personal network. The first such measure is the EI index (Krackhardt and Stern, 1988), which I define as the ratio of the number (E) of alters of a different ethnic group from ego minus the number (I) of alters of the same ethnic group as ego to the total number of nominated alters:

$$EI = \frac{E - I}{E + I} \qquad (Eq. 7)$$

This index measures the diversity a respondent's network between -1 and 1. A value of -1 indicates perfect homophily, where the respondent only nominates friends of their own ethnic

group, while a value of 1 indicates perfect heterophily, where the respondent only nominates friends that are not from their own ethnic group. A value of 0 indicates an equal number of friends of the same ethnic group and those of a different group.

Figure 8 plots the distribution of EI indices for respondents of each ethnicity. Most friendship networks of Malay and Chinese students are between -1 and -0.5, indicating high levels of homophily. Friendships of Indian students mostly have an EI index of around zero. Mixed students seem to cluster around values that indicate high degrees of heterophily, which could be explained by the fact that mixed students are few and far in between, so that few of the friends they nominated are mixed. These patterns of homophily seem to bear a resemblance to findings on baseline homophily in the United States, where larger demographic groups (i.e., Whites in the United States, Malays and Chinese in the sample) have more homogenous networks, while smaller ethnic groups have lower levels of homophily (Marsden, 1987).

Figure 8: Distribution of EI Index by Ethnicity (dotted line = mean)

Henceforth, I move from describing the data to conducting inferential analysis. I alter several variables in the data to make inferences to provide more meaningful inferences. Firstly, I subsume the category for East Malaysians into the "Other" category. In the context of peninsular Malaysia ethnic relations, East Malaysians are a minority and distinct from the three main ethnic groups. The presence or absence of East Malaysians have minimal influence on the dynamics of ethnic relations, and functionally they are comparable to an "other" category. This means that I can combine, without affecting the interpretation of the results, these two categories to overcome the challenge in statistical analysis that the number of respondents who identify as "Other" ethnicity is very small. I also collapse international schools for both the primary and secondary education levels into the "Other" category since they are not a meaningful category on their own and are functionally similar in nature to the schools in the "Other" category (i.e., private schools, home-schools, and others). Finally, I recode the university variable from a categorical variable with five unique values for each university in the sample to a binary variable that indicates either public or private university.

I perform inferential analysis on homophily utilising a metric based on the proportion of friends of each ethnicity in each personal network. I create five new ego-level variables for the proportion of friends who are Malay, Chinese, Indian, Mixed, and Others respectively. Figure 9 plots a heat map for the proportion of each alter ethnicity among all friends nominated summed across each ego ethnicity. The dark shades on the area denoting Chinese-Chinese and Malay-Malay indicate that more than 80 percent of friends that Chinese or Malay respondents collectively nominated are of the same ethnicity. The fact that the darkest shades of each column are along the diagonal line of the heatmap means that the greatest proportion of friends nominated are of the same ethnic group as the respondent, indicating the prevalence of ethnic homophily among all ethnicities, with mixed ethnicity being the exception.

Figure 9: Proportion of Alter Ethnicity Among All Alters Nominated by Ego Ethnicity

I model these proportions as the outcome variable in a quasi-binomial regression model with a logit link function by approximating them as counts of successes in many trials (McCullagh and Nelder, 1989). The explanatory variables included in the model are ethnicity, education pathway (i.e., primary school, secondary school, university), and language. Language is included as a proxy for degree of acculturation. For example, a Chinese student who speaks Chinese primarily and a Chinese student who speaks English primarily could have vastly different ethnic compositions in their friendship networks. I include language to control for such effects.

The results for the three main ethnic groups of interest, presented in Table 2, show that being of the same ethnicity is a statistically significant predictor of friendship for each of these groups. Holding education pathway and language constant, Malay respondents are more likely to have larger proportions of Malay alters, Chinese respondents are more likely to have larger

proportions of Chinese alters, and Indian respondents are more likely to have larger proportions of Indian alters. These results offer evidence to suggest that students of each of the three major ethnic groups exhibit homophilic tendencies in their friendship networks at the university level.

The university variable is statistically significant in the models with Malay and Chinese as the outcome variable. The coefficients suggest that respondents from public universities are more likely to nominate more Malay friends but less likely to nominate more Chinese friends. These findings may result from the larger number of Malay students in public universities and the relative underrepresentation in private universities in the sample. None of the primary school and secondary school variables are statistically significant, neither are any of the language variables.

Table 2: Predictors of Proportion of Friends of Malay, Chinese, or Indian Ethnicity

	Malay	Chinese	Indian
Intercept	−18.369 (1830.881)	−0.644 (0.884)	−3.084 (1.584)
Ethnicity (Ref: Mixed)			
Malay	1.784*** (0.43)	0.548 (0.496)	−0.006 (0.502)
Chinese	−0.551 (0.542)	1.893*** (0.402)	−0.869 (0.461)
Indian	0.18 (0.596)	0.122 (0.508)	1.226*** (0.373)
Other	−0.514 (0.524)	0.473 (0.542)	−0.616 (0.569)
Primary School (Ref: Other)			
SK	0.402 (0.428)	−0.606 (0.447)	0.017 (0.483)
SJKC	0.661 (0.566)	0.171 (0.421)	−0.646 (0.524)
SJKT	−0.579 (0.819)	−1.345 (1.028)	0.199 (0.629)
Islamic	16.569 (1420.266)	−14.374 (1007.679)	−14.1 (1253)
Secondary School (Ref: Other)			
SMK	−0.227 (0.356)	0.157 (0.356)	0.718 (0.442)
CIS	−0.446 (1.041)	0.015 (0.575)	1.133 (0.661)
Bumiputera-excl.	0.567 (0.434)	−1.122 (0.747)	0.669 (0.726)
Public University (Ref: Private)	1.447*** (0.278)	−0.701* (0.321)	−0.466 (0.339)
Language (Ref: Other)			
Malay	16.568 (1830.881)	−1.273 (0.907)	−1.581 (1.682)
Chinese	14.443 (1830.881)	0.882 (0.875)	0.281 (1.619)
Tamil	16.036 (1830.881)	−0.515 (1.088)	1.766 (1.594)
English	15.634 (1830.881)	−0.159 (0.801)	1.317 (1.541)
Dispersion Parameter	0.237	0.324	0.185
Residual Deviance	43.005	56.146	28.941
Degrees of Freedom	183	183	183

$*** p < 0.001,\ ** p < 0.01,\ * p < 0.05$
Standard errors in parentheses.

To test the key hypotheses of this paper and examine the factors that influence homophily, I utilise moderation analysis on exponential random graph models (ERGMs), which allows me to answer the question of how going to a particular type of school moderates the effect of homophily. As the implementation of moderation analysis on egocentric ERGMs is currently not supported, I first simulate a complete network using the parameters of the egocentric ERGM and then perform a regular ERGM on the simulated complete network. Moderation analysis is performed on the regular ERGM rather than the egocentric ERGM.

The results of the egocentric ERGMs are reported in Table 3. I run three different egocentric ERGM model specifications. The results from each run are comparable with statistically significant ethnic homophily across all ethnic groups, while none of the control variables are statistically significant. The results of the egocentric ERGM can be interpreted as follows: Holding university and language spoken constant, Ego ERGM 3 predicts that the odds of a friendship tie between two individuals are $e^{1.879} = 6.547$ times greater if both the ego and the alter are Malay, compared to when one is not Malay. The same interpretation holds for the ethnic homophily term for other ethnicities.

I proceed to simulate a complete network with Ego ERGM 3, which includes controls for both language and university when estimating the parameters for ethnic homophily. The result of the ERGM performed on the simulated network is presented in Table 4, along with the average marginal effects for each of the ERGM terms. The regular ERGM results are comparable in magnitude to that of the egocentric ERGMs, suggesting that the simulated network contains sufficient statistics of the egocentric data (Krivitsky, Handcock and Morris, 2011).

Table 3: Ego ERGM Results

Parameters	Ego ERGM 1	Ego ERGM 2	Ego ERGM 3
Edges	1.556^{***} (0.103)	1.735^{***} (0.194)	0.573 (1.454)
Ethnic Homophily			
Malay	1.880^{***} (0.188)	1.985^{***} (0.185)	1.879^{***} (0.264)
Chinese	1.775^{***} (0.188)	1.718^{***} (0.183)	1.927^{***} (0.305)
Indian	2.539^{***} (0.359)	2.509^{***} (0.322)	2.372^{***} (0.488)
Mixed	0.955^{**} (0.362)	0.909^{**} (0.343)	0.809^{*} (0.352)
Other	2.875^{***} (0.415)	2.947^{***} (0.492)	2.879^{***} (0.457)
Public University (Ref: Private)		−0.191 (0.176)	−0.128 (0.228)
Language (Ref: Other)			
Malay			0.579 (0.786)
Chinese			0.291 (0.749)
Tamil			0.641 (0.957)
English			0.625 (0.738)

Standard errors in parentheses. $*** p < 0.001$, $** p < 0.01$, $* p < 0.05$

Table 4: Results of ERGM based on Simulated Complete Network

Parameter	ERGM Model	AME
Edges	-3.719^{***} (0.139)	-0.176^{***} (0.007)
Ethnic Homophily		
Malay	1.865^{***} (0.107)	0.088^{***} (0.005)
Chinese	1.828^{***} (0.118)	0.087^{***} (0.006)
Indian	2.436^{***} (0.195)	0.115^{***} (0.009)
Mixed	0.821^{*} (0.332)	0.039^{*} (0.016)
Other	2.813^{***} (0.23)	0.133^{***} (0.011)
Primary School (Ref: Other)		
SK	0.143 (0.096)	0.007 (0.005)
SJKC	0.008 (0.091)	0.000 (0.004)
SJKT	0.271 (0.164)	0.013 (0.008)
Islamic	0.039 (0.174)	0.002 (0.008)
Secondary School (Ref: Other)		
SMK	-0.191^{**} (0.064)	-0.009^{**} (0.003)
CIS	0.035 (0.114)	0.002 (0.005)
Bumiputera-excl.	-0.017 (0.088)	-0.001 (0.004)

Standard errors in parentheses. ERGM is run on simulated network based on Ego ERGM 3.
$*** p < 0.001,\ ** p < 0.01,\ * p < 0.05$

Table 5 contains the results from the moderation analysis. $AME_x^{g=0}$ and $AME_x^{g=1}$ indicate the average marginal effect of the ethnic homophily terms when the moderating variable takes on the values of 0 and 1 respectively. ΔAME_x is the second difference, or the difference between $AME_x^{g=1}$ and $AME_x^{g=0}$, which is the key value of interest since it is interpreted as the moderating effect of each school type. N indicates the respondents of the ethnicity that has attended that school in the sample. Some of the combinations of ethnicity and primary or secondary school are rare in the sample, such as Chinese respondents who went to SK ($N = 4$) and Indian students who attend a Bumiputera exclusive secondary school ($N = 1$), and the results for these combinations should be interpreted with caution.

The second difference for SMK is negative and statistically significant across all three major ethnicities, which indicates that attending national secondary school is associated with a lower average marginal effect of ethnic homophily between each of the three major ethnicities respectively. In addition, there is also a statistically significant positive moderating effect of attending SK for the ethnic homophily terms for Malays and Chinese, indicating that attending a national primary school is associated with a higher average marginal effect of ethnic homophily between Malay students and between Chinese students.

Table 5: Results of Moderation Analysis

Moderating Variable	$AME_x^{g=0}$	$AME_x^{g=1}$	ΔAME_x	N
Main Effect: Malay Ethnic Homophily				
Primary: SK	0.080	0.089	0.009^{*} (0.004)	55
Primary: SJKC	0.088	0.088	0.001 (0.004)	1
Primary: Islamic	0.088	0.091	0.003 (0.008)	5
Secondary: SMK	0.103	0.090	-0.014^{***} (0.003)	28
Secondary: Bumiputera-excl.	0.089	0.088	-0.001 (0.004)	27
Main Effect: Chinese Ethnic Homophily				
Primary: SK	0.078	0.088	0.009^{*} (0.004)	4
Primary: SJKC	0.086	0.087	0.001 (0.004)	67
Secondary: SMK	0.101	0.087	-0.014^{***} (0.003)	44
Secondary: CIS	0.086	0.089	0.002 (0.005)	8
Main Effect: Indian Ethnic Homophily				
Primary: SK	0.114	0.127	0.013 (0.008)	11
Primary: SJKT	0.105	0.111	0.007 (0.004)	8
Secondary: SMK	0.135	0.124	-0.011^{**} (0.004)	17
Secondary: Bumiputera-excl.	0.116	0.115	-0.001 (0.004)	1

$*** p < 0.001,\ ** p < 0.01,\ * p < 0.05$
Delta standard errors for second difference in parentheses. Only combinations of ethnicity and primary/secondary with $N > 0$ are included.

VI. Conclusion

Four key findings emerge from the results of the analysis.

Firstly, friendships among Malaysian university students are characterised by ethnic homophily. Across all model specifications, a persistent result is that Malaysian university students tend to have friends that are of the same ethnic group as themselves. Being of the same ethnicity is a statistically significant predictor of having a large proportion of friends of a particular ethnicity. across all ethnic groups, two individuals who share a common ethnicity are statistically significantly more likely to become friends. This implies that ethnicity remains a highly salient attribute shaping social interactions in Malaysia and warrants an inquiry into its causes.

Secondly, the sample shows the choice of which primary education institution to attend is largely influenced by ethnicity, while the choice of secondary education institution is influenced by both primary education and ethnicity. At the primary level, Malays go to SK, Chinese go to SJKC, and Indians go to SJKT. The proportion of students in each educational pathway likens that described in (Raman and Sua, 2010). National secondary schools (SMK) are a key point of convergence for students of all three major ethnicities. A surprising feature of the sample is that a low number of students attend to Chinese Independent School. A few explanations are possible. Since Chinese Independent Schools mostly cater to Chinese students who struggle to follow the curriculum in national secondary schools which is taught in the Malay language (Yeoh, 2013), the relatively low number of students in the sample could be due to the fact that these students pursue tertiary education in Chinese-medium universities, which are not included in the sample. Another explanation is that the United Examinations Certificate, awarded to students who complete their secondary education at a Chinese

Independent School, is not recognised by public universities in Malaysia, thus Chinese students opt for SMK instead. The general pattern of ethnic segregation in the Malaysian education system provides a unique context to evaluate how primary and secondary schooling is associated with friendship formation patterns later in life.

Thirdly, the results indicate support for the hypothesis that students who attend national secondary school are more likely to have a more diverse friendship network. Among all ethnicities, having attended national secondary school is associated with lower levels of ethnic homophily at the university level. Attending national school is associated with lower levels of ethnic homophily at the university level, implying that intergroup contact increases people to have friends of other ethnic groups. The results are consistent with the predictions of the Contact Hypothesis. There are also several implications about the drivers of friendship formation behaviour. I corroborate the results from previous studies that national schools offer an opportunity to foster interethnic understandings (Miller and Brewer, 2013; Thijs and Verkuyten, 2014). The results also implies that long-term exposure to contact with other ethnicities is also able to alter the friendship formation patterns of individuals. Experience in primary and secondary school seems to have lasting effects on the friendship formation patterns later in life. I find evidence that higher levels of contact with other ethnicities through education in national secondary schools is associated with lower levels of ethnic homophily in university friendship networks. Experiences as far back as primary school can potentially affect the friendship formation patterns at the university level, as attending national primary school also has a statistically significant effect on university students' friendships.

Finally, there is no evidence to suggest that attending ethnic-exclusive education institutions affect friendship formation at the university level. The analysis offers some

empirical perspectives on the current political debate on vernacular schools and affirmation action in Malaysia. The operation of vernacular schools is a source of political controversy among Malaysians, with one side favouring the preservation of cultural heritage, while opponents claim that these schools are the cause of ethnic segregation and disharmony in Malaysia. I find no evidence to support the arguments of the critics of vernacular schools. Neither do I find evidence to suggest that attending Bumiputera-exclusive institutions magnify the effects of ethnic homophily at the university level.

Two key limitations underlying the conclusions of this paper. They offer potential avenues for future research.

The measure of homophily in this paper cannot be interpreted as choice homophily in social psychology, where is an individual preference for friends of the same ethnic group. Attributing that the homophilic pattern requires ruling out other possible explanations, such as closure mechanisms and the opportunity pool for meeting friends. Low levels of cross-group ties may induce transitive closure between individuals from different groups, affecting the level of homophily in the network (Quinn, *forthcoming*), while a homogenous friendship network may simply be due to the low numbers of certain ethnic groups at a university instead of inherent homophilic preferences. Both cannot be ruled out due to the lack of data on the underlying ethnic proportions of each ethnicity in the universities and the absence of alter-alter ties in the data. These questions could be addressed with complete network data or utilising novel advancements in homophily estimation using egocentric network data such as Genkin and Koshinen's (*forthcoming*) estimation of homophily using alter-alter ties and conceptualising egos as social foci for alters.

Additionally, the relationship between segregation and homophily is potentially spurious, where segregation can lead to patterns of homophily and homophily can lead to segregation. While I have attempted to tackle this spurious relationship by theorising that segregation in earlier levels of education can lead to homophily later in university. The interplay between the primary schooling and secondary schooling is also unexplored, and it remains an open question on whether homophilic tendencies at the primary school are moderated by experiences in secondary school. Future studies that can disentangle the level of ethnic homophily that are associated with primary and secondary schooling and other confounders such as neighbourhood segregation or segregation within universities would contribute to a more robust understanding of the drivers of the friendship formation.

Ethnic categorisation is rooted in Malaysia's colonial history and remain highly relevant as it has been weaponised in the current political arena (Gabriel, 2015). Some critics may argue that putting ethnicity in the spotlight in this study perpetuates ethnic categorisation and division; however, progress towards promoting interethnic harmony requires an acknowledgement of these realities rather than ignoring them. A nuanced understanding of the dynamics of friendship formation enables policies to be more effectively designed to promote greater harmony between different social groups. These findings are also relevant to multicultural contexts beyond Malaysia.

References

Abraham, C.E.R. (1983) 'Racial and ethnic manipulation in Colonial Malaya', *Ethnic and Racial Studies*, 6(1), pp. 18–32. Available at: https://doi.org/10.1080/01419870.1983.9993396.

Allport, G.W. (1954) *The Nature of Prejudice.* Cambridge, MA: Addison-Wesley.

Amrith, S. (2010) 'Indians Overseas? Governing Tamil Migration to Malaya 1870-1941', *Past & Present*, 208, pp. 231–261. Available at: https://doi.org/10.1093/pastj/gtq027.

Birkett, M. *et al.* (2021) 'Network Canvas: Key decisions in the design of an interviewer-assisted network data collection software suite', *Social Networks*, 66, pp. 114–124. Available at: https://doi.org/10.1016/j.socnet.2021.02.003.

Borgatti, S. and Halgin, D. (2011) 'On Network Theory', *Organization Science*, 22. Available at: https://doi.org/10.1007/978-1-4419-5513-5_2.

Burgess, S. and Platt, L. (2021) 'Inter-ethnic relations of teenagers in England's schools: the role of school and neighbourhood ethnic composition', *Journal of Ethnic and Migration Studies*, 47(9), pp. 2011–2038. Available at: https://doi.org/10.1080/1369183X.2020.1717937.

Collins, R. and Annett, J. (1975) *Conflict Sociology: Toward an Explanatory Science*. Academic Press.

Cornwell, B. (2009) 'GOOD HEALTH AND THE BRIDGING OF STRUCTURAL HOLES', *Social networks*, 31(1), pp. 92–103. Available at: https://doi.org/10.1016/j.socnet.2008.10.005.

Cutrona, C.E. and Russell, D.W. (1990) 'Type of social support and specific stress: Toward a theory of optimal matching', in *Social support: An interactional view*. Oxford, England: John Wiley & Sons (Wiley series on personality processes), pp. 319–366.

Department of Statistics Malaysia (2023) *Department of Statistics Malaysia*, *Quarterly Demographic Statistics* . Available at: https://dosm.gov.my/portal-main/release-content/0903ff04-8b7c-11ed-96a6-1866daa77ef9 (Accessed: 15 August 2023).

DiMaggio, P. and Garip, F. (2012) 'Network Effects and Social Inequality', *Annual Review of Sociology*, 38(1), pp. 93–118. Available at: https://doi.org/10.1146/annurev.soc.012809.102545.

Duxbury, S.W. (2023) 'The Problem of Scaling in Exponential Random Graph Models', *Sociological Methods & Research*, 52(2), pp. 764–802. Available at: https://doi.org/10.1177/0049124120986178.

Feld, S.L. (1981) 'The Focused Organization of Social Ties', *American Journal of Sociology*, 86(5), pp. 1015–1035.

Gabriel, S.P. (2015) 'The meaning of race in Malaysia: Colonial, post-colonial and possible new conjunctures', *Ethnicities*, 15(6), pp. 782–809. Available at: https://doi.org/10.1177/1468796815570347.

Gaertner, S. *et al.* (2000) 'Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation', *Group Dynamics: Theory, Research, and Practice*, 4, pp. 98–114. Available at: https://doi.org/10.1037/1089-2699.4.1.98.

Genkin, M. and Koshinen, J. (forthcoming) 'Homophily Estimation in Nationally Representative Egocentric Data Using an Alter-Based Framework'.

Hoetink, H. (2011) 'Resource Competition, Monopoly, and Socioracial Diversity', in *Resource Competition, Monopoly, and Socioracial Diversity*. De Gruyter Mouton, pp. 9–26. Available at: https://doi.org/10.1515/9783110898170.9.

Hunter, D.R. *et al.* (2008) 'ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks', *Journal of statistical software*, 24(3), p. nihpa54860. Available at: https://doi.org/10.18637/jss.v024.i03.

Kokkonen, A., Esaiasson, P. and Gilljam, M. (2015) 'Diverse Workplaces and Interethnic Friendship Formation—A Multilevel Comparison across 21 OECD Countries', *Journal of Ethnic and Migration Studies*, 41(2), pp. 284–305. Available at: https://doi.org/10.1080/1369183X.2014.902300.

Krackhardt, D. and Stern, R.N. (1988) 'Informal Networks and Organizational Crises: An Experimental Simulation', *Social Psychology Quarterly*, 51(2), pp. 123–140. Available at: https://doi.org/10.2307/2786835.

Kreuter, F., Presser, S. and Tourangeau, R. (2008) 'Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity', *Public Opinion Quarterly*, 72(5), pp. 847–865. Available at: https://doi.org/10.1093/poq/nfn063.

Krivitsky, P.N., Handcock, M.S. and Morris, M. (2011) 'Adjusting for network size and composition effects in exponential-family random graph models', *Statistical Methodology*, 8(4), pp. 319–339. Available at: https://doi.org/10.1016/j.stamet.2011.01.005.

Krivitsky, P.N. and Morris, M. (2017) 'Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US', *The Annals of Applied Statistics*, 11(1), pp. 427–455. Available at: https://doi.org/10.1214/16-AOAS1010.

Lee, H.-A. (2012) 'Affirmative Action in Malaysia: Education and Employment Outcomes since the 1990s', *Journal of Contemporary Asia*, 42(2), pp. 230–254. Available at: https://doi.org/10.1080/09500782.2012.668350.

Lee, H.-A. (2017) 'Majority Affirmative Action in Malaysia':, *Pluralism*.

Lee, S.M. (1989) 'Female Immigrants and Labor in Colonial Malaya: 1860–1947', *International Migration Review*, 23(2), pp. 309–331. Available at: https://doi.org/10.1177/019791838902300207.

Marsden, P.V. (1987) 'Core Discussion Networks of Americans', *American Sociological Review*, 52(1), pp. 122–131. Available at: https://doi.org/10.2307/2095397.

Marsden, P.V. (1988) 'Homogeneity in confiding relations', *Social Networks*, 10(1), pp. 57–76. Available at: https://doi.org/10.1016/0378-8733(88)90010-X.

Matzat, U. and Snijders, C. (2010) 'Does the online collection of ego-centered network data reduce data quality? An experimental comparison', *Social Networks*, 32(2), pp. 105–111. Available at: https://doi.org/10.1016/j.socnet.2009.08.002.

McCarty, C. *et al.* (2001) 'Comparing Two Methods for Estimating Network Size', *Human Organization*, 60. Available at: https://doi.org/10.17730/humo.60.1.efx5t9gjtgmga73y.

McCarty, C., Killworth, P.D. and Rennell, J. (2007) 'Impact of methods for reducing respondent burden on personal network structural measures', *Social Networks*, 29(2), pp. 300–315. Available at: https://doi.org/10.1016/j.socnet.2006.12.005.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. Boston, MA: Springer US. Available at: https://doi.org/10.1007/978-1-4899-3242-6.

McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001) 'Birds of a Feather: Homophily in Social Networks', *Annual Review of Sociology*, 27(1), pp. 415–444. Available at: https://doi.org/10.1146/annurev.soc.27.1.415.

Merlino, L.P., Steinhardt, M.F. and Wren-Lewis, L. (2019) 'More than Just Friends? School Peers and Adult Interracial Relationships', *Journal of Labor Economics* [Preprint]. Available at: https://doi.org/10.1086/702626.

Miller, N.S. and Brewer, M.B. (2013) *Groups in Contact: The Psychology of Desegregation*. Academic Press.

Munck, G.L. (2020) 'Conceptualization and Measurement: Basic Distinctions and Guidelines'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.3889800.

Noor, N.M. and Leong, C.-H. (2013) 'Multiculturalism in Malaysia and Singapore: Contesting models', *International Journal of Intercultural Relations*, 37(6), pp. 714–726. Available at: https://doi.org/10.1016/j.ijintrel.2013.09.009.

Omar, M.A. (2011) 'Asal-usul Orang Melayu: Menulis Semulasejaahnya (The Malay Origin: Rewrite Its History)', *Jurnal Melayu*, 7, pp. 1–82.

Paluck, E.L., Green, S.A. and Green, D.P. (2019) 'The contact hypothesis re-evaluated', *Behavioural Public Policy*, 3(2), pp. 129–158. Available at: https://doi.org/10.1017/bpp.2018.25.

Perry, B.L., Pescosolido, B.A. and Borgatti, S.P. (2018) *Egocentric Network Analysis: Foundations, Methods, and Models*. Cambridge University Press.

Pettigrew, T.F. and Tropp, L.R. (2006) 'A meta-analytic test of intergroup contact theory', *Journal of Personality and Social Psychology*, 90(5), pp. 751–783. Available at: https://doi.org/10.1037/0022-3514.90.5.751.

Quinn, J. (forthcoming) 'Manipulating Initial Network Characteristics and Measuring Changes in the Emergence of Homophily: A Computational Experiment.'

Raman, S.R. and Sua, T.Y. (2010) 'Ethnic segregation in Malaysia's education system: enrolment choices, preferential policies and desegregation', *Paedagogica Historica*, 46(1–2), pp. 117–131. Available at: https://doi.org/10.1080/00309230903528496.

Raykov, T. and Marcoulides, G.A. (2004) 'Using the Delta Method for Approximate Interval Estimation of Parameter Functions in SEM', *Structural Equation Modeling: A Multidisciplinary Journal*, 11(4), pp. 621–637. Available at: https://doi.org/10.1207/s15328007sem1104_7.

Selvadurai, S. *et al.* (2015) 'Debating education for nation building in Malaysia: National school persistence or vernacular school resistance?', p. 1.

Siah, P.C. *et al.* (2015) 'Perception on Chinese values: A comparison of Chinese secondary students studying at national secondary schools and Chinese independent schools in Malaysia', *The Social Science Journal*, 52(1), pp. 62–68. Available at: https://doi.org/10.1016/j.soscij.2014.08.006.

Smith, M.G. (1971) Institutional and Political Conditions of Pluralism. In L. Kuper and M.G. Smith (eds), *Pluralism in Africa*. Los Angeles, CA: University of California Press.

Tajfel, H. and Turner, J. (1979) *The Social Identity Theory of Intergroup Behavior*, *Political Psychology*. Psychology Press, pp. 276–293. Available at: https://doi.org/10.4324/9780203505984-16.

Tey, N., Halimah, A. and Singaravelloo, K. (2009) 'Ethnie interactions among students at the University of Malaya', *Malaysian Journal of Economic Studies*, 46, pp. 53–74.

Thijs, J. and Verkuyten, M. (2014) 'School ethnic diversity and students' interethnic relations', *The British Journal of Educational Psychology*, 84(Pt 1), pp. 1–21. Available at: https://doi.org/10.1111/bjep.12032.

Triandafyllidou, A. (1998) 'National identity and the "other"', *Ethnic and Racial Studies*, 21(4), pp. 593–612. Available at: https://doi.org/10.1080/014198798329784.

Yeoh, E. (2013) 'Socioeconomic Development and Ethno-Cultural Diversity: State Policy and the Evolvement of Pluralism in Malaysia', *International Journal of Business Anthropology*, 4(1), pp. 121–144.

Yeoh, E.K.-K. (2006) 'Ethnic coexistence in a pluralistic campus environment', *GeoJournal*, 66(3), pp. 223–241.

Appendix 1: Survey Interface on Network Canvas

Network Canvas Interviewer
Consent
Please indicate if you would like to proceed with the survey.
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and that I can withdraw from the study at any time, without having to give a reason.

Network Canvas Interviewer
Section 1 of 3: Please Read These Instructions Carefully!
In this section of the survey, you will receive 15 prompts to help you nominate friends that you **met in university**. You can nominate a friend by clicking the icon on the bottom right of the screen and typing their name. (You may use pseudonyms, initials, first names, or nicknames to ensure confidentiality and anonymity)
The first prompt is "Friend(s) you'd study in the library with". After entering the names of your friends **made in university** that come to mind, click on the "down arrow" button on the bottom left of the screen to continue.
Please **spend some time** after reading each prompt to think about who to nominate. You may skip a prompt by clicking on the "down arrow" button on the bottom left of the screen if you do not wish to nominate any friends for that prompt.
You may nominate as many friends as you like for each prompt, as long as you met them **after entering university**. You may also repeat the same friend's name for multiple prompts.

Network Canvas Interviewer
Friend(s) with whom you would study in the library together
Test
Type a name and press enter...

Network Canvas Interviewer
Please drag and drop each of your friend's names into the category that best describes your relationship with them.
Test
Very Close
Close
Not Very Close

Network Canvas Interviewer
Please drag and drop each of your friends' names into the category that reflects the ethnicity that you believe your friend best identifies with.
Test
Malay
Chinese
Indian
East Malaysian Ethnicities
Mixed
Other

Network Canvas Interviewer
Please drag and drop each of your friends' names to the category that best reflects the gender that you believe your friend most identifies with.
Test
Male
Female

Network Canvas Interviewer
Section 3 of 3: Some Questions About You
In this section, you will answer a few questions about yourself.
What ethnic group do you most identify with?
Malay
Chinese
Indian
East Malaysian Ethnicities
Mixed
Other
What kind of school did you attend for the majority of your primary education years?
Sekolah Kebangsaan
Sekolah Jenis Kebangasaan Cina (SJKC)
Sekolah Jenis Kebangsaan Tamil (SJKT)
Sekolah Agama (Islamic School)
International School

Appendix 2: Name Generator Questions in Section 1 of Survey

PROMPTS

1. Friend(s) with whom you would study in the library together
2. Friend(s) you would recommend to a job opening
3. Friend(s) with whom you would participate in volunteering projects
4. Friend(s) you consistently have meals with
5. Friend(s) whom you would want to be your best man / bridesmaid
6. Friend(s) whose birthday(s) you remember
7. Friend(s) with whom you share a similar hobby
8. Friend(s) you would go to the movies with
9. Friend(s) you would go shopping with
10. Friend(s) you would go to a concert with
11. Friend(s) you would travel to another country with
12. Friend(s) you would invite to your home
13. Friend(s) you would play boardgames or video games with
14. Friend(s) you would call when you're going through a difficult time
15. Friend(s) you would have deep conversations with

Appendix 3: Ego Information Captured in Section 4 of Survey

1. What ethnic group do you most identify with?
    a. Malay
    b. Chinese
    c. Indian
    d. East Malaysian Ethnicities
    e. Mixed
    f. Other

2. What kind of school did you attend for the majority of your primary education years?
    a. Sekolah Kebangsaan
    b. Sekolah Jenis Kebangsaan Cina (SJKC)
    c. Sekolah Jenis Kebangsaan Tamil (SJKT)
    d. Sekolah Agama (Islamic School)
    e. International School
    f. Private School / Home School / Others

3. Did you attend any other type of school for primary education? If you did, please list their type and the length of years you attended.

    [Optional with open ended answer field]

4. What kind of school did you attend for the majority of your secondary education years?
    a. Sekolah Menengah Kebangsaan (SMK)
    b. Chinese Independent School
    c. Sekolah Menengah Kebangsaan Agama (SMKA)
    d. Maktab Rendah Sains MARA (MRSM)
    e. International School
    f. Sekolah Berasrama Penuh (SBP)
    g. Vocational College / Technical Institute
    h. Private School / Home School / Others

5. Did you attend any other type of school for secondary education? If you did, please list their type and the length of years you attended.

    [Optional with open ended answer field]

6. Which university do you currently study at?
    a. Universiti Malaya
    b. Universiti Kebangsaan Malaysia
    c. Monash University
    d. Sunway University
    e. University of Nottingham Malaysia

7. What is your current year of study at university?
    a. Foundation / Pre-University
    b. Year 1
    c. Year 2
    d. Year 3
    e. Year 4
    f. Masters
    g. PhD

8. In what year were you born?

    Date in [YYYY] format option available for selection.

9. Which gender do you most identify with?
    a. Male
    b. Female

10. In what language are you most fluent in? (Please only select one)
    a. English
    b. Malay
    c. Chinese
    d. Tamil
    e. Other

Appendix 4: Markov Chain Monte Carlo (MCMC) and Goodness of Fit (GOF) Diagnostics

MCMC Diagnostics of Ego ERGM 1

GOF Diagnostics of Ego ERGM 1

MCMC Diagnostics of Ego ERGM 2

GOF Diagnostics of Ego ERGM 2

MCMC Diagnostics of Ego ERGM 3

MCMC Diagnostics of Ego ERGM 3 (continued)

GOF Diagnostics of Ego ERGM 3