Social Structure, Segregation, and Economic Behavior *

Matthew O. Jackson†

Revised: February 5, 2009

Presented as the Nancy Schwartz Memorial Lecture in April of 2007.

*Financial support from the NSF under grant SES–0647867 is gratefully acknowledged. I thank Toni Calvó-Armengol, Sergio Currarini, Ben Golub, and Paolo Pin for collaboration on various parts of the research discussed below as well as for helpful conversations in shaping my thinking on this subject.

†Department of Economics, Stanford University and the Santa Fe Institute. Email: jacksonm@stanford.edu, http://www.stanford.edu/∼jacksonm/

1 Introduction

Social structure shapes many of our economic interactions as well as our decisions. Examples of situations where our social networks influence our behavior range from everyday decisions of which products to buy to long-term decisions such as which profession and how much education to pursue. Our social networks influence our job opportunities, how exposed we are to criminal behavior, and provide informal insurance, among many other things. This makes it important both to understand how our behaviors depend on our social networks as well as why social networks exhibit certain patterns.

In this lecture I discuss a fundamental and pervasive aspect of social networks that has a significant impact on behavior: "homophily." Homophily refers to the tendency of individuals to associate with others who have similar characteristics as themselves. This tendency has been observed across a variety of dimensions including ethnicity, age, gender, profession, and education level, among others. Given that our opinions, behaviors, and decisions are influenced by those we are in contact with, having a thorough grasp of homophily and how it shapes our social networks and ultimately translates into our behaviors becomes imperative.

This lecture proceeds in three parts. It begins with background on homophily, providing some illustrations and a very brief look at some empirical studies on homophily. It then proceeds to the impact of homophily, discussing the effects of homophily on three different things: how information diffuses through a society, how individuals make education decisions, and social mobility. After discussing the impact of homophily on these different behaviors, I conclude with a discussion of a model of homophily and what we might learn about what leads to homophily.

2 Background on Homophily

Homophily, a term coined by Lazarsfeld and Merton (1954), has been documented across a variety of characteristics including age, race, gender, religion, and profession.[1] To develop a feeling for homophily let me start with a few illustrations. Based on a national survey, Marsden (1987, 1988) reports that only 8 percent of people have any people of another race with whom they "discuss important matters." That percentage is far below what one would expect if such discussions occurred without race playing a role in who talked to whom. Verbrugge (1977) reports on interviews where people were asked who their closest friend was. Only approximately 20 percent of the respondents named a person of the opposite sex

[1] See McPherson, Smith-Lovin, and Cook (2001) for a survey of some of the literature on homophily.

as their closest friend, when clearly that percentage would be roughly 50 percent if gender played no role in friendship formation. Fryer (2006) reports on interracial marriages in the U.S. and notes that only one percent of Whites marry non-Whites, five percent of Blacks marry non-Blacks, and fourteen percent of Asians marry non-Asians; which are again all significantly lower than the percentages that would be exhibited by a society where marriage was blind to race.

As a more detailed example, let us consider the following data from the Adolescent Health data set.[2] These data provide mappings of the friendships in 84 different high schools. One of the high schools is pictured in Figure 1. Nodes in the network are high school students. A link is present between two students if either student named the other as a friend in an interview.[3] Each student could name up to five male and five females as friends. Students also reported themselves as beingin one of five racial categories: Asian, Black, Hispanic, White or other.

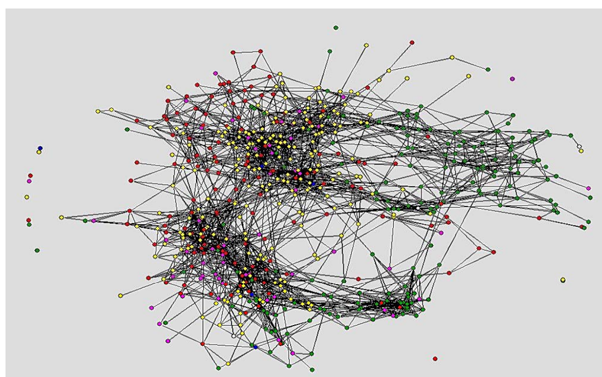

Figure 1: A Network of the Friendships in a High School from the Ad Health Data Set: green=Asian, blue=Black, red=Hispanic, yellow=White, pink=other/unknown.

Table 1 provides the numbers behind Figure 1. As we see, Whites and Asians both have disproportionate fractions of friendships with their own race. This pattern is representative

[2]Add Health is a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).

[3]Less than half of the students who name another student as a friend are named in return. There are various reasons that this might be occurring, including potential measurement error both in terms of which friends a student remembers to name at the time of the interview, caps on the numbers that can be named. This may also represent differences in perceptions of the students as to what constitutes a friend.

of the various high schools in the data set. As Currarini, Jackson and Pin (2007) point out, this high school is fairly typical: homophily is highest for groups that form close to half of the school's population.

Race	As a Percent of Population	Percentage of Friendships with Own Race
White	55	75
Asian	32	65
Hispanic	6	5
Black	1	1
Other/Unknown	6	-

Table 1: Percentage of Links Across Ethnicities in an American School; from the Add Health 1994 Data.

This pattern of segregation across groups gets even more pronounced when we only keep track of "strong" relationships: these are links between individuals who report taking part in at least three activities together in a week. Figure 2 pictures the strong links in a high school which is in this case consists predominately of Black and White students, with a small percentage of Hispanic students.

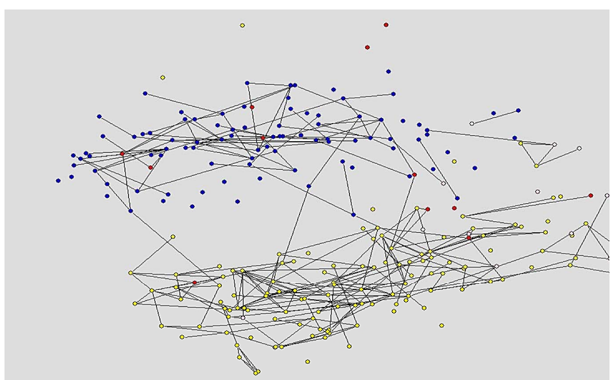

Figure 2: A Network of the Friendships in a High School from the Ad Health Data Set with only "Strong Friendships:" green=Asian, blue=Black, red=Hispanic, yellow=White, pink=other/unknown.

As we see in Figure 2 there are very few strong friendships that link across races.

There are various possible explanations for the patterns observed, including the frequency with which students of one race meet students of another race due to the classes and activities that they participate in, preferences over the characteristics of their friens, peer pressure and group norms. It could be that it is not race per se that affects friendship formation, but instead that race happens to be correlated with other factors such as socio-economic background which then are the factors which influence friendship formation. In order to sort some of this out, Moody (2001) regresses (log linearly) the probability that two nodes are friends over the probability that they are not friends on whether they are of the same race, their socio-economic backgrounds, their grade, gender, their contact through various clubs and extracurricular activities, their observed behaviors (fights, smoking, popularity, skipping school, attitudes), and their similarity in gpa. Looking across the different schools, he still finds a range of coefficients on the race variable, with significant and nontrivial coefficients on the racial variable in schools that have substantial heterogeneity in their racial makeup. While there may still be other factors at work, this shows some correlation between race and friendship, even after controlling for many other factors.

3 The Impact of Homophily

Given some background on homophily, let us now explore some of its implications.

3.1 Homophily's Impact on Communication and Learning

Let me start with a discussion of how homophily affects communication and learning. In particular, let us consider people in a society who are discussing things with their friends and acquaintances over time. They use such discussions to form opinions about whether a product is worth purchasing or not, to determin which political candidate they should vote for, to find out about events, and so forth. Homophily affects the patterns in which people interact and thus influences what opinions end up being formed and how quickly news or information is transmitted through a network. There are many different ways in which people communicate and learn, and so the effect of homophily can depend on the context. To illustrate some of its impact, let me discuss a couple of forms of communication and how homophily impacts them, as analyzed by Golub and Jackson (2008).

I start by describing a simple model that is particularly useful as a model of communication and learning in a network. It is a model that has appeared in various incarnations and traces to back to French (1956), Harary (1959), and DeGroot (1974). In this model each individual i has some belief or behavior at a time t that can be represented by a number $b_i(t)$, and for simplicity let us code these as numbers in [0,1]. So, for instance, we might think of the beliefs as being an agent's estimated probability of some event happening, like the probability of a given company going bankrupt in the coming year. Over time people talk to each other and update their beliefs. The updating takes a very simple form. Each individual has a weight that he or she places on the belief of each of his or her friends. When an individual talks to his or her friends, he or she averages their beliefs with these weights. Agents keep repeating this process, so in the next period they talk to their friends again since their friends' opinions might have changed. In this way, after a few repetitions agents' beliefs begin to incorporate the beliefs of friends of friends, and friends of friends of friends, and so forth, since the friends have also been updating their beliefs to incorporate information from their friends.[4] So, i's belief at time t is represented as

$$b_i(t) = \sum_j T_{ij} b_j(t-1),$$

where T_{ij} is the "trust" or weight that i places on j's belief at time t. These weights sum to 1, so $\sum_j T_{ij} = 1$, and they are nonnegative with $T_{ij} \geq 0$ for all i and j, so that an agent is really averaging beliefs over time, as otherwise beliefs might grow or shrink without bound. An agent i can weight his or her own beliefs, so that $T_{ii} > 0$, in which case i's belief at time t is a weighted average of i's previous belief and i's neighbors' previous beliefs. A special case is one where if i has d_i neighbors that he or she listens to, then $T_{ij} = 1/d_i$ for each j that is linked to i in the network and $T_{ij} = 0$ for j's that are not linked to i. In that case, an agent puts equal weight on each neighbor.

An illustration of this updating process appears in Figure 3, for the case of three agents such that agents 1 and 3 are connected to agent 2 but not to each other, and each agent weights the beliefs of their neighbors (including him or herself) equally. We see the indirect updating over time. In the figure, agent 1 is the only person starting with a positive belief. After the first period that causes agent 2's belief to become positive. Finally, after two periods, this filters through to agent 3 whose belief then becomes positive. The agents' beliefs converge towards a consensus over time.

This opinion updating model has nice properties, and under some simple conditions the beliefs converge, and in fact, all the beliefs will converge to a common consensus.[5] This is easy to see intuitively, at least in the case where all agents put some weight on their own

[4] This process is clearly boundedly rational, in that a Bayesian agent would adjust the weights and updating rule over time to reflect what he or she has already learned and what the anticipated new information is that will come in via different channels. A friend who in turn has many friends would be weighted differently than a friend who has only a few friends. More discussion on this can be found in DeMarzo, Vayanos and Zweibel (2003).

[5] For example, it is sufficient that the network is path-connected, so that there is a directed path from

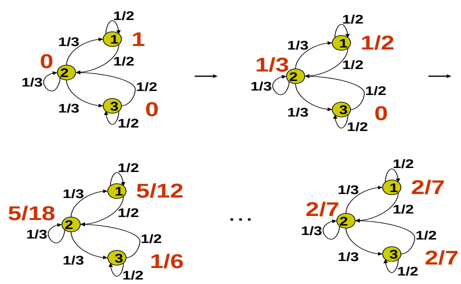

Figure 3: Updating and Convergence in the DeGroot (1974) Model. Agent 2 weighs all agents, including herself, equally. Agent 1 weighs agent 2 and himself equally, and similarly for agent 3. The updating is pictured for initial beliefs of 1 for agent 1 and 0 for each of agents 2 and 3. The limiting consensus beliefs are 2/7 for each agent.

opinions. Consider the agents who have the lowest belief at some time t (for instance agent 3 in Figure 3 at time 2). Those agents' beliefs at time $t+1$ cannot be lower than the lowest beliefs were at time t, since these agents are taking averages over beliefs and their current beliefs are already the lowest in the network. If beliefs have not yet reached a consensus, then at least one of the agents having the lowest belief in society will have a connection to some neighbor who has a belief that is higher than the lowest belief. Thus, that agent's belief will be increased in the next period. In this way, given that agents are path connected, this increase will eventually reach other agents who have the lowest beliefs. Agents who have the lowest beliefs will have their beliefs increase over time and those with the highest beliefs will have their beliefs decrease over time. The fact that they actually converge comes from the fact that the weights (the T_{ij}'s) are steady over time, and so the beliefs are increasing or decreasing a non-vanishing proportion of how much they differ.

This model is particularly useful not only because it has a simple and natural structure and has nice convergence properties, but also because it permits an easy and intuitive characterization of the influence each individual in the society has on the final consensus and how this depends on the network structure. In particular, we can find a number associated with individual i, denoted s_i, that we can interpret as the "influence" of i. In particular,

each agent to each other agent via the T_{ij}'s, and that some agent places weight on him or herself so that $T_{ii} > 0$ for some i. There are weaker conditions for convergence and a full characterization of conditions for convergence and consensus is known. See Chapter 8 in Jackson (2008) for more background and discussion of this model, its properties, and other variations on it.

these influences are such that the limiting consensus belief $b(\infty)$ can be written as

$$b(\infty) = \sum_i s_i b_i(0)$$

That is, the limiting belief is just the sum of the agents' starting beliefs times their influences, where the vector of influences s is derived from the network as captured via the weighting matrix T. In particular, s is simply the (left-hand) unit eigenvector of T. In the special case where each agent weights his or her neighbors equally that was mentioned above, and where if i is neighbors with j then the reverse is also true, it is easy to check that the weights come out to be $s_i = d_i / \sum_j d_j$. For the updating process pictured in Figure 3 these weights are $(2/7, 3/7, 2/7)$ and indeed the limiting consensus belief in that figure is the dot product of $(2/7, 3/7, 2/7)$ and the initial beliefs $(1, 0, 0)$.

Let us now think about homophily in the context of this model. We can get an impression of how this might work by starting with a very simple example with just two agents. Suppose that each agent places weight p on him or herself and weight $1 - p$ on the other agent. This serves as a sort of metaphor for the weight that two different groups place on each other, and so captures some elements of homophily, where we might think of an agent representing a group. Indeed, much of the intuition of how things operate in more complicated networks can be thought of by agglomerating a whole group of agents into a single "super-agent" and then looking at interactions across these super agents. In this case, given the symmetry, $s_1 = s_2$ and so the agents have equal influence, regardless of p. So, for any initial beliefs the same eventual consensus will be reached. The main issue is how quickly the consensus will be reached. Speed of convergence is important for a variety of reasons. It affects the stream of decisions that agents might make based on their information; and it is also the case that agents might only end up communicating a limited number of times and so very slow convergence might translate into very disparate beliefs, while fast convergence would translate into more of a consensus being reached. Let us consider a case where the agents start with differing beliefs, for instance, so that one agent starts with belief 0 and the other starts with belief 1, so, for instance, set $b_1(0) = 1$ and $b_2(0) = 0$, as pictured in Figure 4. It is clear in this case that the limiting consensus belief will be $b(\infty) = 1/2$. But how quickly will the agents reach this consensus? Here it is very clear that if p is close to 1, then convergence will be very slow as the change in beliefs in any period is just $1 - p$ times the gap in beliefs. In contrast, if p is $1/2$ or close to $1/2$, then convergence will be very rapid or even instantaneous. So, substantial homophily slows communication.

The intuition coming out of this example generalizes, as shown by Golub and Jackson (2008). They develop an expression that they call "degree weighted homophily," which is a

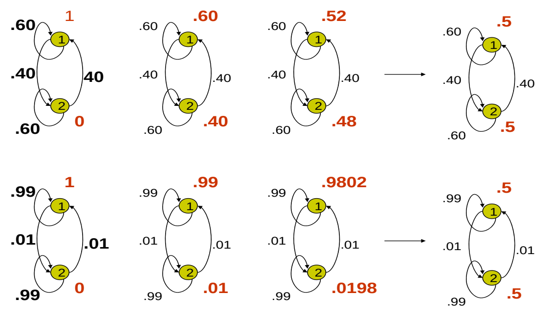

Figure 4: Updating and the Speed of Convergence in the DeGroot (1974) Model with two agents/groups. In the top panel the two agents/groups place large weight on each other and convergence is fast, while in the bottom panel they place low weight on each other and even though they converge to the same consensus limit, they converge much more slowly.

generalization of the p in this example and captures the level of homophily or segregation across groups in a society in terms of the relative weights that different groups place on themselves versus others. They then show that number of periods it takes for a society to get within some ε of its limiting consensus belief starting from the disparate beliefs is proportional to the absolute value of the inverse of the log of this homophily measure. Most importantly, as homophily gets close to one so there is little communication across groups, as in Figure 5, this time to consensus goes to infinity, while as homophily goes to 0, so that groups weight each other just as much as themselves, the consensus time goes to 0.

It is important to note that homophily can slow communication, and also to be able to quantify this precisely. However, the basic idea is not so surprising. What is perhaps more interesting emerges when we start comparing how homophily's impact depends on the basic structure of the communication process. Golub and Jackson (2008) show that the effects of homophily differ dramatically across types of communication. In particular, let us consider an alternative type of communication such that the speed of communication depends on the shortest paths in a network. That is, suppose that the process begins with some agent hearing a piece of news. The news is unambiguous and can simply be communicated by telling a friend. The agent tells everyone whom he or she knows, and they in turn tell everyone whom they know, and so forth. The speed at which information travels from one agent to another just depends on how far away two agents are from each other in the social network, corresponding to the minimal number of links that the information must cross to reach from

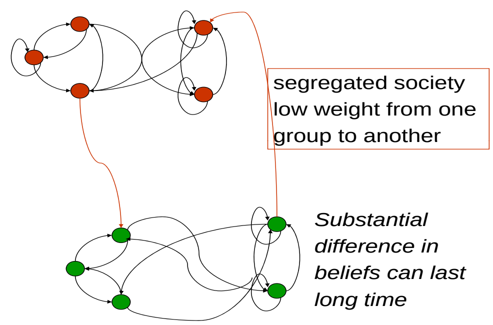

Figure 5: Updating and the Speed of Convergence in the DeGroot (1974) Model with two agents/groups. In the top panel the two agents/groups place large weight on each other and convergence is fast, while in the bottom panel they place low weight on each other and even though they converge to the same consensus limit, they converge much more slowly.

one to another. How the speed of such communication processes depends on homophily can be deduced from an analysis by Jackson (2008b) who examines how homophily affects average distance between agents in a wide class of network formation models. In particular, there it is shown that within this class of network formation models homophily does not affect average distance, and thus does not affect average communication time at all. It does affect which agents are close to a given agent and which ones are far away; but homophily does not affect the overall average distance. The idea behind this is that if we examine two different networks with similar numbers of links, but one with more homophily than the other, the similar number of links means that the neighborhood expansion properties are essentially completely unaffected: how many agents lie at a given distance from a given agent just depends on how many links the given agent has, how many links each of the given agent's neighbors have and so forth.[6]

The contrast between the repeated averaging sort of communication and the shortest path sort of communication is quite stark, as further discussed in Golub and Jackson (2008). In the first case of repeated averaging-based communication, homophily is the critical determinant of the speed of convergence as it determines relatively how much contact or weight a given individual puts on others from another group; but connectivity does not matter: it is relative weights that matter and not how many people with whom a given person interacts. In

[6]There are some complications to worry about in terms of neighborhood overlaps, but it is shown in Jackson (2008) that these are negligible effects.

contrast, with shortest path sorts of communication, the reverse is true: relative weights only affect who hears first but not average speed, while how densely connected the network is affects how quickly information spreads.

Without this sort of analysis it would not have been so obvious that homophily's impact on the speed of learning depends so drastically on the structure of the communication process. In addition, Golub and Jackson (2008) show that there is a nonlinearity in the impact of homophily on learning speeds in the averaging context: a small increase of homophily in a relatively integrated society has little impact, while a comparable increase in a more homophilistic society can decrease the speed of learning quite substantially.

3.2 Education, Human Capital, and other Decisions

We make many decisions about education during our lives: how much to study, whether or not we complete high school, whether we attend university, which profession we pursue, and so forth. These decisions are influenced by our families, friends, and peers, and for a variety of reasons. Family wealth, parental pressure and guidance, learning effects, role models, and complementarities in investment are just some of the things that lead decisions to be sensitive to social surroundings and to exhibit complementarities.

To see such complementarities in more detail, let me discuss an application studied by Calvó-Armengol and Jackson (2004, 2007). They examine investment in human capital and education in the context of a networked labor market. One version of their model can be described as follows. There are different types of jobs, some of which require education and pay high wages and others which do not require education and pay lower wages. Workers occasionally and randomly become unemployed. An unemployed agent has two ways that he or she might hear about a new job: he or she might find a job directly through his or her own efforts or he or she might hear about a job from one of his or her friends.[7] The jobs that unemployed workers hear about via their friends most likely match those friends' status: educated friends are more likely to hear jobs requiring education and uneducated friends are more likely to hear about low wage jobs that do not require education. Thus, if one agent has more educated friends than some other agent, then the first agent is more likely to hear about jobs that require education. All else held equal, this leads to a higher return from investing in education if one expects to have more friends investing in education. Thus, agents' decisions to invest in education exhibit strategic complementarities: an agent's payoff from becoming educated increases with the number of educated friends that he or she

[7] For more background on the role of networks in labor markets, see Ioannides and Datcher Loury (2004).

has. This effect is the driving force the impact of social networks on such decisions, and is an important reason as to why we should care about social interaction patterns.

Now let us examine how decisions of individuals are affected by network structure and homophily in such a context with strategic complementarities. Increased homophily leads different groups, say across race or socio-economic background, to become more segregated. As groups become more segregated it becomes increasingly possible to see very different behaviors across the groups because of these complementarities. This is easily seen via an example that is a bit extreme but provides a stark illustration of the underlying ideas and intuition. Suppose that a society consists of two types of individuals: blues and greens. Let us also examine a simple binary decision where individuals either choose to take an action or not to take it. So, they either pursue higher education or not, etc. Let us also suppose that an individual prefers to take the action if at least one third of his or her neighbors take the action and otherwise prefers not to take the action. Now let us consider two different social networks. One network exhibits substantial homophily, so that the network is such that every green has more than two thirds of his or her links to other greens and similarly blues each have more than two thirds of their links to other blues. This is pictured in Figure 6.

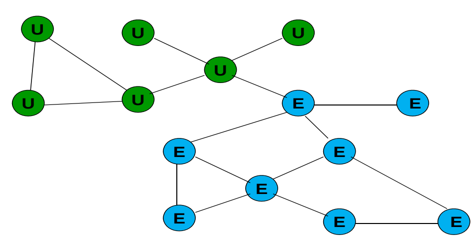

Figure 6: Homophily and education choices: an equilibrium in a situation where each agent prefers to choose to be educated (E) if at least one third of his or her friends are educated (E), and otherwise prefers to choose to be uneducated (U).

The other network is a more evenly integrated society where many agents have at least a third of their friends being of the opposite color, as pictured in Figure 7.

In the homophilistic network of Figure 6 it is possible to have an equilibrium where all of the blues take the action and all of the greens do not, and vice versa. So, the homophily

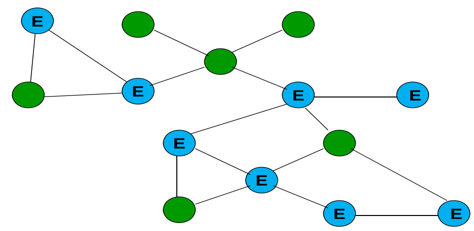

Figure 7: Homophily and education choices: an equilibrium in a situation where each agent prefers to choose to be educated (E) if at least one third of his or her friends are educated (E), and otherwise prefers to choose to be uneducated (U). If all blues choose E then it can be checked that all greens will prefer to choose E as well.

allows the two different groups to pursue completely different actions. However, in the more integrated society such segregated equilibria are impossible - in the network pictured in Figure 7, if all the blues choose E, then it must also be that all the greens choose E. This shows that homophily can affect the pattern of decisions that are taken in a society and how those decisions are distributed across groups, even in networks that are similar in all other ways except for the homophily.

Depending on the specific configuration of the networks, the structure of payoffs to different actions, and other forms of possible heterogeneity across agents, there can be many equilibria.[8] For instance, in the above example it is always an equilibrium for all agents to take the action regardless of the network, or for none of them to take the action. The point here is that homophily can change the pattern of equilibria and in some circumstances can allow for more segregated equilibria to emerge. This example is clearly overly stylized: it will generally be that there is more heterogeneity in terms of the types of neighbors that people have as well as how responsive an individual's decisions are to his or her neighbors' decisions, with some individuals being highly sensitive and others less so, and so forth. Introducing additional noise and variation leads to less stark patterns and can lead to additional complexities in equilibrium structure. Nonetheless, there can still be overall effects that are

[8] Morris (2000) provides a characterization of when it is possible to sustain two different actions in such games in terms of a cohesion condition, but without the attention to homophily. His results can be adapted to this question quite easily. For more discussion and background on such games of complementarities, see Galeotti et al. (2007) and Jackson (2008).

quite powerful.

If in addition to strong homophily in a network, for instance, across races, there are also historical influences so that different races have different pre-existing conditions, then one can see quite different levels of education and investment in human capital across different races, as pointed out by Calvó-Armengol and Jackson (2004). In turn, this can lead to differences in a series of employment outcomes across races such as persistent differences in wages even for those with equivalent skills. For instance, the few members of one race that choose high education for idiosyncratic reasons could end up with lower wages and higher unemployment rates than members of another race who choose the same education level but have more highly educated friends. Developing further understanding of how homophily influences such decisions can help to shape future policies to be more effective in addressing persistent inequalities.

3.3 Social Mobility

Complementarities in behavior in the context of social networks also have other implications. Children end up with similar outcomes and behaviors as their parents on many dimensions, including religious attitudes, political affiliation, education, choice of profession, decisions of whether to smoke, etc., and there are many reasons for such intergenerational correlations ranging from parental guidance, to genetics, to wealth effects.[9] Beyond such reasons for parent-child correlations in behaviors, many such behaviors exhibit significant complementarities with respect to the behavior of an agent's friends, family, and other acquaintances. That is, many such decisions including whether or not to smoke, which profession to choose, where to live, and so forth, are similar to the education decisions discussed above in that an agent's decision could be heavily influenced by the decisions of his or her friends and relatives. To the extent that a parent and child have overlap in their social networks or end up with similar social networks, the parent and child can end up with similar decisions simply because of these complementarities and the network overlap. This is a point made explicitly by Calvó-Armengol and Jackson (2009) in the context of a simple model where a child's social network has some overlap with the parent's social network. Indeed, as Calvó-Armengol and Jackson point out, as one increases the sensitivity of agents' choices to those of their neighbors, one sees an increase in parent-child correlation in behavior even if the parent and child have no interaction with each other whatsoever. Calvó-Armengol

[9] For more on these topics see Thomson (1971), Abramson (1973), Piketty (1995), Johnson (1996), Bisin and Verdier (2000), Berkman and Kawachi (2000) and Galea, Nandi, and Vlahov (2004).

and Jackson also show that such a model can generate correlations in parent-child education decisions that match data from a panel study of education in various European countries. While such effects do not necessarily exclude other reasons for parent-child correlation, they can lead to substantial correlations above and beyond those due to other sources. This means that when analyzing social mobility and various parent-child correlations in behavior, that we be careful to disentangle which mechanisms are at work. The differences between direct parent-child interaction and overlap in social networks can lead to very different policy prescriptions. Again, homophily and segregation in social networks can have extensive and substantial implications for similarity of behaviors across generations and within racial groups.

4 Modeling Homophily

Given the many ways in which homophily can impact behavior, it is also important to have some understanding of what might lead to homophily and to develop models which help to analyze it in more detail. There are various explanations for it, including biases in who meets whom, attitudes and social norms among groups, compatibilities, and preferences (see Blau (1977), Blalock (1982) and McPherson, Smith-Lovin, and Cook (2001) for background). Currarini, Jackson and Pin (2008) analyze a model within which the different roles of two factors that can lead to homophily can be explored. In that model, there are two potential biases that can lead to homophily in the formation of friendships: there can be a bias in the rate at which a person meets others of his or her own type compared to other types, and agents can also have preferences about the fraction of their friends who are like themselves. Clearly either of these biases could lead to homophily on its own. For instance if members of one group mainly meet members of that same group then their friends could tend to mainly be from that group. Similarly, if agents have a preference for friendships with people of their own type then they could tend to have a mix of friendships that includes more of their own type than would be representative of their type's percentage of the population as a whole.

To get a better feeling for the roles of these different types of biases in generating homophily, Currarini, Jackson and Pin (2008) exhibit two additional empirical observations regarding homophily. In particular, they examine racial homophily within the high school friendship networks from the Adolescent Health data set discussed above. With 84 high schools of social network data, they are able to examine how friendship patterns depend on the racial composition of a high school. Currarini, Jackson and Pin (2008) document two tendencies that are statistically highly significant. The first observation is that how homophilistic a given group is depends on the fraction of the high school that it comprises. If a racial group comprises half of a high school then it exhibits substantial homophily. In contrast if a group comprises a very small or very large percentage of a high school then it exhibits little homophily, even after normalizing for the size of the group. The high school described in Table 1 is quite representative: in that school Blacks and Hispanics comprise very small fractions of the population and they exhibit no homophily in their friendships, and in contrast Whites and Asians comprise fractions that are closer to half of the high school population and they exhibit substantial homophily. Thus, how homophilistic a group is depends on how large a fraction it makes up of the whole high school, and in particular there is a *non-monotonic* relationship where the groups exhibiting maximal homophily are of middle sizes, and groups that are very large or very small exhibit negligible homophily.[10] The second observation is that the numbers of friendships that individuals have depend on how large their groups are as a fraction of their high schools. If a group comprises a very small fraction of a high school then its members will have on average between four and five friends per capita, while if a group comprises a very large fraction of a high school then its members will have on average more than eight friends per capita.

Currarini, Jackson and Pin (2008) show that if one includes biases in both preferences and meeting rates then one can generate friendship patterns that exhibit homophily and are are also closely consistent with both of these additional patterns. In contrast, if one only includes a bias in friendships then one can generate the differences across groups in numbers of friends, but one cannot match the pattern of how homophily depends on group size. If one only includes a bias in the meeting process then one can match the pattern of how homophily depends on group size, but cannot generate the differences across groups in numbers of friends. Thus, at least in the context of their model, Currarini, Jackson and Pin (2008) find that both sources of bias must be present in order to match the empirical high school friendship patterns. Without such modeling it would not be clear that bias in both the meeting process and in preferences might be needed to generate observed patterns in the data. While this is just a step in understanding what factors contribute to homophily, it shows that the data contain important clues as to the forces behind homophily and also that modeling can help to understand the potential forces that might interact to generate homophily and observed network patterns.

[10] It is important to note that this is true of a normalized measure of homophily, since very large groups necessarily have some limit on how many friendships they can even have with other groups. The normalization divides the measure of homophily by how homophilistic a given group could be, and so provides a measure that is not intrisically biased by group size, but nonetheless ends up exhibiting strong group-size effects in the data.

5 Concluding Remarks

It is clear not only from the voluminous literature on the subject, but even from casual empiricism, that our behavior is substantially influenced by our friends and relatives. As our social networks exhibit significant patterns of homophily, it is important to understand how homophily translates into behavior and also the reasons behind homophily. Here I have discussed a few settings where we can model and understand the potential impact of homophily, including its impact on communication and learning, decisions to become educated, and social mobility. I have also discussed a starting model of network formation designed to study the roles of different sorts of biases in generating homophily. These analyses show that there are systematic and intuitive implications of homophily, and also that we can begin to sort out some of the factors that contribute to it. Developing a fuller understanding of how homophily and other social patterns influence behavior, and why our social networks exhibit such patterns, presents a healthy agenda for future research.

References

[1] Abramson, P. (1973) "Intergenerational social mobility and partisan preferences in Britain and Italy," *Comparative Political Studies* 6, 224-234.

[2] Berkman, L. and I. Kawachi (2000) *Social Epidemiology*, New York: Oxford University Press.

[3] Bisin, A. and T. Verdier (2000) "Beyond the melting pot: cultural transmission, marriage, and the evolution of ethnic and religious traits," *Quarterly Journal of Economics* 115, 955-988.

[4] Blalock Jr., H.M., (1982) *Race and Ethnic Relations* , Englewood Cliffs, NJ: Prentice-Hall.

[5] Blau, P.M., (1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, New York: Free Press.

[6] Calvó-Armengol, A. and M.O. Jackson (2004) "The Effects of Social Networks on Employment and Inequality," *American Economic Review*, vol. 94, no. 3, 426-454, June.

[7] Calvó-Armengol, A. and M.O. Jackson (2007) "Networks in Labor Markets: Wage and Employment Dynamics and Inequality," *Journal of Economic Theory*, Vol. 132, No. 1, 27-46.

[8] Calvó-Armengol, A. and M. O. Jackson (2008), "Like Father, Like Son: Labor Market Networks and Social Mobility," *American Economics Journal: Microeconomics*, forthcoming.

[9] Currarini, S., M.O. Jackson, and P. Pin (2007) "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica*, forthcoming.

[10] DeGroot M.H. (1974) "Reaching a consensus," *Journal of the American Statistical Association*, 69, 118 - 121.

[11] DeMarzo, P., D. Vayanos and J. Zwiebel (2003) "Persuasion Bias, Social Influence, and Unidimensional Opinions," Quarterly Journal of Economics, 118(3): pp. 909 - 968.

[12] French, J. (1956): "A Formal Theory of Social Power,' Psychological Review, 63, 181–194. F¨uredi, Z. and J. Koml´os (1981): "The Eigenvalues of Random Symmetric Matrices,' Combinatorica, 1, 233–241.

[13] Galea, S., A. Nandi, and D. Vlahov (2004) "The Social Epidemiology of Substance Use." *Epidemiologic Reviews,* 26: 36-52.

[14] Galeotti, A., S. Goyal, M.O. Jackson, F. Vega-Redondo, and L. Yariv (2007) "Network Games," forthcoming in the *Review of Economic Studies*.

[15] Golub, B. and M. O. Jackson (2007): "Na¨ive Learning in Social Networks: Convergence, Influence, and the Wisdom of Crowds,' Preprint, available at http://www.stanford.edu/ jacksonm/naivelearning.pdf

[16] Golub, B. and M.O. Jackson (2008) "How Homophily Affects Communication in Networks" arXiv:0811.4013v1.

[17] Harary, F. (1959): "A Criterion for Unanimity in French's Theory of Social Power,' in STudies in Social Power, ed. by D. Cartwright, Ann Arbor: University of Michigan Press.

[18] Ioannides, Y.M. and L. Datcher Loury (2004) "Job information networks, neighborhood effects and inequality," *Journal of Economic Literature,* XLII, December, 1056–1093.

[19] Jackson, M.O., (2008) "Social and Economic Networks," Princeton University Press.

[20] Jackson, M.O., (2008b) "Average Distance, Diameter, and Clustering in Social Networks with Homophily," in the *Proceedings of the Workshop in Internet and Network Economics (WINE 2008), Lecture Notes in Computer Science*, edited by C. Papadimitriou and S. Zhang, Springer-Verlag, Berlin Heidelberg, and arXiv:0810.2603v1 [physics.soc-ph],.

[21] Johnson, D. (1996) "Fathers presence matters: a review of the literature," *National Center on Fathers and Families*, working paper.

[22] Lazarsfeld, P.F, and R.K. Merton (1954) "Friendship as a social process: a substantive and methodological analysis," in *Freedom and Control in Modern Society*, ed. M Berger, 18–66, New York: Van Nostrand.

[23] Marsden, P.V., (1987) "Core Discussion Networks of Americans," *American Sociological Review* , 52, 122–131.

[24] Marsden, P.V., (1988) "Homogeneity in confiding relations," *Social Networks*, 10, 57–76.

[25] McPherson, M., L. Smith-Lovin, and J.M. Cook (2001) "Birds of a Feather: Homophily in Social Networks," *Annu. Rev. Sociol.*, 27, 415–44

[26] Moody, J., (2001) "Race, school integration, and friendship segregation in America," *American Journal of Sociology*, 107 (3), 679–716.

[27] Morris, S. (2000), "Contagion," *The Review of Economic Studies*, 67(1), 57-78.

[28] Piketty, T. (1995) "Social mobility and redistributive politics," *Quarterly Journal of Economics* 110, 551-584.

[29] Piketty, T. (2000) "Theories of persistent inequality and intergenerational mobility," in *Handbook of Income Distribution, Vol. 1*, Anthony B. Atkinson and François Bourguignon, eds., Elsevier: Amsterdam.

[30] Thomson, K. (1971) "A cross-national analysis of intergenerational social mobility and political orientation," *Comparative Political Studies* 4, 3-20.