The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

A dissertation presented by

Susan Godlonton

to

The Commerce Graduate Committee

in partial fulfillment of the requirements

for the degree of

Masters of Commerce

in the subject of

Economics

University of Cape Town

University of Cape Town
Cape Town, South Africa
April 2008

Advisor: Dr. Justine Burns

Title: Social Networks and Academic Achievement: A study of a South African University

Social Networks and Academic Achievement – The Case of a South African University

Susan Godlonton

University of Cape Town

Abstract

This paper examines the role of study partnerships in determining academic success in undergraduate economics examinations. The study uses data from a South African University – the University of Cape Town. The data is unique as it has information on actual study partner linkages which are used to construct the peer networks unlike most other studies that use proxied networks. A key finding of this paper is that the "quality" of one's peer network (as measured by the average academic performance of the individual's study network) is the most important (of those considered) characteristic of one's study network that improves individuals' performance. The impact observed is sizeable and non-linear – it is particularly beneficial to poorer performing students to interact in study partner networks that perform better on average relative to better performing students. Other characteristics of the study partner networks considered include "quantity" measures of the study network as well as the racial composition. While the "quantity" measure does positively impact individual performance, racial diversity negatively impacts performance particularly for poorer performing students. However, the impact of these alternative network measures is very small in magnitude.

1 Introduction

Thirteen years into the new dispensation and South Africa continues to face huge educational disparities which are translated into considerable labour market inequalities particularly between Black and White individuals (Moll (1996), Keswell and Poswell (2002)). Due to increasing returns to education (Keswell and Poswell (2002), Erichsen and Wakeford (2001), and Moll (1996)) in South Africa, it is important to understand channels of influence that assist students to progress through the education system. One key channel of influence which has largely been ignored is the role of peers. Understanding the nature, interaction, and influence of peer effects on educational outcomes is important as it may provide insights to potential institutional structures that can propagate academic success particularly for those from disadvantaged backgrounds.

Driving the increasing returns to education in South Africa is the lack of skilled workers and an oversupply of unskilled workers[1] (Burger and Woolard (2005)). Due to this mismatch in the demand and supply for skilled labour, graduates not only receive higher returns but they are more likely to find and maintain higher paid employment than non-graduates. Moleke (2005) finds that 59.5 percent of all graduates find jobs immediately, while an additional 28.4 percent found employment with 6 months of completing their studies. Bhorat (2006a) shows that while tertiary graduate unemployment (for youth aged 15 – 24) is quite high in South Africa at 17 percent, youth unemployment of matriculants[2] in the same age cohort is 66 percent. Clearly, obtaining a tertiary education is very important in determining wages and employment, therefore understanding factors determining success in tertiary education is important.

The current South African literature has investigated the impact of a number of individual-specific controls on tertiary performance in Economics. Males are found to perform consistently better than females (Edwards (2000), Smith (2004), Van Walbeek (2004) and Parker (2006)), particularly in

1 Increasingly this group of unskilled workers that are unemployable constitute individuals with a complete secondary schooling (matric – 12 years of completed schooling) and even tertiary graduates that did not receive the educational inputs demanded by the labour market. The rising trend in graduate unemployment is largely attributable to: (i) type of qualification obtained; and (ii) "quality" of the institution at which qualifications were obtained. While an increasing trend in graduate unemployment is of great concern, graduates still face much brighter employment prospects and as such improving the throughput of individuals is important (Bhorat (2006a) and (2006b)).

2 A matriculant is someone who has completed secondary schooling.

University of Cape Town

answering Multiple Choice Questions (Van Walbeek (2004)). Mathematical ability ensures greater success, as does previous academic performance at school (Edwards (2000), van Walbeek (2004) and Parker (2006)). General student motivation as measured by lecture attendance (Smith (2004), Van Walbeek (2004)) or the number of hours studied (Parker (2006)) are also significant determinants of the success of the student.

This literature has examined tertiary success by focusing on individual study decisions in isolation, ignoring the potential influence that social networks, constituted by peers or study partners, might have on these individual decisions. A growing body of literature suggests that networks matter for economic and social outcomes, since they allow for complex social interactions to take place between individuals, facilitating information spillovers and learning between network members (Banerjee(1992), Bikhchandani, Hirshleifer and Welch (1992)). While parental background is no doubt important, it has been found that non-kinship ties are of particular relevance in urbanized areas and for more educated individuals for a range of outcomes including education (Fischer (1982))[3]. Evidently understanding the nature of peer effects is important as such interactions may be a key input in the tertiary education production function. In particular the existence of such effects may create the opportunity for socially enhancing interventions in the higher education system in South Africa.

University of Cape Town

A recurring difficulty in analysing peer effects is the ability to distinguish them from correlated effects arising due to the self-selection of one's peers. Sacerdote (2000) and Zimmerman (1999) employ the ideal empirical strategy to identify peer effects. Both studies employ a quasi-experimental strategy that utilizes the random assignment of students to roommates in Dartmouth and Williams College respectively.

Zimmerman (1999) uses the random assignment (in terms of academic ability) of students to roommates to test for aggregate peer effects at the dormitory room-level. His findings show that students in the middle of the distribution (based on SAT scores) perform worse when paired with a low

[3] Studies have found that kinship ties only account for 50 percent of reported personal relations, emphasizing the importance of peers. (Wellman (1979) and (1982), Fischer (1982)) In the South African context, Schoer (2005) found using the Khayelitsha-Mitchell's Plain data that 49 percent of all successful job methods were through networks external to the household.

performer (SAT score in 15th percentile). He also finds that top students appear to be least affected by their peers (as defined as roommates).

Sacerdote (2000) investigates the impact of roommates in determining academic effort as well as the decision on whether a student selects to join a fraternity at Dartmouth College. He finds that a one standard deviation increase in ones' roommates' GPA is associated with a 0.05 increase in one's own GPA. He does not find evidence of a peer effect (where peer is defined as roommate) on the choice of major, but finds a peer effect (where peer is defined either as a roommate or as someone on the same corridor as the individual) in fraternity membership. A sample restricted to only roommate pairs showed that 27 percent of the pairs joined the same fraternity[4].

A weakness with many existing studies (including these quasi-experimental approaches of Zimmerman (1999) and Sacerdote (2000)) that examine the role of peer effects is the estimation of peer effects based on proxied networks. This arises due to the lack of explicit data on the actual contacts with which the individual interacts. Most studies trying to explore peer effects define an individual's network using a selection of observable characteristics. Two common group defining criteria are geographical[5] proximity (Glaeser et al. (1996) and Bertrand et al. (1999)) and race (Loury (1977))[6]. A notable problem in defining the network in this manner is the poor definition of "membership"; in particular it is unclear whether the proxied network is an accurate representation of the true peer effects.

A notable exception in the Economics literature is that of Conley and Udry (2005). They employ both explicit (informational neighbours) and approximated contacts (geographical neighbours) in their analysis of the uptake of technology in the farming of pineapples in Ghana. Therefore, in this study they do not have difficulty defining peers. They find that approximated networks underestimated the underlying social interactions. This highlights the need to consider the actual rather than proxied

4 The null hypothesis associated with this test that there is no peer effect would result in only 5 percent of the peers joining the same fraternity.
5 Geographical proximity can be interpreted in a broad sense. That is, peers are defined either as those living in the same area, same residence room/corridor, attend the same school. In all cases while students may be close to this identified peer network positionally, that is not sufficient for the existence of influence.
6 Notable exceptions to this are the economic impact of disease transmission (Morris, 1996 and 1997) and Mental health (Ueno, 2004).

University of Cape Town

networks in order to determine the role of peers in a social setting. Specifically, they find that inexperienced farmers will increase their fertilizer input by approximately 4 cedis[7] (geographical neighbours) or 6 cedis (informational neighbours) per plant when there is a one standard deviation increase in their constructed index that measures the fertilizer use by farmer neighbours that is associated with good news (the learning effect from peers). While the Conley and Udry (2005) analysis does a significantly better job than most economic studies in identifying actual contacts, the respondents were presented with a list of 7 randomly selected individuals in the village (that were also part of the study) and asked to identify which of them they had approached for advice for their farm. So while the Conley and Udry (2005) study captures actual networks it is a subset of the actual interactions. In this paper all interactions will be studied and will be discussed in Section 2.

Measuring the extent to which social networks affect decision making is challenging because social group formation is usually endogenous, complicating causal interpretation: that is, if belonging to a social group is a matter of deliberate choice, it is difficult to assign causality to the impact of the group itself (Manski 1993). In addition, individuals may make simultaneous contemporary decisions making it difficult to determine the causal behavior; this is often called the reflection problem. This study will adopt a fairly standard approach to measuring peer effects in that it will regress own outcomes (measure of academic success) on peer outcomes to determine the impact. As outlined by Manski (1993) peer effects determined in this manner are difficult to interpret. The first issue, and perhaps the most difficult to resolve, is that individuals self-select into groups. Secondly, it is difficult to identify the causal effect if two individuals affect each other simultaneously as outcomes will be jointly determined. The third and final issue is that it is difficult to separate out contextual and endogenous effects. Contextual effects pertain to the effect of one student's background characteristics on the respondents' outcome. Endogenous effects arise due to similar behavioural patterns of students – for example, students may jointly determine their study patterns due to similar underlying preferences of hard work or their value of social interaction. The explicit strategies undertaken to resolve these interpretation issues for the purposes of this analysis will be discussed more in depth in Section 2.

University of Cape Town

[7] Cedis is the currency of Ghana and therefore this measure represents the increase in per-plant value of fertilizer.

This paper seeks to offer two key contributions to the existing literature. First, it will provide an analysis of the role of explicit social networks (instead of proxied networks) in determining academic success. Secondly, it extends the existing literature on higher education in South Africa by accounting for the role of social interactions as an input in the education production function.

Section 2 will discuss the sample design and the data used in this analysis. The structural properties of the global study partner network, as well as individual level network characteristics are discussed in Section 3. Section 4 presents a discussion of the main findings, and Section 5 concludes.

2 Study Design and Data

In this study, the entire Economics 110 undergraduate class (of 2005) at the University of Cape Town was selected as the population group of interest. All students registered for this class were requested to complete a survey questionnaire that elicited information regarding their study partnership networks. This data was matched to administrative data from the University which contains demographic information on all the students in the class (including those that opted out of completing the questionnaire). The data used in this analysis is drawn from these two sources: (i) The University of Cape Town's Heritage Database; and the (ii) self-administered questionnaire.

Unfortunately, while adopting a quasi-experimental approach similar to that of Zimmerman (1999) and Sacerdote (2000) would have been ideal it is not possible in the context of this study as students are not randomly assigned study partners with which they are encouraged/ forced to interact with. Students are compelled to attend tutorial groups but the assignment of students to such groups is not administered in a random manner. Due to class schedules tutorial groups are highly segregated in terms of streams, which are highly correlated with academic ability. For example, there are tutorial groups comprised of mostly Actuarial Science students and others of Politics, Philosophy and Economics students. There are considerably different admission requirements for these two particular streams and therefore tutorial groups are highly clustered according to academic ability.

University of Cape Town

2.1 University of Cape Town's - Heritage system data

The University of Cape Town's Heritage system provides individual-level demographic information on the entire first year Economics 110 class. It provides accurate school level information as well as all information related to the final marks for each course taken by a student. In addition - tutorial, test and exam information is recorded by the course convenor and also used in the analysis. This data enables validation of the information collected on the survey questionnaires. It also enables one to input demographic and test performance information of the respondents study partners'. As it includes all student exam performance measures since entering the University it provides multiple academic performance measures measured at different points in time. This is particularly useful as many studies depend on outcome measures of peers determined at the same point in time.

This is the first modification to the standard approach of estimating peer effects that tries to resolve the reflection problem. Rather than regressing the impact of individual academic success on peer outcomes measured at the same point in time this data enables one to regress exam performance of an individual on test performance measures of his peers. This ensures that the estimated effect can be interpreted as a causal effect, as it does not face the problem that the two outcomes are determined jointly.

University of Cape Town

2.2 Self-administered questionnaire

2.2.1 Design:

The survey was administered by the tutors (advanced economics students registered and employed by the University of Cape Town) during tutorial periods that were compulsory. The administration of the questionnaire was constructed to be incentive compatible for the tutors to ensure that they encouraged and gave students adequate time to complete the questionnaire thoroughly during the tutorial. Tutors were paid per questionnaire. Two different rates were paid – a higher rate for completed questionnaires and a somewhat lower rate for incomplete questionnaires. The fact that the tutorials were compulsory (and attendance was monitored carefully) is critical. Had the questionnaire been administered during the lecture (which is not compulsory) a biased sample of students would have resulted.

The key focus of the questionnaire was to collect peer information to complement the administrative data available from the Heritage data set. Students were asked to identify those students in the Economics 110 class that they considered to be study partners. The bulk of the social network information obtained was from the question: "Please tell us about all your friends and associates that also take Economics 110. Please indicate all friends, individuals who you study with and those that you ask for help with your economics studies." (Refer to Appendix A for the complete questionnaire.)

University of Cape Town

2.2.2 Sampling:

A key consideration before conducting the questionnaire was to address the issue of sampling. This is particularly complicated when attempting to use regression analysis in combination with social network analysis due to conflicting sampling concerns. Random sampling of the underlying network of peers is flawed because the possibility that two individuals randomly selected are connected is negligible (Scott, 1991). Social network theorists have identified three alternative methods to sample such networks: (i) snow-ball sampling; (ii) ego-centric sampling; and (iii) socio-centric sampling.

Snow-ball sampling is the process whereby a group of individuals is selected and interviewed. These respondents identify their social network – in this context – who they study with. The nominations made by respondents are then followed up and interviewed. This iterative procedure continues for a fixed number of repetitions. A major drawback to implementing this process is that the initial sample is unlikely to be a representative sample of the underlying network as in practice a non-random initial sample is selected (and is preferable) to ensure greater connectedness among the individuals.

Ego-centric sampling is the most common approach. This method selects a core group of individuals of interest. Information will be gathered through these respondents about their peers but these peers will not be interviewed themselves. This results in an incomplete network as not all peer linkages are identified. Figure 1 highlights how a partial network arises, and the fundamental problem with this method for capturing intermediary linkages. Two key groups X and Y are connected by individuals A and B. If individuals C and D (and no other individual in groups X and Y) are selected to be

interviewed then the link (that between A and B) connecting these two groups will not be identified in the data. If the role of these intermediary linkages is important, as shown in the work by Granovetter (1983) in the context of job search, then this method will not enable one to correctly estimate the impact of one's peers.

The last approach, socio-centric sampling, requires identifying a population of interest and interviewing all members in the population. In this case both the characteristics of the global network and the indirect linkages between individuals can be accurately captured. It is for these reasons that a socio-centric approach was undertaken in this study.

University of Cape Town

2.2.3 Population of Interest

The population selected was that of the first year Economics 110 class at the University of Cape Town in 2005 and constitutes a total of 1525 individuals. The first year class was selected to reduce concerns about the endogeneity of friendship formation and academic performance. This was the second measure taken to resolve the reflection problem as discussed by Manski (1993). It is well-known that individuals exhibit homophillic[8] sorting behaviour in peer selection (Moody, 2001). If this type of sorting occurs with respect to academic performance, for example, that well-performing students associate with other well-performing students, and poor-performing students associate with other poor-performing students – then a positive effect of the performance of one's network on one's own performance is not necessarily indicative of a peer effect. In this case, finding a positive coefficient on one's study partner's academic performance indicates a correlation between their academic ability.

By selecting the first year class and conducting the survey at the beginning of the year (only four months into the year) the likelihood that the study suffers from selection bias is minimized.[9]. The four month period ensured that the students had sufficient time to form study partnerships but the

8 That is, individuals sort into groups with individuals that share similar characteristics.
9 As noted in the previous section, ideally one would like to exploit some quasi-experimental strategy that exploits random peer associations, similar to that used by Sacerdote (2000) and Zimmerman (2001). Unfortunately as mentioned this is not possible in the current context. Even if such a randomized strategy were possible in the current context if we were to adopt this instead then it might not adequately capture the study networks relied upon by the individuals.

academic performance of their peers in Economics 110 would not yet have been revealed. This is because none of the Economic 110 results for an exam had been received by the students at this point in the semester. The median for the period of time knowing their reported peers is 28 weeks (7 months). The mean is 38 weeks but there is considerable variation in the length of time study partners have known each other with a standard deviation of 31.65 weeks. In fact 73.39 percent of all study partner linkages had not existed for longer than a year.

3 Sample Descriptive Statistics and Network Properties

Of a class size of 1525, there were 893 students that completed the questionnaire resulting in a response rate of 58.5 percent[10]. It should be noted that an additional 5.6 percent did not go on to write the examination as they either deregistered / decided not to write the final exam or did not meet the minimum requirements to sit the exam. It is not a perfect socio-centric approach as we do not have information on the linkages of the entire sample. However, it does capture a considerable portion of the underlying network.

Despite non-random sampling the resulting sample of students that completed the questionnaire is a very good representation of the underlying population along a range of observable characteristics (Refer to Table 1). If one considers the two largest racial groups the population comprises 28.6 percent Black and 44.8 percent White students compared to the sample which comprises 26.8 percent Black and 45.8 percent White students. The majority of the population is male (60.9 percent) and this is consistent in the sample although the proportion of males is 2 percent lower (58.9 percent) (Table 1). The representative nature of the sample is further supported by Figures 2a and 2b which illustrate that the distribution of both current and past academic performance is similar when comparing the sample to the population.

Table 1 also presents the descriptive statistics of individual and parental inputs of economic

10 While this response rate may be lower than ideal, it is not substantially lower than other studies that collect this type of information – specifically the identity of social connections. The Add-Health study (which is used by Moody (2001), Bearman, Moody and Stovel (2002) and Ueno (2004)) had a response of 75 percent for the in-school component, part of which included the module on friendships.

University of Cape Town

production function. The sample is comprised mostly of first year students (89.3 percent), who are male (58.9 percent), and are registered in the Commerce Faculty (61.1 percent). The respondents are on average 19.4 years of age. The majority of students (66 percent) report that they speak English in the home[11]. Schools writing the IEB (Independent Examination Board) constitute private and well-off public (or semi-public) schools - 17.6 percent of the sample attended such schools. Interestingly, despite the rule that tutorials were compulsory for the course, students only attended on average 91.4 percent of the time – in general they missed one tutorial. The average parental education of registered students is high, with approximately 14 – 15 years[12] of schooling. There is little variation in the years of schooling of the parents of the respondents and therefore dummies were created for the presence or absence of parents holding a degree.

University of Cape Town

3.1 Social Network Descriptive Statistics – Network Attributes

As this analysis considers the impact of various network measures it is important to understand the underlying structure and features of the study partner networks in the Economics 110 class. Various network properties will be examined and discussed in this section and comparisons to other studies with similar data will be made.

Network Attribute 1: Vertices

A vertex can represent individuals or organisations and is the fundamental unit of any network study. It is also sometimes referred to as an actor or a node in the network literature. In this study, a vertex represents a student from the Economics 110 class that completed a questionnaire and provided information on their study partners. Table 2 shows that there are 868 vertices in this study. While the size of this network pales in comparison to studies conducted using email contacts (Rows 5 and 6 of Table 2) or co-authorship networks (Rows 3 and 4 of Table 2), it is sizeable relative to studies in which network data is collected using a questionnaire (Rows 2 and 7 of Table 2).

11 This percentage is computed on the basis that if a student speaks multiple language in the home of which English is one of those mentioned, the student is considered to speak English at home.
12 This is equivalent to the parent completing Matric (grade 12 – final year of secondary schooling in the South African system) as well as obtaining a dilpoma / further certification.

Network Attribute 2: Edges (Directed and Undirected)

An edge connects two vertices – it represents the existence of a study partnership between two students in this study. There are two types of edges: (i) directed; and (ii) undirected edges. A directed edge has an origin and an end point – for example, in this study it is known who nominated whom as a study partner. Therefore, edges are directed as the beginning point is the respondent of the questionnaire and the end point is the study partner nominated. Table 2 shows that there are 2627 directed edges. That is, there are 2627 observed study partner linkages in this analysis.

Undirected edges simply represent a connection between two individuals without capturing information about the direction. For example, when studying sexual networks (Bearman et al, 2002) it is difficult to determine the direction of the connection between two individuals who have engaged in sex, in such a case edges will be undirected.

It is possible to convert directed edges to undirected edges (by ignoring the direction and simply considering the connection made) as is done by Conley and Udry (2005)[13]. This would enable one to increase the sample size of those students for which the study network is known. However, there are a number of problems in doing this. Firstly, one loses the informational advantages of knowing the direction. This is important in this analysis as this may lead to the construction of incorrect study partner links. While Student A may seek help from Student B and therefore consider Student B a study partner, Student B might not seek help from Student A. In this case, if Student A completed the questionnaire and Student B did not then by creating a link to Student B from Student A would be a misrepresentation of Student A's actual study network.

A second issue one faces if one uses the undirected edges for the basis of analysis is that it is only possible to generate partial networks for those who did not complete the questionnaire. For example, if a student studies with some students who completed the questionnaire and others who did not, the network that would be constructed for an individual who did not complete the questionnaire would only

[13] Conley and Udry (2005) use a sample roster of 7 randomly selected individuals and ask farmers whether they share advice with any of the individuals on the roster. They establish an information linkage if either farmer mentions that they do share farming advice.

University of Cape Town

include individuals who did complete the questionnaire. In this case, the analysis would constitute partial network information for some of the sample and complete network information for others. It is for these reasons that the analysis will only considered the directed edges in the analysis.

Network Attribute 3: Degree (Indegree and Outdegree)

Mean degree refers to the average number of edges connected to a vertex. In this study, this refers to the average number of study partners that a student reports. As this study can identify the direction of the edges (of the nomination) there are 2 types of degree measures – outdegree and indegree. Outdegree refers to the number of study partners reported by a respondent. Indegree refers to the number of contacts who nominated the respondent as a study partner. In this case one can not use the indegree measure as it will be a noisy indicator[14] as not all students in the Economics 110 class completed the questionnaire. The outdegree measure will however be accurate as all results are based only on the group who reported who their study partners are. This is reported in Table 2 and is 3.89. This means that each respondent reported on average 3.89 study partners.

An outdegree of 3.89 is small compared to a number of other studies – although larger than that found in sexual network studies (0.871 and 1.66 – Table 2). This is to be expected as individuals are likely to have more study partners than sexual partners. It is considerably smaller than the mean degree in co-authorship studies – which again is to be expected as such networks are likely to be smaller and more cohesive.

Network Attribute 4: Geodesic Distance

The geodesic distance measures the average shortest distance between any two individuals in the network. The last column of Table 2 compares this measure across the different studies. In this study, this measure is 9.517. Therefore, in the Economics class, select any two students in the class – say, Student A and Student B. If Student A was given some information pertaining to the exam and passed it on, then this information would on average need to be passed along through 9.517 other students in

[14] This measure will be noisy as it would only characterize a partial description of the network. It faces similar problems to that discussed with the undirected edges. That is, not only is this measure only a partial representation of the indegree measure for individuals, but the linkages also make less intuitive sense if one believes that study partnerships are in some cases unidirectional.

University of Cape Town

the class before reaching Student B (assuming the information travelled along the shortest trajectory).

The geodesic distance in this study of 9.517 is larger than all the other reported studies (except for the sexual network case – which is 16.01). It is not however considerably larger than the Ueno (2004) adolescent friendship network which reports a geodesic distance of 7.3. The geodesic distance reported in this study is probably an overestimate of the actual measure of the network, as not all students completed the questionnaire the network is incomplete. As the network is incomplete there are "holes" in the network that do not exist in reality – these unobserved connections could link individuals along shorter paths reducing the mean geodesic distance. However, this is a global network measure (i.e. it can not be computed for each student but only for the entire network) and therefore is not used in the regression analysis, so any biased estimate of this measure should not be of concern in interpreting the results that follow.

Network Attribute 5: Betweenness – Measuring the informational linkages of the individual

The measure used is Freeman's betweenness measure[15] (Freeman, 1977 and 1979). It measures the extent to which the individual is well connected to other individuals through both direct and indirect linkages. The more geodesic paths an individual lies on the more connected he/she is to the rest of the network. Intuitively, the betweenness proportion for individual i is a sum of probabilities (where the probability measures the likelihood that individual i lies on a geodesic between two other individual's – namely j and k) divided by the maximum number of pairs of nodes (individuals) in the network.

University of Cape Town

---

15 The standardized betweenness measure is computed using the formula (Freeman 1977 presented in Wasserman and Faust, 1994): $$C_B(n_i) = \left(\frac{\sum_{j<k} g_{jk}(n_i)}{g_{jk}} \right) \Big/ \left(\frac{(g-1)(g-2)}{2} \right)$$

where:

g_{jk} = the number of geodesics linking 2 individuals j and k;

i = a distinct individual required for the flow of information between individuals j and k;

$g_{jk}(n_i)$ = the number of geodesics connecting individuals j and k using i;

$\left(\frac{(g-1)(g-2)}{2} \right)$ = the maximum number of pairs of individuals in the underlying network.

Intuitively, the betweenness proportion for individual i is a sum of probabilities (where the probability measures the likelihood that i lies on a geodesic between two other individual's – namely j and k) divided by the maximum number of pairs of nodes (individual's) in the network.

It captures the intermediary role that individuals play in the study networks as they facilitate and enable information flows between otherwise unconnected individuals. If for example, we are to track the flow of information from student i to student j, and it is noted that student k is situated along the geodesic path of this flow of information, then student k is likely to have a higher betweenness measure due to this intermediary role. The importance of indirect linkages has been recognized in the social network literature for many decades but only in the late 1970s was it quantified (Wasserman and Faust, 1994). Granovetter's (1973, 1982) research played a fundamental role in highlighting the importance of indirect linkages in his analysis of the flow of information in job search networks. In this study Freeman's betweenness measure is reported in Table 3 – Panel A and is 0.372. This means that the probability that a student in the sample lies on a shortest path connecting any 2 students is 37.2 percent.

Network Attribute 6: Homophillic Sorting Behaviour

As discussed in the Introduction it is commonly know that individuals sort themselves according to criteria – that is on average we see evidence of homophillic sorting behaviour in social networks. In order to be aware of potential self-selection biases it is useful to understand the precise nature of the sorting behaviour in the study partner networks being examined. Three particular attributes will be considered in depth: (i) Nationality; (ii) Race; and (iii) Gender.

Nationality Dimension – South African vs Foreign

One type of distinction along which students may sort is nationality. Unfortunately, due to data constraints a very crude measure of this is used and this analysis will only distinguish between South Africans and Foreign nationals. One method to look at the extent of sorting is to construct the Ethno-linguistic Fractionalization[16] index, using nationality (or alternative individual attributes) to measure the extent of associations across groups. This index has frequently been adopted by economists exploring ethnic heterogeneity (Easterly and Levine (1997), and Collier (1998) and (1999)). Using this measure

[16] The ethno-linguistic fractionaliztion index is computed using the following formula:

$$ELF = 1 - \sum_{i=1}^{n} [\frac{n_i}{N}] \times [\frac{n_i - 1}{N - 1}]$$

where n_i represents the number of students of the ith race/nationality group in the study network of an individual, and N is the total number of members of the students' study network.

University of Cape Town

Table 3 – Panel A shows that there is a 15 percent probability that 2 randomly selected students from an individual's study network are of different nationalities.

A second method is to look at the proportion of study partners that share the same characteristic as the respondent. As Table 3 – Panel B shows foreign nationals are more likely to associate with South Africans (56 percent)), than South African are with foreign nationals (8.2 percent). This is not too surprising as there are more South Africans (81.4 percent) in the data and therefore both groups should associate more with South Africans. However, what is perhaps more interesting is that the proportions do not reflect the underlying nationality composition of the sample. That is, taking into consideration the nationality composition South African still associate with other South African more and Foreigners associate with South Africans considerably more.

Lastly, an alternative way to investigate the nature of these sorting attributes is to construct an *assortativity coefficient* (Newman, 2002 and 2003a). The assortative coefficient is defined as:

$$r = \frac{\sum_i e_{ii} - \sum_{ijk} e_{ik} e_{jk}}{1 - \sum_{ijk} e_{ik} e_{jk}}$$

where: i, j and k represent the different types of individuals; and e_{ij} as the fraction of connections between individuals of type i and type j. Similarly, e_{jk} represents the proportion of ties between individuals of type j and k. (Newman, 2002)

For example, in computing the assortativity coefficient for the South African citizenship-non South African citizenship type, there are 2084 study partnerships reported by South Africans to other South Africans, 186 nominations by South Africans reporting foreign study partners. Foreigners have 157 South African, and 201 Foreign study partners. Therefore, as there are a total of 2628 study partners e_{SS} is 0.79, e_{FF} is 0.08, e_{SF} is 0.07 and e_{FS} is 0.06. Using these figures one can apply the formula given to calculate the coefficient:

University of Cape Town

$$r = \frac{\left[\left(e_{SS} + e_{FF}\right) - \left(e_{SS}^2 + e_{FS}e_{SS} + e_{SF}e_{FS} + e_{SS}e_{SF} + e_{FF}e_{FS} + e_{FS}e_{SF} + e_{SF}e_{FF} + e_{FF}^2\right)\right]}{\left[1 - \left(e_{SS}^2 + e_{FS}e_{SS} + e_{SF}e_{FS} + e_{SS}e_{SF} + e_{FF}e_{FS} + e_{FS}e_{SF} + e_{SF}e_{FF} + e_{FF}^2\right)\right]} = 0.46$$

The closer the value of this coefficient is to 1, the greater the extent of homophillic association with respect to the defined "type" exists. If it equals one then perfect assortative mixing is present (all individuals only associate with others similar to themselves). In this case, the coefficient of 0.46 does not suggest that the network is highly segregated by nationality.

Racial Dimension

The racial dimension appears to be the most important of the three attributes considered in terms of its role in sorting behaviour. This is supported by evidence in three different ways across Panels A – C of Table 3. Table 3 – Panel A reports the Ethno-linguistic Fractionalization index and indicates that there is a 46.4 percent probability that a randomly selected student and his/her study partner are of different races.

Panel B of Table 3 shows that 84.6 percent of all study partners nominated by black students are black, while the proportion of white study partners of white students is 86.8 percent. As is evident from Table 3 – Panel B, Black and White students exhibit the greatest degree of within race study partners. This is an expected finding as these students have greater opportunity to study with others like themselves in the Economics 110 class due to the sheer size of both of these population groups in the class.

Lastly, Table 3 – Panel C indicates very high assortativity coefficients particularly when the sample is restricted to only Black and White respondents the measure is 0.82 which is close to 1, which would indicate perfect homophillic sorting. The overall sample has a race assortativity coefficient of 0.65 which is still high and indicative of homophillic sorting.

Gender Dimension

Panel B of Table 3 shows that men study with mostly men (80.3 percent of a male respondent's study partners are on average male), whereas women study mostly with other men (on average 67.3

University of Cape Town

percent of female respondents' study partners are male). Men constitute 58.9 percent of the sample, as such these figures still indicate that most respondents prefer to study in groups where men are in the majority. The assortativity coefficient is 0.48 and similar to that found for the nationality dimension. It does not indicate very high homophillic sorting in the gender dimension in the study networks (Table 3 – Panel C).

4 Findings and Discussion

As noted the empirical approach undertaken to measure the impact of the peer effect is to regress own outcomes (exam performance) on various network measures. The general form of the regression specification is:

$$\ln(Exam)_i = \alpha + \beta' X_i + \delta' Z_i + \varepsilon_i$$

where:

X_i = A vector of individual background characteristics including: race, gender, citizenship, age, parental education, and previous academic ability.

Z_i = A vector of the study partner network attributes of the individual. This vector differs across specifications depending on the attribute of the network examined. Due to the expected non-linear nature of the peer effects, many specifications include appropriate interaction terms which interact the study partner network measures with the students' own academic ability.

University of Cape Town

4.1 "Quantity" vs. "Quality" of Study Partnerships

One interesting aspect of the importance of study partners is to consider the relative value of the "quantity" of study partners relative to the "quality" of the respondents study partners (as measured by academic achievement). Measures that are used to quantify the "quantity" effect of study partners include: Outdegree, Indegree, and Freeman's Betweenness measure. The first two variables measure direct contacts, whereas the third measures indirect contacts – that is the access the student has to the broader network through their contacts. As noted earlier, it is important to include interaction terms of the various "quantity" and "quality" measures with own test performance in the specifications as the impact of one's network is likely to exhibit non-linearities as the effects vary considerably depending

on one's own ability.

4.1.1 "Quantity" Network Measures

Table 4 presents the results that analyze the relative importance of various "quantity" measures of an individual's study network. All of the network measures used are constructed based on the information provided by the respondents as to who their study partners were in the Economics 110 class. Therefore, all measures are based on true study partner relations. Columns 1 and 2 capture the impact of direct contacts; Column 3 the effect of indirect contacts; and Column 4 controls for both indirect and direct contacts.

Result 1: Access to a large indirect network is more important than a large immediate network (particularly for poorer performing students).

From Table 4 - Column 1, it is evident that having more direct study partners (that is having reported many study partners) has a small positive impact on an individual's academic performance. For students obtaining the mean grade on the first test (4.086 in logged terms) the effect of an additional study partner is an improvement of 0.009 percent[17]. While this effect is small in magnitude it is statistically significant[18]. This effect is equivalent to a 10 percent increase in a student's own tutorial attendance[19]. The network effect is small relative to the impact of a students own test 1 performance – for example, a 1 percent increase in a student's test 1 performance predicts a 0.636 percent[20] increase on the final exams for students with 3 study partners.

The results presented in Column 2 – Table 4 indicate that being reported as a study partner by many Economics 110 students has no statistically significant effect[21]. As discussed in Section 3, we should expect that the true indegree measure is a noisier measure of the "quantity" network measure. Recall

[17] $E[\ln(Exam)_i \mid Outdegree_i = x] - E[\ln(Exam)_i \mid Outdegree_i = (x-1)] = \beta_1 + \beta_2 \times Test_i \times \Delta Outdegree_i = 0.111 - 0.025 \times 4.086 \times 1 = 0.009$

[18] Controlling for a quadratic term for outdegree does not affect the results in any significant manner, and is not itself statistically significant or sizeable.

[19] It should be noted, that students attend 91.4 percent of tutorials on average. Thus there is not much room for tutorial attendance to be an important determinant in magnitude. The considerably high tutorial attendance is reflective of the fact that the tutorials were in fact mandatory. However, through the course of the semester missing one tutorial would not have led a student becoming disqualified to sit the exam and this is likely to have been well understood by the students.

[20] $\frac{\partial \ln(Exam)_i}{\partial Test1_i} = \beta_{Test1} + \beta_{Test1*Outdegree} \times Outdegree = 0.711 - 0.025 \times 3 = 0.636$

[21] Controlling for a quadratic term for indegree does not affect the results in any significant manner, and is not itself statistically significant or sizeable.

University of Cape Town

that this measure is constructed such that it sums the number of time the individual was nominated by a fellow student. It is noisier as not everyone completed the questionnaire and is therefore only a partial measure of the true indegree, measure and captures study partner linkages that may be unidirectional.

The results presented in Table 4 - Column 3 shows that indirect contacts are more important in determining academic success (to the extent we believe the causality) than the number of direct contacts. The measured coefficient indicate that two students (say – Students A and B) similar in every respect except their study network will have different academic outcomes. In the case where both students obtained the mean grade on the first test; and Student A is connected indirectly through their study partners to 10 percent more of the class relative to Student B then the academic performance of Student A will be 0.007 percent[22] greater than that of Student B.

The positive impact of interconnectedness is less important for students who perform better on test 1. For example, students obtaining 75 percent[23] on the first test would only improve by 0.001 percent[24] for every additional 10 percent access to the class. Students at risk of failing, for example, a student who scored 40 percent on the first test would gain 0.0164 percent[25]. This finding – that the positive impact of interconnectedness persists when controlling for the number of direct study partners (Table 4 – Column 4) albeit somewhat smaller in magnitude (0.013, 0.005 and 0.0009 logged exam point improvement for students achieving 40 percent, 59.5 percent - Mean grade on Test 1, and 75 percent respectively). Intuitively, the role of indirect linkages in determining academic success is plausible as students with informational linkages to more students may be better resourced in that they may have acquired additional past papers, advice from older student who had previously taken the course, or incurred knowledge spillovers from better run tutorial groups.

University of Cape Town

[22] $E[\ln(Exam)_i | Btweeness = x + 0.1, OwnTest1 = 4.086] - E[\ln(Exam)_i | Btweeness = x, OwnTest1 = 4.086] = (0.1) \times (1.068 - 0.245 \times 4.086) = 0.0067$

[23] Obtaining 75 percent is equivalent to achieving a distinction at the University.

[24] $E[\ln(Exam)_i | Btweeness = x + 0.1, OwnTest1 = 4.317] - E[\ln(Exam)_i | Btweeness = x, OwnTest1 = 4.317] = (0.1) \times (1.068 - 0.245 \times 4.317) = 0.001$

[25] $E[\ln(Exam)_i | Btweeness = x + 0.1, OwnTest1 = 3.688] - E[\ln(Exam)_i | Btweeness = x, OwnTest1 = 3.688] = (0.1) \times (1.068 - 0.245 \times 3.688) = 0.0164$

4.1.2 "Quality" Network Measures

Table 5, presents results incorporating "quality" measures of the study networks of the first year students. Measures employed to control for the "quality" aspect of these study networks include: (i) The average test 1 performance (logged) of the individual's study network; (ii) the standard deviation of the (logged) test 1 scores of the individual's study network. Regressions also control for a "quantity" network measure (interconnectedness) and interaction effects of these two "quality" measures with each other and the individual's own performance[26].

Result 2: Good academic performance by one's study partners improves own academic outcomes

Unsurprisingly Column 1 of Table 5 shows that an individuals' exam performance is positively associated with good performing study partner networks. A 1 percent increase in the average performance of a students' study partner network is associated with a 0.140 percent improvement in the individuals own performance, controlling for the "quantity" effect of the network (Table 5, Column 1). This effect is sizeable and significant but considerably smaller than the coefficient on own test performance – a 1 percent increase in an individual's test 1 performance is associated with a 0.575 percent higher outcome by the student on the final exam (Table 5, Column 1).

Result (2a): Good academic performance by one's study partners is particularly beneficial for students at the lower end of the distribution

The negative coefficient on the interaction term suggests that there is a non-linear effect such that those at the lower end of the distribution benefit most from study with better-performing peers. The impact of the average performance of one's study partners is decreasing in own test 1 performance. The interaction effect is only significant at the 10.8 percent significance level. However, it is jointly significant with own test performance and the networks' average performance. These results indicate that for students on the margin (own test 1 performance of 50 percent), a 1 percent increase in the average performance of their peers would transmit into a 0.23 percent increase in the individual's own

[26] Regressions are not sensitive to the choice of the selected "quantity" network measure used. Due to collinearity issues, only one measure is used, and in the regressions presented here include the betweenness measure.

University of Cape Town

final exam performance[27] (Table 5, Column 2).

In fact, only for students achieving higher than 82 percent on their first test is there a negative impact from studying with smarter peers. The negative impact is also negligible as even for the highest performing student a 1 percent increase in the performance of his/her peers' results in only a 0.09 percent decrease in the individual's own outcome[28](Table 5, Column 2).

Result (2b): The magnitude of this "quality" network effect dominates the "quantity" network effect

This peer "quality" network effect – that is the positive association of peer performance with individual performance is considerably greater than the "quantity" effects discussed in Section 4.1.1. One way of comparing the different magnitudes is to consider the network change required to induce a student to obtain a 1 percecnt higher grade on the final exam. A change of 1 percent is too large to be accomplished using a change in the quantitative nature of the network – i.e. even if a student is indirectly connected to every other student in the class, the change in the network structure will not be sufficient to enable a 1 percent higher grade. However, by exposing a student to a network that performs 4.19 percent[29] better than his/her current network the student will incur the desired 1 percent increase (if the individual obtained 50 percent on the first test). Clearly, the "quality" of one's study network is more important than the "quantity" in determining successful academic outcomes.

Result (2c): Variation in the performance of one's study partners has a negligible effect on individual's academic performance

The second measure to proxy for the "quality" of the student's study network that was used is the standard deviation of their peers' marks on the first test. In the simplest specification (Column 3, Table

University of Cape Town

27 $\frac{\partial \ln(Exam)_i}{\partial \ln(SP's AvgPerf)} = \beta_{\ln(SPAvgPerf)} + \beta_{Test1*\ln(SPAvgPerf)} \times Test1 = 2.08 - 0.472 \times 3.912 = 0.23$

28 The maximum logged test score is 4.60517 in the sample, therefore in this case the marginal effect of a 1 percent increase in the average performance of the individual's study network results in a 0.09 percent decrease in the individual's own performance on the final exam (Table5, Column 2).
$\frac{\partial \ln(Exam)_i}{\partial \ln(SP's AvgPerf)} = \beta_{\ln(SPAvgPerf)} + \beta_{Test1*\ln(SPAvgPerf)} \times Test1 = 2.08 - 0.472 \times 4.605 = -0.09$

29 $1 = (\Delta AvgSP's) \times (2.08 - 3.902 \times 0.472) \Leftrightarrow \frac{1}{0.238} = (\Delta AvgSP's) \Leftrightarrow (\Delta AvgSP's) = 4.19$

5), having a homogenous study group in terms of peer's academic performance has a negligible impact (both in magnitude and statistical significance) on own performance.

When accounting for interaction effects with individual's performance, and with average peer performance, the negligible impact of variation in peers' performance persists. In comparing two individuals with similar individual and parental attributes, both with study partners who on average achieved 60 percent, where student A has 2 study partners (one who obtained 30 percent on test 1, the other 90 percent) and student B also has 2 friends (both obtaining 60 percent on test 1) student B, both students perform equally as well.

4.1.3 Robustness Checks of "Quality" Network Effects:

Results (2a) – (2c) persist when controlling for possible interaction effects of the two "quality" network measures and the individuals' own test 1 performance (Columns 2 – 6 of Table 5). In some specifications the significance of the average peer performance becomes insignificant – however, due to the presence of the interaction terms in such specifications this is likely to be driven by the presence of collinearity. A joint test of the "quality" network measures across specifications in Table 5 consistently rejects (at least at the 10 percent significance level) the null that they are jointly insignificant. Due to these findings, Column 2 of Table 5 is used as the baseline regression for later specifications.

The results (2a) – (2c) also persist when conducting quantile regressions, the results of which are presented in Table 6. The estimated effect of the average of one's peers is almost twice as important for students at the 25^{th} percentile (students improve by 0.17 percent per 1 percent increase in their peers average performance) relative to those at the 75^{th} percentile (students improve by 0.089 percent per 1 percent increase in their peers average performance). More generally Table 6 shows that there is a monotonically increasing relationship between individual's performance and the performance of one's peers, mirroring result 2a.

Result 2b – that the magnitude of the "quality" effect dominates the "quantity" network effect is also

University of Cape Town

evident in Table 6. The coefficient on the betweenness measure is dwarfed by that on the average performance of the study work especially at the 75th percentile regression. As noted earlier, across the percentiles the standard deviation of one's study network is not a significant factor in determining one's own outcome – as such Result 2c is reiterated.

4.2 Composition of Networks

Table 7 explores the impact of the composition of the study network on academic performance, with particular focus on the racial composition[30]. To explore the role of diversity in race the Ethno-linguistic Fractionalization formula was adopted as discussed in Section 3. Recall, this measures the probability that two randomly drawn study partners from a respondent's study network are of different races[31], therefore the higher the index is to 1 the more diverse the study network is.

Result (3a): Racial diversity in study networks is a less important factor than the average performance of peers in producing better outcomes

Table 7 – Column 1 shows that increasing racial diversity within study networks negatively impacts the individual's own performance. This result is statistically significant and is robust to the inclusion of interaction effects (Table 7, Columns 2 – 7). This may not be too surprising due to factors such as culture and/or language barriers or differential learning strategies that make the transfer of knowledge across races more difficult. As Column 2 of Table 7 indicates this effect is partially offset for students who perform better, however, this effect is not completely offset for the highest performing student[32].

Results 2a) – 2c) indicated that the "quality" of the study partner network matters and should be accounted for in the racial composition analysis too. Accounting for this the finding of decreasing returns to racially diverse networks persists (Table 7 – Columns 2 – 7). For example, take a student on

30 Nationality is interpreted very narrowly for the purposes of this discussion referring only to the distinction between South Africans and non-South Africans.

31 The herfindahl concentration index is computed using the following formula:

$$H = 1 - \sum_{i=1}^{n} [\frac{n_i}{N}] \times [\frac{n_i - 1}{N - 1}]$$

where n_i represents the number of students of the ith race/nationality group in the study network of an individual, and N is the total number of members of the students' study network.

32 $\frac{\partial \ln(Exam)}{\partial Herfindahl} = -0.120 + 0.022 \times OwnTest1 = 0 \Leftrightarrow OwnTest1 = 5.455$
The maximum logged score is 4.60517.

University of Cape Town

the margin obtaining 50percent on test 1, if the racial composition of their study partner network is modified such that there is a 50 percent higher probability that 2 individuals selected from the network are of different races, then the student is predicted to perform 0.017 percent[33] worse on the final exam. Therefore, while there does appear to be some negative returns to diversity the size of this impact is small in magnitude.

Result (3b): Racially diverse study networks negatively impact students at the low end of the academic distribution

Although the negative impact is small it is most severe for poorer performing students (Table 7, Column 2, 4 – 7) and for students with study partner networks that do not perform well academically (Table 7, Column 3, 7). This last finding – that racially diverse networks are particularly ineffective for students with low average performing networks - is not robust to the inclusion of other interactions with the academic performance measures.

Table 8 presents the quantile regressions that account for the racial composition of the network. These results support the finding that the negative impact is largest for the poor performing students – most notably the negative coefficient on the racial diversity index (measured by the herfindahl index) is only significant in the 25th percentile regression. The coefficient of -0.172 is also considerably larger in magnitude than that observed for the median and 75th percentile regression – -0.032 and -0.067 respectively.

The predicted negative association of student performance with racially diverse networks remains very small even in the 25th percentile regression in which all coefficients related to the diversity index are highly statistically significant. For example, if there are two students who obtained 50 percent on test 1 and are in all ways alike except the racial composition of their network, where one student (Student A) has a study partner network such that there is a 50 percent higher probability that two individuals selected from the network are of different races relative to the other student (Student B), then Student

[33] $E[\ln(Exam) \mid Herfindahl = x + 0.5, OwnTest1 = 3.912] - E[\ln(Exam) \mid Herfindahl = x, OwnTest1 = 3.912] = -0.12(0.5) + (0.022)(0.5)(3.912) = 0.170$

University of Cape Town

A is predicted to perform 0.012 percent[34] worse on the final exam.

5 Conclusion

This analysis presented evidence that study partner networks do positively influence academic achievement in a tertiary institution in South Africa and that the effects are moderate. Using the unique data collected of actual study partnerships rather than proxied relations this paper showed that the average academic performance of a students' network is particularly important in determining individual academic success. This effect is non-linear, as it is decreasing in one's own prior academic performance. Thus, study partner networks can play a particularly important role for poorer performing students.

The quantity of peers; interconnectedness to the class; and racial composition also influence individual outcomes, but they all are significantly less important factors (in magnitude) in determining success. The variation in academic performance of a student's study network does not appear to matter. The findings suggest that the influence of peers is driven primarily by performance of peers rather than the size of one's network or the racial composition of the network.

These findings are more modest than anticipated. They suggest that substantially more work needs to be conducted on understanding the role of peers in this setting. In particular, more research needs to be conducted to understand how the formation of these study networks occur, if the underlying process is better understood it could be informative for potential institutional changes.

The findings do suggest that a role for institutional interventions may exist. For example, if low ability students could be paired up with high ability students the throughput of students might be improved. While University policies cannot dictate the formation of study partnerships, the institution can influence the structure of the tutorial groups and encourage study networks within these groups. Tutorial groups could be designed in a more effective manner to ensure better outcomes particularly

[34] $E[\ln(Exam) \mid Herfindahl = x + 0.5, OwnTest1 = 3.912, SPAvg = 3.912] - E[\ln(Exam) \mid Herfindahl = x, OwnTest1 = 3.912. SPAvg = 3.912]$
$= (0.5)[(-0.172) + (-0.339 \times 3.912) + (0.377 \times 3.912)] = -0.0117$

University of Cape Town

for poorer performing students. However, before dramatic interventions can be appropriately designed and implemented a thorough cost-benefit analysis needs to be conducted.

The current study is subject to issues regarding endogenous friendship formation and there may be some cause for concern about the causality of these results despite the actions taken to reduce these concerns. Future work could therefore draw on these findings and design a randomized experiment in which students are randomly assigned to a treatment and comparison group. Students assigned to the treatment group could be grouped into tutorial groups such that low ability students are paired with high ability students. Students in the comparison group could be randomly assigned to tutorial groups. In adopting this experiment, this and potentially other "treatments" (such as tutorial group size and composition) could be tested to see whether the tutorial system can be adapted in a manner that supports students (particularly at the lower end of the distribution) and ensure higher throughput rates. In so doing, this would offer more conclusive evidence on whether institutional changes can impact these social networks to induce positive learning effects.

University of Cape Town

6 References

Add-Health (2003): "Add Health: The national longitudinal study of adolescent health – Study Design", [Online] Available at: http://www.cpc.unc.edu/projects/addhealth/design.

Banerjee, A. (1992): "A Simple Model of Herd Behavior", *Quarterly Journal of Economics, Vol. 107(3), pp. 797 – 818.*

Bearman, P., Moody, J., and Stovel, K. (2004): "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks", *American Journal of Sociology, Vol. 110(1), pp. 44-91.*

Bertrand, M., Luttmer, E., and Mullainathan, S. (2000): "Network Effects and Welfare Cultures", *Quarterly Journal of Economics, Vol. 116 (1), pp. 351 – 377.*

Bhorat, H. (2006a): "Youth unemployment in post-Apartheid South Africa", presented at Harold Wolpe Memorial Trust open dialogue, SA Jewish Museum, Cape Town.

Bhorat, H. (2006b): "Graduate Unemployment in Post-Apartheid South Africa: Nature and Possible Policy Responses", Research Report Compiled for Standard Bank and Business Leadership South Africa.

Bikhchandani, S; D. Hirshleifer and I.Welch (1992): "A Theory of Fads, Fashion, Custom and Cultural Change in Informational Cascades", *Journal of Political Economy, Vol. 100(5), pp. 992-1026*

Burger, R.L. and Woolard, I. (2005): "The state of the labour market in South Africa after the first decade of democracy". *Journal of Vocational Education and Training, Vol. 57(4), pp. 453-476.*

Conley, T. and Udry, C. (2005): "Learning about a New Technology: Pineapple in Ghana", *American Journal of Agricultural Economics Vol. 83 Issue 3, pp.668-673.*

Collier, P. (1998): "The Political Economy of Ethnicity", in Annual World Bank Conference on Development Economics, Washington DC

Collier, P. (1999): "Ethnicity, Politics and Economic Performance", *Economics and Politics, Vol. 12(3), pp. 225 – 245.*

Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N. (2001): "Size-dependent degree distribution of a scale-free growing network", Physical Review E: Statistical, Nonlinear and Soft Matter Physics, *Vol. 63(6).*

Easterly, W. and Levine, R. (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions", *Quarterly Journal of Economics, Vol. 112(4), pp. 1203 – 1250.*

Edwards, L (2000): "An Econometric Evaluation of Academic Development Programmes in Economics", *South African Journal of Economics, Vol. 68(3), pp. 455-483*

Erichsen, G. and Wakeford, J. (2001): "Racial Discrimination in South Africa Before and After the First Democratic Election", *DPRU Working Papers, No. 01:49.*

Fischer, C. (1982): "To dwell among friends", *Chicago: University of Chicago Press.*

Freeman, L.C. (1977): "A set of measures of centrality based on betweenness", *Sociometry. Vol. 40, pp. 35 – 41.*

Freeman, L.C. (1979): "Centrality in Social Networks: Conceptual Clarification", *Social Networks, Vol. 1(3), pp. 215 – 239.*

University of Cape Town

Glaeser, E.L., Sacerdote, B. And Sceinkman, J.A. (1996): "Crime and Social Interactions", *Quarterly Journal of Economics, Vol. 111(2), pp. 507 – 548.*

Granovetter, M. (1973): "The Strength of Weak Ties", *American Journal of Sociology, Vol. 78(6), pp. 1360 - 1380.*

Granovetter, M. (1982): "The Strength of Weak Ties: A Network Theory Revisited", *Sociological Theory, Vol. 1, 1983, pp 201-233.*

Keswell, M. and Poswell, L. (2002): "How important is education for getting ahead in South Africa?", *Centre for Social Science Research Working Paper, No. 22.*

Manski, C. (1993): "Identification of Endogenous Social Effects: the Reflection Problem", *Review of Economic Studies, Vol. 60(3), pp.531 - 542.*

Mlatsheni, C. and Rospabe, S. (1999): "Why is youth unemployment so high an unequally spread in South Africa", Unpublished mimeo.

Moleke, P. (2003): "Employment experiences of graduates", Employment and Economic Policy Research Report, Pretoria: Human Science Research Council.

Moll, P. (1996): "The Collapse of Primary Schooling Returns in South Africa, 1960 – 1990", *Oxford Bulletin of Economics and Statistics, Vol. 58 pp. 185 – 209.*

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

Morris, M, et al. (1996): "Bridge populations in the spread of HIV/AIDS in Thailand" *AIDS, Vol. 10(11), pp. 1265 – 1271.*

Morris, M, et al. (1997): "Adolescent sexual networking and HIV transmission in rural Uganda." *Health Transition Review, Vol. 7, pp 89 – 100.*

Newman, M.E.J., Forrest, S. and Balthrop, K. (2002): "Email networks and the spread of computer viruses, Physical Review E: Statistical, Nonlinear and Soft Matter Physics, Vol. 66.

Newman, M. E. J (2001): "Coauthorship Networks and Patterns of Scientific Collaboration", *Proceedings of the National Academy of Sciences of the United States of America (PNAS), Vol. 101(1), pp. 5200-5205.*

Newman, M. E. J. (2002): "Assortative mixing in networks", *Physical Review E: Statistical, Nonlinear and Soft Matter Physics, Vol. 89.*

Newman, M. E. J. (2003a): "Mixing patterns in networks", *Physical Review E: Statistical, nonlinear and soft matter physics, Vol. 67.*

Newman, M. E. J. (2003b): "The structure and function of complex networks", *SIAM Review (45): pp 167-256.*

Newman, M. E. J. and Park, J. (2003): "Why social networks are different from other types of networks", *Physical Review E: Statistical, Nonlinear and Soft Matter Physics, Vol. 68.*

Parker, K. (2006): "The Effect of Student Characteristics on Achievement in Introductory Microeconomics in South Africa", *South African Journal of Economics, Vol. 74(1).*

Sacerdote, B. (2000): "Peer Effects with Random Assignment", *Quarterly Journal of Economics, Vol. 116(2) pp.681 – 704.*

University of Cape Town

Schoer, V. (2005) "Job search strategies and social networks: Evidence from the Khayelitsha/ Mitchell's Plain Survey", *Masters Dissertation, University of Cape Town.*

Scott, J. (2000) "Social Network Analysis: A Handbook", *London: Sage Publications.*

Smith, L. (2004): "A multivariate evaluation of mainstream and academic development courses in first-year microeconomics at the University of Cape Town: A comparative study", *Doctoral dissertation, University of Cape Town.*

Ueno, K. (2004): "Friendship Integration and Adolescent Mental Health". *Doctoral dissertation, Vanderbilt University.*

Van Walbeek, C. (2004): "Does Lecture Attendance Matter? Some Observations from a First-Year Economics Course at the University Of Cape Town". *South African Journal of Economics, 72(4): 861-883*

Wasserman, S. and Faust, K. (1994): "Social Network Analysis: Methods and Applications", *Cambridge: Cambridge University Press.*

Wellman, B. (1979): "The Community Question: The Intimate Networks of East Yorkers", *American Journal of Sociology, Vol. 84.*

Wellman, B. (1982): "Studying Personal Communities", in Marsden and Lin (eds) Social structure and Network Analysis. *Beverly Hills, CA: Sage Publications.*

Zimmerman, D. (2002): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *Review of Economics and Statistics, Vol. 85(1), pp. 9 – 23.*

University of Cape Town

Table 1: Descriptive Statistics

	Population[1] N = 2825		Sample[2] N = 868	
	Mean	Standard Deviation	Mean	Standard Deviation
Individual Attributes:				
African	0.286	0.452	0.268	0.443
Coloured	0.118	0.323	0.119	0.324
Indian	0.117	0.322	0.115	0.320
White	0.448	0.497	0.458	0.498
Male	0.609	0.488	0.589	0.492
First year student	0.868	0.339	0.893	0.309
Commerce Faculty	0.588	0.492	0.611	0.488
Logged matric points[3]	4.004	0.136	4.014	0.139
South African	0.814	0.390	0.814	0.390
Proportion of tutorials attended (Number of tutorials/total (9) tutorials)[4]	0.895	0.186	0.914	0.162
Logged test 1 mark	4.067	0.289	4.086	0.285
Logged test 2 mark	3.813	0.426	3.843	0.406
Logged exam mark	3.967	0.280	3.973	0.286
Age	19.423	1.598	19.427	1.603
Age squared	379.827	73.368	379.977	73.582
English speaker[5]	0.660	0.474	0.660	0.474
IEB matriculation[6]	0.181	0.385	0.176	0.381
Parental Attributes:				
Mothers education[7]	14.040	2.693	14.027	2.693
Fathers education[7]	14.801	2.527	14.795	2.531
Mother has a University degree[8]	0.508	0.500	0.505	0.500

University of Cape Town

1 The population is defined as all 1525 students that were registered for Economics 110 at the University of Cape Town in 2005.

2 The sample constitutes all students of the registered Economics 110 that completed the additional questionnaire administered by the tutors.

3 South African tertiary institutions make admission decisions based on a points system which allocates points on the basis of the grade of the subject taken (Higher Grade and Standard Grade) as well as the symbol obtained on the particular grade.

4 The proportion of tutorials attended is based on tutorial attendance taken by the tutors during the tutorials. It is calculated by dividing the number of tutorials attended by a respondent by 9 (the total number of tutorials). Students were expected to attend all tutorials as it was "compulsory", despite this there is still considerable variation in attendance of students at tutorials.

5 This dummy is coded as 1 if the respondent said he/she speaks English at home (regardless of whether of not the individual speaks more than one language at home) and 0 if this is not the case.

6 IEB = Independent Examinations Board. This dummy is equal to 1 if the respondent attended a school that writes the IEB matric exams and 0 if he/she did not. All private schools and a number of the ex-Model C (generally the better off public schools) learners write the IEB exams.

7 Mothers and Fathers education is measured as the number of years of schooling. Parents that have a diploma without matric are considered to have 14 years of schooling, if they had completed a diploma after matriculating it is assumed that they have had 14 years of schooling, degree holders are coded as having 15 years of schooling, while parents with a postgraduate qualification are considered to have had 17 years of schooling.

8 These dummies equal 1 if the parent received a full undergraduate degree, and 0 otherwise.

Figure 1: Illustration of Importance of Intermediaries:

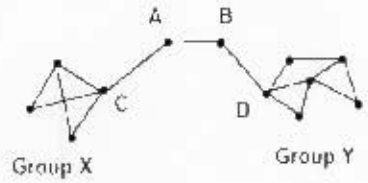
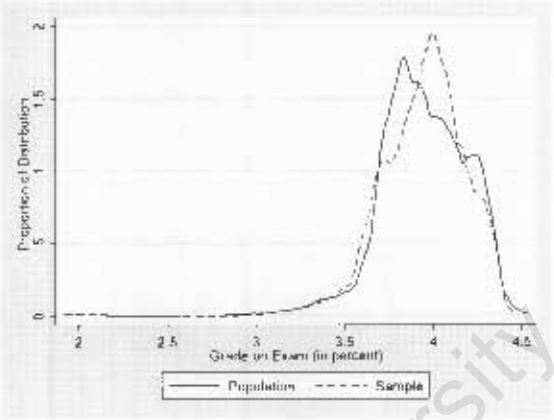
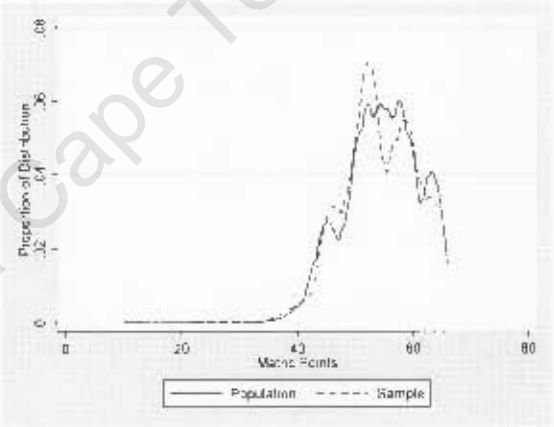


Figure 2: Comparative Academic Performance Distributions (Population vs. Sample)[1]

a) Kernel Density of Current Academic Performance



b) Kernel Density of Past Academic Performance



University of Cape Town

1 The population is defined as all 1525 students that were registered for Economics 110 at the University of Cape Town in 2005.
2 The sample constitutes all students of the registered Economics 110 that completed the additional questionnaire administered by the tutors.

Table 2: Global Structural Properties of Social Networks

Network	Type	Number of vertices[1]	Number of edges[2]	Mean degree[3]	Mean geodesic distance[4]
Economics 110 Class	Directed[6]	868	2627	3.89 (Outdegree)	9.517
Student Sexual Networks (Bearman, Moody and Stovel, 2002)	Undirected[7]	573	477	1.660	16.010
Biology Coauthorships (Newman, 2001)	Undirected	1520251	11803064	15.530	4.920
Physics Coauthorships (Newman, 2001)	Undirected	52909	245300	9.270	6.190
Email messages (Dorogovtsev et al, 2001)	Directed	59912	86300	1.440	4.950
Email address books (Newman, 2002)	Undirected	16881	57029	3.380	5.220
Adolescent sexual contacts in Uganda (Morris et al, 1997)	Undirected	389	339	0.871	
Adolescent Mental Health – Adolescent friendships (Ueno, 2004)	Directed	10023	73168	7.300	7.680

University of Cape Town

1 A vertex refers to the basic unit of a network, in this study that pertains to a student in the Economics 110 class.
2 An edge refers to the linkage between to vertices, where the linkage could indicate a sexual partnership as in Morris (1996), friendship as in Ueno (2004) or study partner (current study).
3 Degree measures the average number of edges connected to a vertex, ie. The number of study partners that a student has.
4 Mean geodesic distance measures the average shortest distance between two vertices (individuals) in a network.
5 The clustering coefficient measures the cohesiveness of the network, how clustered the relationships are.
6 A directed network is one in which the direction of the edge is known.
7 An undirected network does not account for the direction of its edges.

Table 3: Social Network Measures[1]

Panel A: Descriptive Statistics of Network Measures (Sample consists of all students who nominated study partners on the questionnaire)

	Mean	Standard Deviation
Indegree[1]	2.142	1.879
Outdegree[2]	3.891	2.469
Betweenness (Normalized)[3]	0.372	0.389
Average of logged test 1 scores of nominated study partners[4]	4.108	0.213
Absolute value of standard deviation of logged test 1 results of nominated study partners[5]	1.884	1.074
Average of logged test 1 scores of all study partners[6]	4.103	0.202
Logged test 1 result*Average of logged test 1 scores of nominated study partners[7]	16.816	1.639
Ethno-linguistic Fractionalization nationality measure8	0.150	0.250
Ethno-linguistic Fractionalization race measure[9]	0.464	0.274

Panel B: Proportion of Study Partners in respondent's network similar to the respondent

	Proportion of Study partners similar to the respondent along the observable characteristic		
	All nominations[10]	First 3 nominations[11]	Most influential nominations academically[12]
Nationality:			
South African	0.918	0.920	0.922
Foreigner	0.439	0.411	0.410
Race:			
Black	0.846	0.856	0.816
Coloured	0.450	0.456	0.474
Indian	0.624	0.650	0.664
White	0.868	0.870	0.878
Other	0.356	0.366	0.291
Gender:			
Male	0.803	0.831	0.802
Female	0.327	0.260	0.306

Panel C: Assortativity Coefficients[13]

Dimension	Directed Nominations	Undirected nominations	First 3 study partners	Most influential on academics
Nationality	0.460	0.460	0.570	0.500
Race	0.650	0.650	0.670	0.640
Race (excluding foreigners)	0.640	0.650	0.660	0.650
Race (subsample only Black and White relations)	0.820	0.830	0.770	0.820
Race (white and non-white relations only)	0.630	0.590	0.590	0.570
Gender	0.480	0.480	0.510	0.500

1 Indegree measures the number of Economics 110 students that identified this individual as part of their study network.
2 Outdegree measures the number of Economics 110 students that the respondent identified as individuals with which he/she studies.
3 Betweenness measures the interconnectedness of the individual, the greater the value of this coefficient the more connected the individual is directly and indirectly to other members of the network.
4 This is calculated as the average performance of the respondent's study partners' (only those that the individual nominated) logged Economics 110 test 1 results.
5 This is calculated as the standard deviation of the respondent's study partners' (only those that the individual nominated) logged Economics 110 test 1 results.
6 This is calculated as the average performance of the respondent's study partners' (study partners include both those nominated by the individual as well as those that nominated him/her) logged Economics 110 test 1 results.
7 The interaction of the respondents test 1 result (logged) and the average performance of his study network (only including those nominations made by the respondent).
8 The Ethno-linguistic Fractionalization formula was adopted. This index is used to measure foreign vs local diversity.
9 The Ethno-linguistic Fractionalization formula was adopted. This index is used here to measure racial diversity.
10 All nominations refers to the fact that the study network for each individual is defined as all those individuals that the respondent indicates as study partners.
11 The only study partnerships recognized by this are the first 3 individua's nominated as study partners by each respondent.
12 One of the questions asked on the questionnaire was "Please rate the strength of influence this person has on your academic acheivement on a scale of 1 (very little) to 5 (substantially)." The most influential study partners considered are those that the respondent reported as being more influential than the calculated mean influence.
13 Assortativity coefficients are defined in Section 3.1.

University of Cape Town

Table 4: "Quantity" Network Measures

Variables	Logged exam mark			
	(1)	(2)	(3)	(4)
Proportion of tutorials attended	0.092	0.092	0.095	0.085
	(0.033)*	(0.027)**	(0.028)**	(0.025)**
Test 1 (logged)	0.711	0.668	0.701	0.742
	(0.088)***	(0.035)***	(0.071)***	(0.056)***
Outdegree (Number of study partners reported by an individual)	0.111			0.052
	(0.043)*			(0.127)
Test 1 (logged) * Outdegree	-0.025			-0.011
	(0.011)*			(0.031)
Indegree (Number of times individual was nominated as a study partner by other students)		0.108		0.042
		(0.107)		(0.039)
Test 1 (logged) * Indegree		-0.025		-0.010
		(0.026)		(0.009)
Betweenness (normalized) – Measures connectedness of individuals beyond their immediate network			1.068	0.842
			(0.269)**	(0.293)**
Test 1 (logged) * Betweenness			-0.245	-0.193
			(0.068)**	(0.075)*
Constant	1.353	1.524	1.397	1.214
	(0.468)**	(0.266)***	(0.420)**	(0.373)**
Observations	619	619	619	619
R-squared	0.43	0.42	0.43	0.43

University of Cape Town

Notes:
1. A single asterisk indicates significance at the 10 percent level, a double asterisk indicates significance at the 5 percent significance level, and three asterisks indicates significance at the 1 percent level. Robust standard errors are reported in parentheses.
2. The regression also controls for: race, commerce faculty, first year student status, gender, South African citizenship, age, parental education measured by whether or not parents hold a degree, English-speaking (self-reported). The number of matric points is not controlled for as this drops over 100 students in the regression, including it is also problematic for preserving the represenatitivity of the sample as the missing values are primarily for non-South African students.

1. The normalized betweenness measure is computed using the formula (Freeman 1977 presented in Wasserman and Faust, 1994):

$$C_B(n_i) = \left(\frac{\sum_{j<k} (g_{jk}(n_i))}{g_{jk}}\right) / \left[(g-1)\frac{(g-2)}{2}\right]$$

where:

g_{jk} = the number of geodesics linking 2 individuals j and k;

i = a distinct individual required for the flow of information between individuals j and k;

$g_{jk}(n_i)$ = the number of geodesics connecting individuals j and k using i;

$\left[\frac{(g-1)(g-2)}{2}\right]$ = the maximum number of pairs of individual's in the underlying network.

Intuitively, the betweenness proportion for individual i is a sum of probabilities (where the probability measures the likelihood that i lies on a geodesic between two other individual's – namely j and k) divided by the maximum number of pairs of nodes (individual's) in the network.

Table 5: Quality Network Measures

z	Logged exam mark						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proportion of Tutorials attended	0.058	0.051	0.054	0.054	0.046	0.046	0.047
	(0.047)	(0.046)	(0.055)	(0.056)	(0.055)	(0.054)	(0.056)
Test 1 performance (own, logged)	0.575	2.532	0.575	0.651	2.852	2.845	3.754
	(0.036)***	(0.970)*	(0.037)***	(0.081)**	(1.165)*	(1.190)*	(1.828)
Age		-0.028	-0.012	-0.012	-0.013	-0.014	-0.02
		(0.007)**	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
Mom has a degree		0.025	0.017	0.017	0.019	0.018	0.02
		(0.011)*	(0.011)	(0.010)	(0.014)	(0.014)	(0.015)
Dad has a degree		0.038	0.025	0.025	0.026	0.026	0.03
		(0.013)**	(0.013)	(0.012)	(0.012)*	(0.013)	(0.013)
English-speaking		0.075	0.056	0.056	0.055	0.055	0.06
		(0.038)	(0.041)	(0.040)	(0.043)	(0.044)	(0.044)
Average of study network (logged)	0.140	2.080	0.142	0.140	2.290	2.294	3.184
	(0.069)	(0.969)*.	(0.066)*	(0.065)*	(1.105)	(1.106)	(1.863)
Own performance * Average performance of study network		-0.472			-0.523	-0.522	-0.741
		(0.229)			(0.261)	(0.266)	(0.433)
Betweenness (normalized)	0.037	0.039	0.032	0.034	0.037	0.037	0.037
	(0.019)	(0.017)*	(0.026)	(0.026)	(0.024)	(0.025)	(0.025)
Standard deviation of logged performance of study network (absolute value)			0.006	0.154	0.219	0.237	2.176
			(0.015)	(0.113)	(0.130)	(0.193)	(3.081)
Own performance * Standard deviation of study network				-0.036	-0.052	-0.051	0.528
				(0.026)	(0.030)	(0.033)	(0.716)
Study Network performance * Standard deviation of study network						-0.005	0.473
						(0.051)	(0.753)
Own performance * Average performance of study network * Standard deviation of study network							0.115
							(0.174)
Constant	1.107	-6.567	1.810	1.981	-2.903	-4.681	0.63
	(0.564)	(4.190)	(0.276)***	(0.645)**	(4.292)	(2.668)	(8.951)
Observations	538	538	538	538	538	538	538
R-squared	0.45	0.45	0.45	0.45	0.45	0.45	0.45

University of Cape Town

Notes:
1. A single asterisk indicates significance at the 10 percent level, a double asterisk indicates significance at the 5 percent significance level, and three asterisks indicates significance at the 1 percent level. Robust standard errors are reported in parentheses.
2. The regression also controls for: race, commerce faculty, first year student status, gender, South African citizenship, age, parental education measured by whether or not parents hold a degree, English-speaking (self-reported). The number of matric points is not controlled for as this drops over 100 students in the regression, including it is also problematic for preserving the represenatitivity of the sample as the missing values are primarily for non-South African students.

Table 6: Quantile Regressions - Quality and Quantity Network Measures

Variables	Logged exam mark		
	25th Percentile	Median	75th Percentile
Proportion of Tutorials attended	0.071	0.068	0.086
	(0.079)	(0.058)	(0.062)
Test 1 performance (own, logged)	0.675	0.599	0.483
	(0.053)***	(0.040)***	(0.041)***
Betweenness (normalized)	0.043	0.034	0.003
	(0.035)	(0.023)	(0.021)
Average of study network (logged)	0.170	0.120	0.089
	(0.081)**	(0.025)*	(0.052)*
Standard deviation of logged performance of study network (absolute value)	0.009	-0.005	0.008
	(0.015)	(0.010)	(0.009)
Constant	0.899	1.275	2.080
	(0.445)**	(0.321)**	(0.321)***
Observations	545	545	545

Notes:
1. A single asterisk indicates significance at the 10 percent level, a double asterisk indicates significance at the 5 percent significance level, and three asterisks indicates significance at the 1 percent level. Robust standard errors are reported in parentheses for Column 2.
2. The regression also controls for: race, commerce faculty, first year student status, gender, South African citizenship, age, parental education measured by whether or not parents hold a degree, English-speaking (self-reported). The number of matric points is not controlled for as this drops over 100 students in the regression, including it is also problematic for preserving the represenatitivity of the sample as the missing values are primarily for non-South African students.

University of Cape Town

Table 7: Race Network Measures

Variables	Logged Exam Mark						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proportion of Tutorials attended	0.077	0.070	0.043	0.040	0.032	0.034	0.024
	(0.041)	(0.042)	(0.049)	(0.055)	(0.052)	(0.058)	(0.058)
Test 1 performance (own, logged)	0.623	0.613	0.577	0.523	2.706	0.468	2.900
	(0.051)***	(0.050)***	(0.036)***	(0.090)***	(0.761)**	(0.106)**	(0.875)**
Betweenness (normalized)	0.039	0.036	0.026	0.026	0.027	0.025	0.026
	(0.016)*	(0.014)*	(0.015)	(0.016)	(0.014)	(0.017)*	(0.014)
Herfindahl Index	-0.082	-0.120	-0.136	-0.139	-0.149	-0.947	-1.118
	(0.034)*	(0.031)**	(0.034)**	(0.034)**	(0.032)**	(0.623)	(0.714)
Herfindahl Index * Own test 1 performance		0.022		0.125	0.114	0.041	0.012
		(0.008)*		(0.376)	(0.295)	(0.243)	(0.242)
Average of study network (logged)			0.140	0.189	2.345	0.068	2.457
			(0.068)	(0.097)	(0.868)*	(0.062)	(0.931)*
Herfindahl Index * Study Partner's average test 1 performance			0.027	-0.098	-0.086	-0.015	0.014
			(0.007)**	(0.287)	(0.291)	(0.239)**	(0.239)
Own performance * Average performance of study network					-0.526		-0.588
					(0.192)*		(0.230)*
Herfindahl Index * Own performance * Average performance of study network						0.047	0.057
						(0.037)**	(0.042)
Constant	1.835	2.131	1.681	1.893	-6.226	1.991	-0.609
	(0.314)***	(0.371)***	(0.794)	(0.700)*	(3.906)	(0.726)*	(12.855)
Observations	579	579	538	538	538	538	538
R-squared	0.42	0.42	0.45	0.46	0.46	0.46	0.47

University of Cape Town

Notes:

1. A single asterisk indicates significance at the 10 percent level, a double asterisk indicates significance at the 5 percent significance level, and three asterisks indicates significance at the 1 percent level. Robust standard errors are reported in parentheses for Column 2.
2. The regression also controls for: race, commerce faculty, first year student status, gender, South African citizenship, age, parental education measured by whether or not parents hold a degree, English-speaking (self-reported). The number of matric points is not controlled for as this drops over 100 students in the regression, including it is also problematic for preserving the represenatitivity of the sample as the missing values are primarily for non-South African students.

Table 8: Quantile Regressions – Racial Composition Network Measures

Variables	Logged Exam Mark		
	25th Percentile	Median	75th Percentile
Proportion of Tutorials attended	-0.016	0.038	0.082
	(0.078)	(0.060)	(0.067)
Test 1 Performance (own, logged)	0.551	0.515	0.439
	(0.084)***	(0.074)***	(0.085)***
Herfindahl Index	-0.172	-0.032	-0.067
	(0.060)***	(0.048)	(0.052)
Betweenness (normalized)	0.038	0.023	0.002
	(0.030)	(0.024)	(0.024)
Average of Study Network (logged)	0.347	0.171	0.152
	(0.098)***	(0.079)**	(0.089)*
Herfindahl Index * Study Partner's Average Test 1 Performance	-0.339	-0.156	-0.143
	(0.172)**	(0.135)	(0.159)
Herfindahl Index * Own Test 1 Performance	0.377	0.162	0.149
	(0.171)**	(0.135)**	(0.161)
Observations	545	545	545
Pseudo R-squared	0.29	0.28	0.28

Notes:
1. A single asterisk indicates significance at the 10 percent level, a double asterisk indicates significance at the 5 percent significance level, and three asterisks indicates significance at the 1 percent level. Robust standard errors are reported in parentheses for Column 2.
2. The regression also controls for: race, commerce faculty, first year student status, gender, South African citizenship, age, parental education measured by whether or not parents hold a degree, English-speaking (self-reported). The number of matric points is not controlled for as this drops over 100 students in the regression, including it is also problematic for preserving the represenatitivity of the sample as the missing values are primarily for non-South African students.

University of Cape Town

Appendix A: Questionnaire:

University of Cape Town
School of Economics
Economics 110F
Questionnaire on social networks

The information collected here will be confidential and will not be matched to your name in the database. The information that will be collected here will be used to understand the dynamic within the Economics 110F class better. We would appreciate it if you could answer as fully and honestly as possible as it will contribute to the usefulness and validity of the study.
THANK YOU!

A. Background information:

1. Full Name: (no nicknames) _____
2. Surname: _____
3. Student number: _____
4. Gender: (please tick the appropriate box) Male Female
5. What population group are you? Black White Coloured Indian Other
6. Date of birth ? _____
7. What is your physical term address? _____
   (Please be sure to indicate town/ _____
   suburb and province) _____
8. What is your physical home address? _____
   (Please be sure to indicate town/ _____
   suburb and province) _____
9. Which language do you speak at home? _____
10. What field of study are you currently registered for? (e.g. BCom (PPE)) _____
    _____
11. Are you a South African citizen? Yes No

B. Education:

1. What school did you attend for high school? (Please provide the full name) _____
   _____
2. Is this school: Public Private
3. What were the ANNUAL school fees?
   _____
4. In which suburb is this school located? _____
5. In which town is this school located? _____
6. In which province is this school located? _____

University of Cape Town

7. How many subjects did you take? _____
8. How many of these were taken on Higher Grade? _____
9. What was your aggregate % mark in your matric finals? _____

C. Family Background Information:

1. What is the highest level of education your MOTHER has completed?

No schooling	Grade 1/Sub A	Grade 2/Sub B
Standard 1/Grade 3	Standard 2/Grade 4	Standard 3/Grade 5
Standard 4/Grade 6	Standard 5/Grade 7	Standard 6/Grade 8
Standard 7/Grade 9	Standard 8/Grade 10	Standard 9/Grade 11
Standard 10/Grade 12/Matric	NTC I	NTC II

Diploma/Certificate (of at least 6 months) with less than matric
Diploma/Certificate (of more than 6 months) with Matric
Degree
Postgraduate Degree/Diploma
Other, please specify _____
Don't know

2. What is your MOTHER'S occupation? (Please give as much information as possible, for e.g. Electrical Technician. If your mother is deceased please state her occupation before she died.)

_____

3. What is the highest level of education your FATHER has completed?

No schooling	Grade 1/Sub A	Grade 2/Sub B
Standard 1/Grade 3	Standard 2/Grade 4	Standard 3/Grade 5
Standard 4/Grade 6	Standard 5/Grade 7	Standard 6/Grade 8
Standard 7/Grade 9	Standard 8/Grade 10	Standard 9/Grade 11
Standard 10/Grade 12/Matric	NTC I	NTC II

Diploma/Certificate (of at least 6 months) with less than matric
Diploma/Certificate (of more than 6 months) with Matric
Degree
Postgraduate Degree/Diploma
Other, please specify _____
Don't know

4. What is your FATHER's occupation? (Please give as much information as possible, for e.g. Electrical Technician. If your father is deceased please state his occupation before he died.)_____

____

5. What make and model car does your mother drive? _____
6. What make and model car does your father drive? _____
7. Do your parent(s) OWN (not rent) the house that you wrote down as your permanent home address? Yes No
8. If yes, please could you indicate the value of the house in which they reside (if known).

_____

University of Cape Town

9. Do you have any assets in your name? Yes No

   (a) If yes, what are these assets? _____

   (b) Please could you estimate the value of these assets? _____

10. Have you received an inheritance? Yes No

11. What was the value of this inheritance ? _____

12. Do you expect to receive an inheritance at some point? Yes No

13. What is the expected value of this inheritance? _____

14. Are you on financial aid? Yes No

15. Where do you think that your family fits into the income distribution, on a scale of 0 to 10? Where 0 indicates the poorest of the poor and 10 is the richest of the rich. _____

D. Future Plans:

1. What do you expect to do after completing your undergraduate degree? (Please be specific, Saying that you would like to get a job is insufficient)

   _____

2. Do you have a job placement secured for when you graduate? Yes No

   (a) If yes, please specify what it is and at which company. _____

3. What career do you intend to pursue? _____

University of Cape Town

. Social Networks in Economics 110F

Please tell us about all your **friends and associates** that also take **ECONOMICS 110F**. Please indicate all **friends, individuals who you study with,** and **those that you ask for help** with your Economics studies. **NB! Please ONLY provide information for those friends and associates in your ECONOMICS 110f.**

	1. Name (NO Nicknames)	2. Surname	3. Contact Number (preferably cell phone)	4. e-mail address	5. What field of study is he/she currently registered for?	6. What school did he/she attend?	7. In which town/city and province is this school situated?
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

University of Cape Town

The following code list will be required for **QUESTION E. 20** , that follows on the next page.

Code list for types of interaction:

1. Outdoor Activities (Hiking, Mountain climbing etc.)
2. Play sport together/Go to gym together
3. Go to Movies
4. Go out for drinks
5. Go out clubbing
6. Studying together
7. Have sex with each other
8. Go for coffee together
9. Watch sport together
10. Have braais with one another
11. Going out for meals together
12. Hanging out together (talking etc.)

University of Cape Town

Now, please tell us about all your **friends and associates** in UCT or elsewhere (not necessarily students) with whom you **interact with on a regular basis**. (Please do NOT repeat those already recorded in the previous question).

	Name (NO Nicknames)	Surname	Contact Number (preferably cell phone)	e-mail address	If he/she is not a registered UCT student, where is he/she currently registered, or what is his/her occupation?	What field of study is he/she currently registered for?	What school did he/she attend?	In which town/city and province is this school situated?
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

University of Cape Town

The following code list will be required for **QUESTION F. 21** , that follows on the next page.

Code list for types of interaction:

13. Outdoor Activities (Hiking, Mountain climbing etc.)
14. Sport
15. Go to Movies
16. Go out for drinks
17. Go out clubbing
18. Studying
19. Sex
20. Go for coffee
21. Watch sport together
22. Have braais
23. Going out for meals
24. Hanging out together (talking etc.)

University of Cape Town

Question Number	Using those individuals captured in the previous table, copy across the names in the first col				
		1	2	3	4
1	Full name (NO nicknames)				
8	Gender	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female
9	What population group?	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White
10	Is this person related to you?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
11	If yes, what is the nature of the relation?				
12	If you are siblings, are you identical twins?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
13	How long have you known the individual? (Please indicate either in weeks, months OR years)	_____ Weeks _____ Month _____ Years	_____ Weeks _____ Month _____ Years	_____ Weeks _____ Month _____ Years	_____ Weeks _____ Month _____ Years
14	Did you attend the same high school?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
15	Do you live with this person? (In residence, or share a flat etc)	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
16	Is this person a neighbour AND/OR does he/she live within close walking distance?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
17	How many times did you have contact with this person in the last 3 months?				
18	If you answered 0 times above, did you try make contact with this person that was unsuccesful?				
19	How many of the times (in the last 3 months) you made contact was it face-to-face contact?				
20	Typically when you make contact what form does it take? (can select multiple options – see code list attached)				
21	How many times did you GO TO THIS PERSON'S PLACE OF RESIDENCE in the last 7 days?				
22	How many times did you GO OUT TOGETHER in the last 7 days?				
23	How many times did you TALK TO THIS PERSON ABOUT A PROBLEM in the last 7 days?				
24	How many times did you SPEAK ON THE PHONE to this person in the last 7 days?				
25	How many times did you SMS this person in the last 7 days?				
26	How many times did you EMAIL this person in the last 7 days?				
27	Do you lend class notes/summaries to this person?				
28	Does this person lend class notes/summaries to you?				
29	Please rate the strength of your relationship with this person on a scale of 1(acquaintance) to 10 (best friend).				
30	Please rate the strength of influence this person has on your academic achievement on a scale of 1(very little) to 5 (substantially).				
31	Please rate the strength of influence this person has on you in general on a scale of 1(very little) to 5 (substantially).				
32	On average how many HOURS a WEEK do you spend TOGETHER on the following activities?				
(a)	Studying together/Formal Study Group				
(b)	Lectures/Classes				
(c)	Socially				
(d)	Sport				
(e)	Cultural Activities				
(f)	NGO work				
(g)	Committees/Meetings				
(h)	Business/Work				

University of Cape Town

umn to the first row in this table, and then answer the questions for each individual.

5	6	7	8	9	10
☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female
☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
________ Weeks ________ Month ________ Years	________ Weeks ________ Month ________ Years	________ Weeks ________ Month ________ Years	________ Weeks ________ Month ________ Years	________ Weeks ________ Month ________ Years	________ Weeks ________ Month ________ Years
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

University of Cape Town

Question Number	Using those individuals captured in the previous table, copy across the names in the first co				
		1	2	3	4
1	**Full name (NO Nicknames)**				
9	**Gender**	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female
10	**What population group?**	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White
11	**Is this person related to you?**	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
12	**If yes, what is the nature of the relation?**				
13	**If you are siblings, are you identical twins?**	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
14	**How long have you known the individual? (Please indicate either in weeks, months OR years)**	_____ Weeks _____ Month _____ Years	_____ Weeks _____ Month _____ Years	_____ Weeks _____ Month _____ Years	_____ Weeks _____ Month _____ Years
15	**Did you attend the same high school?**	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
16	**Do you live with this person? (In residence, or share a flat etc)**	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
17	**Is this person a neighbour AND/OR does he/she live within close walking distance?**	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
18	**How many times did you have contact with this person in the last 3 months?**				
19	**If you answered 0 times above, did you try make contact with this person that was unsuccesful?**				
20	**How many of the times (in the last 3 months) you made contact was it face-to-face contact?**				
21	**Typically when you make contact what form does it take? (can select multiple options – see code list attached)**				
22	**How many times did you GO TO THIS PERSON'S PLACE OF RESIDENCE in the last 7 days?**				
23	**How many times did you GO OUT TOGETHER in the last 7 days?**				
24	**How many times did you TALK TO THIS PERSON ABOUT A PROBLEM in the last 7 days?**				
25	**How many times did you SPEAK ON THE PHONE to this person in the last 7 days?**				
26	**How many times did you SMS this person in the last 7 days?**				
27	**How many times did you EMAIL this person in the last 7 days?**				
28	**Please rate the strength of your relationship with this person on a scale of 1(acquaintance) to 10 (best friend).**				
29	**Please rate the strength of influence this person has on your academic achievement on a scale of 1(very little) to 5 (substantially).**				
30	**Please rate the strength of influence this person has on you in general on a scale of 1(very little) to 5 (substantially).**				
31	**Of the friends listed in this section (1 – 10), write down the corresponding column numbers to indicate which of these this person is also friends with.**				
32	**On average how many HOURS a WEEK do you spend together on the following activities?**				
(a)	**Studying together/Formal Study Groups**				
(b)	**Lectures/Classes**				
(c)	**Socially**				
(d)	**Sport**				
(e)	**Cultural Activities**				
(f)	**NGO work**				
(g)	**Committees/Meetings**				
(h)	**Business/Work**				

University of Cape Town

lumn to the first row in this table, and then answer the questions for each individual.

5	6	7	8	9	10
☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female	☐ Male ☐ Female
☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White	☐ Asian ☐ Black ☐ Coloured ☐ Indian ☐ White
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
______ Weeks ______ Month ______ Years	______ Weeks ______ Month ______ Years	______ Weeks ______ Month ______ Years	______ Weeks ______ Month ______ Years	______ Weeks ______ Month ______ Years	______ Weeks ______ Month ______ Years
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

University of Cape Town