DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

How Network Externalities Can Exacerbate Intergroup Inequality

(Article begins on next page)

How Network Externalities
Can Exacerbate Intergroup Inequality[1]

Paul DiMaggio
Princeton University

Filiz Garip
Harvard University

Abstract

We describe a common but largely unrecognized mechanism that produces and exacerbates intergroup inequality: the diffusion of valuable practices with positive network externalities through social networks whose members differentially possess characteristics associated with adoption. We examine two cases, the first to explicate the implications of the model, the second to demonstrate its utility in analyzing empirical data. In the first, the diffusion of Internet use, network effects increase the utility of adoption to friends and relatives of prior adopters. An agent-based model demonstrates positive, monotonic relationships, given externalities, between homophily bias and intergroup inequality in equilibrium adoption rates. In the second, rural/urban migration in Thailand, network effects reduce risk to persons whose networks include prior migrants. Using longitudinal individual-level migration data, we find that network homophily interacts with network externalities to induce inequality in migration rates among otherwise similar villages.

Introduction

We introduce a class of social mechanisms that influence intergroup inequality, usually by making such inequality larger and more durable, although potentially effacing inequality as well. These mechanisms come into play when (a) a good, service, or practice influences individual life chances; (b) that good, service or practice is charact- erized by network externalities, such that the benefits to adopters are higher, or the risks are lower, if persons to whom adopters have social ties have already adopted; and (c) actors' social networks are differentiated with respect to some characteristic associated with adoption. We illustrate our argument with case studies of two very different practices --- technology adoption in the U.S. and internal migration in Thailand – to suggest the scope within which we believe these mechanisms operate. The first case employs a computational model to explicate the implications of the argument. The second illustrates the argument's utility in addressing empirical puzzles.

This project emerged from our respective efforts to solve concrete empirical problems --- whether intergroup differences in Internet adoption would persist; and why similar Thai villages diverged over time in rates of rural-urban migration. These two empirical questions, we came to realize, entail analytically similar social mechanisms. In each case individual choices (to use the Internet or to migrate) are influenced by prior choices by members of ego's social network; and intergroup inequality (in technology

use or in rates of migration) is amplified to the extent that ego's social network is homo-
philous with respect to socioeconomic status. [2]

Our approach is consistent with the view that sociological explanation can be
advanced by identifyng "social mechanisms" that entail (a) goal-directed individual
actions and (b) consequent social interactions that (c) yield higher-level outcomes that (i)
are emergent (i.e., that cannot be recovered simply by aggregating individual actions that
combine to produce them) and that (ii) vary depending upon the initial social structure
(ordinarily depicted in terms of social networks) (Hedstrom 2005; Tilly 2006).   Despite
this perspective's growing prominence, it has played a relatively minor role in empirical
research on social inequality. Such research has often treated inequality as the aggregate
product of individual efforts to obtain useful educations, good jobs, and adequate
incomes, positing that people with similar initial endowments have similar experiences
that lead them to similar outcomes. Efforts to incorporate actors' structural locations
(position in networks or in the labor market) into such models ordinarily convert social
structure into individual-level variables in income-determination models (Lin, Ensel and
Vaughn 1981; Baron and Bielby 1980).

By contrast, we focus on the production of system-level inequality as a
consequence of individual choice under varying structural conditions.   Simon (1957)
introduced the idea that inequality reflects the depth of organizational hierarchies (itself a
function of organization size and span of control). White (1970) explored the impact of
vacancy rates on emergent properties of systems of inequality. Boudon (1973; Breen and

---

[2] We recognize that other less common mechanisms (*e.g.* heterophilous preference for alters with high-status characteristics one does *not* possess in populations with weakly or negatively correlated status parameters) could also combine with externalities to generate inequality.

Goldthorpe 1997) developed a competitive model of individual investments in schooling

to explain why intergroup inequality persisted in the face of educational expansion.

Western, Bloome and Percheski (2008) explored the impact of family structure and

homogamy on income inequality. In this paper, we develop a model of inequality

generation that draws on elements from three social-science literatures: on network

externalities; on innovation diffusion under conditions of interdependent choice; and on

homophily in social networks.

Network Externalities, Diffusion Models, and Social Homophily

In 1892, John F. Parkinson, an entrepreneur who sold hardware and lumber, became the

first telephone subscriber in Palo Alto, California. (This account is from Fischer 1992:

130-34.) Parkinson placed the phone in his business. A line to the Menlo Park telephone

exchange a few miles away connected him to other Bay Area firms.   By 1893, a realtor

and a butcher had joined him and soon a local pharmacy took a subscription, placing its

phone in a quiet room where customers could use it for a fee.   By 1897, Palo Alto had

nineteen telephone subscribers, including several – Parkinson, two physicians and two

newspaper editors – with telephones in their homes.

   It is no accident that early subscribers were businessmen and professionals for

whom the telephone was a means of staying in contact with suppliers, customers, and cli-

ents.   Why get a phone for social reasons unless you could call your family and friends?

Indeed, many years would pass before telephone companies recognized the telephone's

potential as an instrument of sociability (Fischer 1992: chapter 3).   Not until 1920 did

subscription rates approach 50 percent even in prosperous Palo Alto (*ibid.*: 141).

Naturally, telephones were more common in professional and business households,

whose members were more likely to have friends and relatives with telephone service. By 1930, blue-collar households caught up in Palo Alto, where many blue-collar workers were independent tradesmen whose clients had phones, but not in neighboring towns where most blue-collar residents worked in factories (*ibid.* 146-47).

Even after telephone service took off, inequality was tenacious. The United States did not approach 90 percent household penetration until 1970. As late as 1990, the poorest Americans often lacked service, which was close to universal at incomes of $20,000 or more (in 1990 dollars) but declined precipitously below that.[3] Race and ethnicity had independent effects on telephone service even controlling for income (Schement 1995). Something other than cost – perhaps interaction effects related to network composition -- must have driven these differences.

Tangible networks are not the only ones that exhibit positive externalities. The basic mechanism applies to purely social networks too. Consider the problem of socioeconomic and racial inequality in advanced-placement (AP) course enrollments in U.S. high schools (Kao and Thompson 2003; Klopfenstein 2004). Imagine a class with three hundred students, divided into three hierarchically arrayed SES tiers. Each student must decide whether to take an AP course. Each knows that AP courses will help gain admission to selective colleges and may reduce the time it takes to earn a college degree. Each also knows that to take advantage of this benefit, he or she will have to invest a lot of time in learning enough to pass the AP exam.

---

[3]After 1990, many Americans, including young, affluent ones, substituted cell phone service for land lines. In 2000, household penetration declined for the first time since the Depression, but apparently as a result of cell-phone substitution, not a decline in telephone access.

Several factors influence students' decisions: whether they expect to attend college, whether they qualify for admission to AP courses, whether team sports or after-school work pose competing demands, and so on. Based only on such considerations, imagine that the probability of choosing to enroll in an AP course is 80 percent in the top tier, 50 percent in the middle tier, and 20 percent at the bottom.

Now, consider the network externalities associated with taking AP courses. If your friends take the course too, you can study together, collaborate on homework, share a scientific calculator, or use a friend's high-speed Internet connection to prepare presentations. Such considerations increase the benefit of taking the course (you will learn more and enjoy socializing), reduce the rate at which you discount expected benefits (you will be more likely to pass the AP exam), and reduce your financial and temporal costs.

If friendships were distributed randomly – if a top-tier kid were as likely to have a bottom-tier friend as one in his or her own group – then, on average, half of each student's friends would plan to take an AP course. If kids in each tier had the same number of friends and effects were additive, more students would take AP courses than if decisions were made in isolation. And because lower-tier kids would know as many AP-course-takers as upper-tier kids, the outcome might be more equal than if choices did not interact.

In real high schools, however, friendships are never randomly distributed. Friendships are homophilous: similar kids hang out with one another more than with kids from other groups (Currarini, Jackson and Pin 2010; Quillian and Campbell 2003). If we build homophily into our example, externalities may exacerbate inequality, not reduce it.

Most positive effects from mutual influence will accrue to students in the top tier, because their friends disproportionately enroll in AP courses. Because bottom-tier kids have only a few friends in AP courses, externalities benefit them less. Combine these influences – big effects on advantaged students' choices, small ones on those of the disadvantaged – and initial gaps increase.

The point of these examples is to convey the intuitions behind and suggest the scope of applicability of models we describe more systematically below. Each example has three elements: a *choice* (purchasing telephone service, signing up for an AP test); *positive network externalities* (your telephone is more valuable if you can call your friends with it; you get more out of AP courses with less effort if your friends take them, too); and *social homophily* (which makes positive externalities redound more decisively to groups that possess an initial advantage). Combine these three elements – choice, externalities, and homophily (or, more generally, networks differentiated by factors associated with adoption) –in a diffusion process and the result often exacerbates social inequality by reinforcing pre-existing forms of privilege through differential adoption of a new product or practice. Below, we examine these elements more closely and discuss work that has informed our perspective.

Network Externalities

A product, service, or behavior possesses network externalities insofar as its value to an actor depends upon the number of other actors who consume the product or service or engage in the practice. The term derives from work in economics of communication and innovation, which has focused on a network's aggregate value as a nonlinear function of scale (and the difficulty of internalizing that value) (Katz and Shapiro 1985; Arthur 1989;

Shy 2001). By contrast, we focus on network effects *from the standpoint of individual decision makers* (to whom the value of action increases with network size).   DiMaggio and Cohen (2004) distinguish between *general* network externalities (when each user's benefit from adoption is influenced only by the number, and not by the identities, of other adopters) and *identity-specific* network externalities (where each user's benefit is a function of the identities of those who adopt).   Specific externalities may be *status-based* (*e.g.,* positive when users are higher-status than oneself and negative when they are lower-status) or *network-based* (when perceived benefit is a function of the number or percentage of *members of one's own network* who have already adopted).   Whereas economists ordinarily emphasize general network externalities, we are primarily concerned with *specific* externalities, because of their implications for social inequality. Thus in order to adapt the notion of network externalities to the study of social inequality, we *narrow* it by focusing upon identity-specific externalities but *broaden it* by applying it to practices as well as goods and services.

Networks are endemic to communications technologies and utilities: telephones, fax machines, e-mail clients, instant messaging, and social media like Facebook or Twitter. But as our AP-course example suggests, we believe that network externalities characterize a much broader range of choices: between public, secular private, and re-ligious secondary schools (kids like to go to the same schools as their friends and parents benefit from car-pooling and information-sharing); fertility (raising children is easier if your friends have kids [or if you make friends with people who do]) (Buhler and Fratczak 2007); divorce (the more divorces in one's social network, the more people are available

for support and companionship and the larger the pool of potential mates) (Booth, Edwards and Johnson 1991); and health-related behaviors (Christakis and Fowler 2008).

Social Homophily

Social networks are homophilous with respect to a trait to the extent that connected pairs of actors share that characteristic.[4] Since Lazarsfeld and Merton (1954) coined the term, studies in many societies and organizations have demonstrated pervasive homophily with respect to race and ethnicity, gender, age, educational attainment, religion, and other factors (McPherson, Smith-Lovin, and Cook 2001).

Rogers (2003:307) depicts homophily as a barrier to diffusion, arguing that where homophily is strong, adoption of innovations is more likely to be limited to elites. Several authors note a relationship between homophily and inequality in educational or occupational attainment, but have not placed this relationship in the context of an explicit model of diffusion or behavioral change. Buhai and van der Leij (2008) model occupational segregation as a result of social homophily combined with network effects on access to jobs. Quillian (2006) argues that racially homogeneous friendship networks reduce the academic achievement of capable Black students relative to that of equally able white peers. From the group perspective, homophily facilitates sustaining monopolies over scarce resources (Tilly 1999). From an individual perspective, however, it is precisely *heterophilous* ties (especially to persons of higher status) that enhance mobility (Granovetter 1973; Lin, Ensel and Vaughn 1981).

---

[4] We use " homophily" to refer to the tendency for networks to consist of persons similar in status for whatever reason, purposely conflating structural, choice, and inbreeding homophily because we focus on homophily's effects and do not model its causes.

We apply these insights to the system level to contend that homophily tends to increase intergroup inequality, whereas heterophily tends to reduce it. We argue that *the interaction between network externalities and social homophily is a critical mechanism for the production of social inequality* in access to novel goods and engagement in innovative practices. To the extent that adoption of such goods and practices contributes to individual success and well-being, this mechanism increases intergroup social inequality.

<u>Models of Diffusion</u>

In order to understand the implications of research on externalities and homophily for macro-level intergroup inequality, we must specify the mechanisms by which individual choices interact. The most useful instruments to this end are threshold models of diffusion. Threshold models constitute a subset of contagion models that emphasize the distribution of adoption thresholds, often depicted as a function of individual attributes and network characteristics, within an at-risk population.[5] An early exposition of ideas influencing such models appears in Liebenstein's (1950)   essay on interdependence in consumer demand. He distinguished between positive and negative externalities, which he termed, respectively, "bandwagon" effects and "snob" effects; and between cases in which ego's demand is a function of aggregate consumption and those in which some consumers influence ego's decisions more than others.   Leibenstein recognized that his analysis was limited by his reliance upon comparative statics, the assumption that "the

---

[5] We do not presume that choices are "rational" either phenomenologically or in effect. People may adopt courses of action unreflectively; or they may adopt them on the basis of poor information about costs, benefits, and risks.

order of events is of no significance" (*ibid.*: 187).   Progress came through relaxing this assumption.

Coleman, Katz and Menzel (1957) introduced dynamic models based on network-driven interdependence in their study of physicians' adoption of tetracycline. They found that adoption was brisker and penetration greater among physicians with many ties to other doctors than among less connected practitioners.   This difference, they argued, reflected choice interdependence among the well-connected, inducing a "snowball" or "chain-reaction" pattern as use of the new drug spread.

Building on Schelling's (1971) work on residential segregation, Granovetter (1978) produced a significant advance towards the class of models employed here.   His models applied to situations in which (a) agents must make a binary choice over a number of successive time points; (b) "The costs and benefits to the actor … depend in part on how many others make which choice" (1422); (c) each agent has a threshold at which she or he will choose to act; (d) agents respond more directly to actions of personal contacts than to those of strangers; and (e) outcomes (the proportion of agents choosing to act and, especially, whether this proportion reaches the critical mass necessary to sustain some collective behavior) depend upon the distribution of thresholds, not just the mean. Granovetter assembled all the parts necessary for the models developed here, with two exceptions: thresholds are exogenous and actors vary only in thresholds and network position.[6]

---

[6] Different mechanisms may account for similar forms of diffusion (Van den Bulte and Lilien 2001). Van den Bulte and Stremersch (*ibid.*) and Rossman, Chiu and Mol (2006) introduce innovative ways to use information on multiple innovations to distinguish among varying mechanisms.

We draw on this work, but diverge from many models by treating externalities as specific, in most conditions based on decisions by members of one's own social network and not by the number of adopters in general. Moreover, we innovate by adding status homophily as a key variable and by focusing on group-specific diffusion rates (and their implications for social inequality), rather than on rates for the population as a whole.

We combine insights from work on network effects, social homophily, and threshold models to argue that diffusion of goods and practices with strong, identity-specific network externalities through status-homogeneous networks tends to exacerbate social inequality. To do so, we present two cases. First, we explicate our argument using an agent-based model to predict group-specific diffusion paths for Internet use, varying the type of network effect and the extent of homophily. Second, we demonstrate the argument's empirical utility through an analysis of variation in rates of rural-urban migration in Thailand, to test the hypothesis that differences in status homophily among otherwise similar villages combined over time with network externalities to generate divergent migration patterns from small initial differences.

Table 1 here

These cases are very different (Table 1). Internet diffusion is a conventional instance of new-product adoption where network effects directly enhance the technology's value (*i.e.*, the value of the network to which the technology provides access) to the agent. Rural-urban migration is a longstanding practice that became much more widespread in Thailand during the 1980s. Network effects are indirect, in that connections to prior adopters are not themselves the source of value, but rather provide

information that increases returns to and reduces risks associated with migration. In the Internet case we track intergroup inequality based on educational attainment, income and race. In the migration case, we explain inequality in migration rates among villages. Nonetheless, the two cases share the requisite characteristics for the model to apply: network effects and interdependent choice; status homophily within networks; and sequential choice by numerous agents.

Case 1: The Internet: Transitional Inequality or Permanent Divide?

To understand network effects in the emergence of intergroup inequality in access to and use of the Internet, we employ agent-based models (Macy and Willer 2002). Agent-based modeling is particularly useful when theory-based predictions of individual choices are available but global results of interactions among choices are not readily apprehended through intuition or amenable to formal mathematical analysis, and when detailed longitudinal micro-data are unavailable. Such models enable us to explore the systems implications of behavioral mechanisms and the robustness of those mechanisms to changes in the values of key parameters. Agent-based modeling entails a tension between realism and generalizability. We base as many parameters as possible on available data, while varying two theoretically central variables as, in effect, experimental conditions. The purpose is *not* to *simulate* Internet diffusion (which would require more complex models) but to understand macro-patterns produced by particular micro processes.

The Problem

We begin with an empirical puzzle. Early in the Internet era, many believed that, by making useful information accessible at low cost, the Internet would enhance inform-ational, and thus social, equality (Cairncross 1997).   Others warned that the Internet could exacerbate inequality, because the wealthy and highly educated had resources that enabled them to employ the Internet more extensively and productively than their lower-status peers, thus widening the "knowledge gap" (and the gap in rewards that knowledge brings) observed for other media (Bonfadelli 2002).

When, after a decade of commercial availability, Internet use remained more com-mon among Americans with college degrees and relatively high incomes than among the less educated and the poor, many concluded that such inequality (the "digital divide") was an enduring problem. Critics of this view, however, pointed out that intergroup ineq-uality in adoption rates occurs whenever groups start at different baselines or reach critical mass at different times (Leigh and Atkinson 2001).

Figure 1 here

Consider the top panel of Figure 1, which depicts stylized adoption curves for two populations, A and B, with differences measured at three points in time (*x, y,* and *z*). The *x* axis represents time and the *y* axis represents the percentage of adopters. Group B begins the adoption process later than and reaches takeoff after group A.   If we compare early rates (time *x*), we see no inequality. Because A reaches takeoff before B, we then see increasing inequality, reaching a maximum at *y*, just before B reaches takeoff and just as A approaches equilibrium. Trend analysis between *y* and *z*, by contrast, shows declining inequality and (correctly) predicts convergence.

The bottom panel of Figure 1 presents a similar depiction of the adoption history of two other populations. Group C's trajectory is identical to that of Group A in the top panel. Group D gets into the game at the same point as group B (top), but adopts more slowly and plateaus at a lower level of penetration. As in the previous case, inequality peaks at point *y*. But it remains high even at point *z*. In this case, a pessimistic conclusion based on measurement at points *x* and *y* (intergroup inequality is a real problem) is more accurate than an optimistic prediction based on measurement at points *y* and *z*. The problem is this: At point *x* we cannot know if we are in the top or the bottom panel of Figure 1 – i.e., whether intergroup inequality will increase or decline – unless we understand the mechanisms driving diffusion.

How might we tell? One can examine trends in inequality by comparing penetration rates for different groups over this span (Fig.2). Data on Internet use have been collected by the Current Population Survey (CPS) from 1997 to 2007.[7] We focus on Internet use at home, rather than at work, school, or elsewhere, because home use ordinarily provides the greatest autonomy and opportunities for learning (DiMaggio, Hargittai, Celeste and Shafer 2004). Moreover, Internet use at home is consequential, boosting earnings net of technology use at work (DiMaggio and Bonikowski 2008).

Figure 2 here

---

[7] Between 1997 and 2003, the items were sponsored by the National Telecommunications and Information Agency, the federal agency responsible for implementing congressionally mandated universal telecommunications service. The 2007 items were asked as part of a module on school enrollment.

Using odds ratios (Fig. 2) as a measure, we see that some types of inequality

declined between 1997 and 2007.[8] Men's advantage over women was much greater

before 2000 than thereafter. Some movement towards greater equality occurred between

non-Hispanic whites and, respectively, African Americans and Hispanics; and between

the elderly and those in younger cohorts. By contrast, college graduates' advantage over

high school graduates without college persisted throughout this period.

Similarly, analysis of successive cross-sections with statistical controls demonstrated

declining net effects on adoption of gender and metropolitan (as compared to rural)

residence, but persistent effects of education and income (DiMaggio and Cohen 2004).

Such studies cannot tell us if the digital divide is a permanent or transitional problem,

however. For that we need a theoretical account of mechanisms that generate intergroup

inequality in technology use. We build upon such an account with an agent-based model

that highlights one critical mechanism: choice interdependence influenced by network

externalities under conditions of homophily.

The Model

Our model produces seven artificial worlds (i.e., experimental conditions), each

populated by 2,257 heterogeneous agents. The worlds vary in (a) the presence and type

of network externality and (b) the degree of homophily. We use them to explore how

inequality in Internet use between groups based on race (African-American or white),

income, and education varies over time as a function of the presence and type of network

effect (none, general, or identity-specific) and as a function of the strength of pressures

---

[8] By "odds ratio" we refer to the ratio of the odds of adoption of the first group to the odds of adoption for the second group: $(p_1/1-p_1)/(p_2/1-p_2)$.

towards homophily (i.e., homophily bias).   To ensure that the distribution of parameters

and associations among parameters are realistic, we base our agents upon 2257 African-

American and white respondents to the 2002 General Social Survey, which included

items on network size, race, education, and income.[9]

Each agent is assigned a *reservation* price – an adoption threshold at which it will

purchase Internet service. At the end of each period (roughly equivalent to one month),

each agent compares the price of Internet service to its reservation price, and either

adopts or declines to adopt. Agents may adopt because economies of scale drive prices to

or below their reservation price; because more members of their networks adopt (raising

their reservation price); or due to a combination of both. After each period, the

cumulative percentage of agents with different races, income levels, and educational

degrees who have adopted is reported and added to a graph that records 100 periods. Our

analysis focuses on the impact of two variables – the nature and extent of network

externalities and the degree of homophily -- on the extent of intergroup inequality in

adoption throughout the diffusion process.

We explicate model details in Appendix A (available online).   Two points are

critical: (1) adoption is driven by the relationship between the cost of Internet service and

---

[9] Respondents were first asked "Not counting people at work or family at home, about how many other friends, or relatives do you keep in contact with at least once a year?" We use the network size measure generated by the follow-up to this question: "Of these friends and relatives, about how many would you say you feel really close to, that is close enough to discuss personal or important problems with?" GSS reports income as a series of ranges: We treat income as uniformly distributed within each interval and randomly assign individuals to points in their distributions. Individuals who reported family incomes of $110K or more (about 10 percent) were randomly allocated to incomes up to $650,000 based on CPS data on actual income distributions in that range. (CPS is top-coded at $250K, but the mean income in the top-income category is reported to be $450K. We assume the range in the top category to be [$250K, $650K], producing a mean of $450K if income is uniformly distributed.) Race is either White or Black and education is measured in years.

each agent's reservation price (i.e., what the agent will pay) and (2) this relationship

varies across experimental conditions. Absent externalities, reservation prices are a

function of income and remain constant; therefore only reductions in cost can induce

adoption. In the general-externalities condition, reservation prices are affected both by

income and by change in response to the percentage of agents who have already adopted.

In this condition, new adopters affect all at-risk agents in the same way (although

susceptibility to this influence rises with income). In the five conditions characterized

by identity-specific externalities, reservation prices are influenced by income and by the

percentage of the members of ego's personal network who have already adopted. Each

additional adoption increases the reservation price for agents linked to the adopter and for

no one else. Five conditions with identity-specific externalities vary in the extent to

which network composition is biased towards homophily (Skvoretz 1990), ranging from

no homophily bias (i.e., network partners selected at random from the population)

through maximal homophily (all network partners chosen from a pool of those most

similar to ego based on a composite measure of similarity in education, income, and race,

with weights based on results from McPherson, Smith-Lovin and Brashears [2006]).[10]

Results

We argue that individual choices affect the reservation prices of others. In conditions

with identity-specific externalities, adoption cascades through micro-networks, retarded

by poverty (people without much money have lower reservation prices) and either

---

[10] The authors produced an Errata when GSS reported miscoding 41 responses, but the differences are
small and immaterial for our purposes.

retarded or accelerated by homogeneity within ego networks.   This intuition leaves open

questions about how local adoption pressures lead to global intergroup inequality. To

what extent do externalities accelerate adoption rates for groups with initial advantages?

To what extent does homophily retard adoption among less well endowed agents? What

level of homophily is required to have an effect? The models enable us to address such

questions.

We follow the adoption paths of 2,257 agents for 100 time-periods under seven

scenarios (1) no network externalities, (2) general network externalities, (3) identity-

specific network externalities and no homophily (h=0) (random nets), and (4-7) identity-

specific externalities with homophily bias set, respectively, to h=0.25, 0.5, 0.75 and 1.0.

For each condition, we undertake 1000 simulation trials, reporting the mean values for

each time period.

The analysis includes four steps: First, we look at the impact of the seven conditions

on overall diffusion trajectories, for the sample as a whole and for groups within it, focus-

ing on the rapidity of takeoff (i.e., the slope) and on equilibrium adoption rates.   Second,

we examine variation among the seven conditions in inequality between pairs of groups

defined on the basis of race, education, and income. Third, we undertake logistic regress-

ions predicting adoption at each period to examine the net impact of race, education and

income on individual choices throughout the process. Fourth, we use regression analysis

of 7000 trials to describe how externalities and homophily interact to affect group-level

rates and intergroup inequality.

Overall adoption rates.   Absent externalities, Internet adoption never reaches

critical mass: prices decline and adoption proceeds glacially, with only 10 percent

penetration after 100 periods (Fig. 3). Adoption rates in other conditions resemble a conventional sigmoid curve, rising slowly, then sharply, and eventually leveling off. Adoption under general externalities (i.e. when reservation prices decline as the proportion of adopters in the population increases) takes off between periods 30 and 40, rising sharply to the highest adoption level of any condition, nearly 65 percent at equilibrium. Introducing specific externalities *without* homophily scatters the effects of new adoptions throughout the population, rather than applying them to everyone as for general externalities. As a result, some agents are marooned in non-adopting networks, with adoption slower (the takeoff begins after period 40) and lower at equilibrium.

Figure 3 here

Combining network externalities with homophily induces more rapid takeoffs, but lower penetration overall. When homophily is modest (h=0.25) the trajectory is similar to that for the general-externalities condition through period 55, then plateaus more quickly. Strong homophily (h=1.0) stimulates early adoption sharply, as adoption cascades among homogeneous networks, inducing takeoff around period 20. It limits the reach of network effects, however, producing an equilibrium adoption level below that of any other condition with externalities. Middling homophily levels (h=0.50 and h=0.75) yield middling results. The rule, then, is that the stronger the bias towards homophily in a process driven by identify-specific externalities, the steeper the slope and the lower the equilibrium penetration rate (shown in the inset figure in the lower right), with differences in penetration across *h* small but monotonic.

Forms and Degrees of Inequality. We begin with inequality in rates of Internet adoption by income and then address inequality based on education and on race. We divide agents into three equal-sized income classes, high (>$55,000), medium ($30,000-$54,999), and low (<$30,000). Because the model includes a direct income effect *and* (where applicable) interactions between income and network effects, income has a very strong influence on adoption. In analyses available upon request, the top income class has an equilibrium adoption rate of more than 90 percent in all conditions with extern-alities, with the highest rates associated with the highest level of homophily. Rates for the middle class range from 63 to 70 percent, with rates highest for general externalities and descending as homophily bias increases. Equilibrium rates for the lowest income class decline from 26 percent with general network effects and 23 percent with identity-specific externalities but random networks (no homophily) to just 18 percent with specific externalities and complete homophily. In other words, given identity-specific network externalities, homophily increases adoption among the most prosperous and suppresses adoption among the least privileged.

Figure 4 depicts the ratio of the odds of adoption for the highest to the odds for the lowest income group over the course of diffusion. Differences are sharpest when identity-specific externalities combine with maximum homophily bias (an odds ratio at equilibrium of almost 100:1); in that condition, just after period 20, when all of the high-income agents have adopted but hardly any of the low-income agents have done so, the odds ratio reaches 300:1, dropping as low-income adoption increases, and leveling off at period 40.

Figure 4 here

Given identity-specific externalities, increments in homophily enhance the slope and the equilibrium rate of adoption for the highest-income group and increase that group's advantage at equilibrium over the lowest-income tertile. All effects are mon-otonically related to homophily bias: The more homophily, the more rapidly inequality increases, the higher the rate at which it peaks, the greater the gap between the peak and the equilibrium level, and the higher is inequality at equilibrium. Inequality in the general-externalities condition is similar to that at moderate levels of homophily bias (h=0.5), but the peak and variance are considerably smaller. It follows that estimates of intergroup inequality early in a diffusion process are often poor indicators of long-term differences.

Educational inequality. There are four education subgroups: Agents with college degrees or more; agents with some college who did not graduate; high school graduates who did not attend college; and high–school non-completers. Unlike income, education has no direct effect upon reservation price; by design, its sole influence comes from the correlation of education with income. Therefore inequality among different education levels is less extreme than among income levels. In the general-externalities condition (where penetration is greatest), equilibrium rates are just over 80 percent for college graduates, almost as high for agents with some college, 61 percent for high school graduates without college, and 42 percent for agents lacking high school degrees (graphs available upon request). Unlike the case for income, general externalities are associated with relatively low equilibrium levels of inequality, slightly below that observed with random networks under the identity-specific externalities condition. Otherwise, the

results for income and education are very similar: the greater the tendency toward

homophily, the steeper the slope of odds ratios between education levels, the more

marked the difference in equilibrium rates by class, and the greater the difference

between peak and equilibrium inequality (Fig. 5).

Figure 5 here

Racial inequality. The system has two races, Blacks and Whites. Race does not

affect agents' reservation prices directly, but is associated with them through its correlat-

ion with income.   Income is lower for Blacks than for Whites. Equilibrium diffusion for

whites (graphs available upon request) is highest (68 percent) under general externalities,

but very similar (63-65 percent) under all specific-externality conditions.   Equilibrium

adoption rates for Blacks, although lower in all conditions, are, as for whites, highest (50

percent) under general externalities. In contrast to whites' rates, however, rates for

Blacks decline monotonically with homophily under identity-specific externalities, from

44 percent with random nets to 39 percent when homophily bias is at its maximum.

Figure 6 here

Racial inequality, measured as an odds ratio, is lowest (about 2.3:1) under general

externalities (Fig. 6). As with other forms of inequality, under specific-externality

conditions the slope of inequality, peak inequality, equilibrium inequality, and the

difference between peak and equilibrium inequality all increase monotonically as homophily bias increases.[11]

Net Effects of Income, Race, and Education. Thus far we have compared diffusion trajectories and estimated intergroup inequality separately for groups based on each individual attribute. Here we report results of analyses using repeated cross-sectional regressions to plot *net* effects of income, race, and education on adoption over time, under different conditions.[12]

Logistic regressions were run to predict individual-level adoption under each of the seven conditions (with results for each period averaged across 25 trials per condition) from periods 10 through 80 (Figs. 7-9).[13] We focus on the relative magnitude of these estimates across conditions. A unit increase in logged income -- for example, an increase from $20,000 to $54,364 -- raises the equilibrium odds of adoption between 80 percent (given specific externalities and a random network) to 120 percent (with specific externalities and maximum homophily bias). Each extra year of education increases the odds of adoption at equilibrium by from 1 percent (general externalities) to 7 percent (specific externalities with maximum homophily bias) in Figure 8. Being White increases odds of adoption by 11 percent (general externalities) to 50 percent (specific externalities with maximum homophily) compared to being Black.

----

[11] Networks generated by our model are less racially segregated than those reported in the General Social Survey because, as noted above, we used a composite social-distance measure to generate the networks characterized by homophily.

[12] Separate models were run with an added control for network size, but this did not materially affect the results.

[13] Twenty-five trials were used (after exploratory analyses established their results did not vary from those using larger numbers of trials) due to computational time and capacity constraints. Results were averaged only through period 80 because the models reached equilibrium by that point.

Figures 7, 8 and 9 here

Inspecting differences among conditions reinforces inferences from descriptive results presented earlier. Under the identity-specific–externality conditions, net effects of income, education and race all increase monotonically with homophily bias, providing further evidence that identity-specific externalities and homophily interact to exacerbate intergroup inequality.   Under general externalities, net effects of race and education are essentially null and income effects are moderate, comparable to their impact at the median of homophily bias. Education effects gather strength over the course of the diffusion process, especially at low levels of homophily. By contrast, especially at high levels of homophily bias, race and income effects weaken over time.

We also modeled adoption at equilibrium with interaction effects (table available upon request).[14]   As before (and necessarily given the design of the model), income is highly significant in every condition, and becomes more so as homophily increases.   In models without interactions, race and education reach statistical significance only in conditions with identity-specific externalities; the effects of race peak when $h=0.75$ and of education when $h=1.0$. Thus the combination of specific externalities and homophily induces *net* increases in inequality based on variables merely correlated with income, even though the model does not specify a direct effect of these on adoption choices; and the effect strengthens as homophily increases.   Positive interactions between being white and income and between education and income in most conditions with identity-specific

---

[14] In order to calculate standard errors, we did not average across 25 runs (as we did for the coefficients plotted in Fig. 10), but used results from the implementation of each model with the median coefficient for education (as those coefficients were the most variable across runs).

externalities suggest that such network effects not only exacerbate each form of intergroup inequality, but also *compound* effects of different advantages.

   Statistical impact of externalities and homophily on inequality. Finally, we examine statistical effects of externalities and homophily on diffusion trajectories and on inequality at equilibrium.   Cases are 7000 runs of the model, 1000 for each of seven conditions.   Independent variables are dummies for each of six conditions. (Identity-specific externalities without homophily is the reference category). Dependent variables are equilibrium adoption levels (global and for specific population subgroups) (Table 2) and indicators of intergroup inequality (logged odds ratios comparing paired subgroups) (Table 3).

Tables 2 and 3 here

   Results confirm qualitative interpretations from prior analyses: no takeoffs without externalities[15], highest diffusion levels under general externalities, negative effects of homophily (under specific externalities) on overall adoption levels, and positive effects of homophily on inequality at equilibrium. These analyses demonstrate that specific externalities increase inequality in adoption rates by significantly reducing adoption by less privileged groups, while exerting modest, often positive, effects on more privileged groups' adoption.   Moreover, the greater the status distance between groups, the more homophily exacerbates inequality. Income effects are strongest for the top

---

[15] We operationalize takeoff as the first time point when at least 1 percent of the population adopted Internet since the previous period. According to this definition, takeoff for adoption (i) does not occur by T=100 with no network externalities, and occurs (ii) at T=37 with general externalities, and (iii) between T=36 and T=20 with specific externalities where homophily ranges from 0.25 to 1.

tertile as compared to the bottom tertile, with the effect on top-to-middle inequality much smaller and on middle-to-bottom even smaller, albeit significant. Similarly, the effects of increases in homophily bias on inequality between college graduates and high-school nongraduates exceed the still significant effects on inequality between other subgroups.

Conclusions to Technology Adoption Case

Although the model's purpose is to clarify the behavior of a particular kind of diffusion process by highlighting and varying key dimensions of that process, and *not* to *simulate* the empirical details, it does capture some important dimensions of Internet diffusion. Most important, as the model would predict, intergroup inequality in Internet access has persisted as group-specific penetration rates have plateaud, dashing hopes that the digital divide would vanish on its own. This is true even though our model exaggerates income's impact on adoption compared to that of education, and neglects the fact that adoption reflects not just specific network effects (the appeal of communicating with one's friends and kin through e-mail, messaging, and social media) but also general externalities (the desire to benefit from such sites as Wikipedia or eBay that do not involve interacting with friends but which become more valuable as the overall user base increases in size). When we modified our model (results available upon request) to capture a mixed-mechanism process (both specific and general network effects) by adding a term that permitted education (as a proxy for demand for information) to interact with overall adoption levels to reduce the reservation price, the effects of education relative to income increased, but the qualitative results were unchanged.[16]

---

[16] A realistic simulation would more adequately model *disadoption,* which tends to reinforce inequality as users at the edge of the technology's niche drop out more than others (Popielarz and McPherson1995). A

To summarize: Like other technology-diffusion models, ours demonstrates that network effects steepen the slope and increase the extent of diffusion.   Our results also demonstrate a less often appreciated effect: When externalities are specific to potential adopters' own networks – when, as in the case of most communication devices, ego cares less about how many people have adopted than about whether her or his friends and family have adopted – externalities interact with homophily to exacerbate the effects of social inequality on adoption. Our analyses help resolve the question with which this section began by demonstrating that under almost any range of plausible conditions, and barring dramatic public-policy intervention, intergroup adoption rates will not converge. This conclusion reinforces inferences drawn from other methods, and enables us to understand more clearly the mechanisms that perpetuate inequality.

Case 2: Internal Migration in Thailand

Our second case demonstrates the utility of the argument for solving empirical puzzles. Whereas the Internet case used a formal model to explore change over time in inequality among groups defined by race and socioeconomic status in a national population, our second case draws on empirical data collected over two decades to explore change over time in inequality among villages in a region of Thailand.   Whereas the point of the first case was to document the impact of variation in externality strength and homophily on intergroup inequality, the goal of the second is to explore the tendency for homophily to

version of our model that included group-specific disadoption rates (results available on request) generated lower equilibrium adoption rates and produced greater inequality, but otherwise retained the qualitative features reported earlier. A realistic simulation would also take account of the role of institutions like schools and workplaces in introducing people to technology (DiMaggio and Bonikowski 2008), public policies aimed at reducing SES differentials in access, endogeneity of networks to technology use, change in the Internet (including significant quality improvements) over the diffusion period, interdependence among technologies and spatial heterogeneity in cost and availability. It was not our goal to produce such a model, which would obscure analysis of the mechanisms our study seeks to understand.

generate inequality (in the sense of high-variance outcomes) where network effects are constant and socioeconomic differences at both the individual and village levels can be rigorously controlled. Rather than attempting to provide a full explanation of rural-urban migration, analyses of the Thai data focus on effects of ego networks and homophily on aggregate migration *at the village level*, attending to other factors only as necessary for model specification.

Hypotheses

Migration flows may begin for many reasons: attempts to increase individual or household income, demand for low-wage workers in destination regions, or deteriorating economic conditions in sending communities. The conditions *initiating* migration flows, however, may differ greatly from those that *sustain* them. Although economic factors may continue to stimulate migration, prior migration itself may produce new conditions in origin communities that then act as independent causes of migration. As ties connecting individuals in origin to migrants in destination spread, they serve as conduits of information or assistance for potential migrants (Massey 1990). Remittances from migrants change the income and land distributions within origin communities, propelling other individuals to migrate as well (Stark and Taylor 1989). Migration may become culturally sanctioned, taking on a symbolic meaning as a rite of passage or affirmation of identity (Levitt 1998). Thus migration flows are partially decoupled from the conditions that initiate them, through what Massey (1990) calls 'cumulative causation of migration'. Studies from different contexts have illustrated this process (Dunlevy 1991; Massey and Espinosa 1997).

Recent research, however, points to spatial and temporal heterogeneity in migration outcomes not readily explicable within the cumulative causation framework (Curran et al. 2005; Fussell and Massey 2004). Recent evidence from Thailand, based on data used here, depicts cumulative causation but also reveals dramatic heterogeneity in migration patterns across communities. Whereas migration to urban places reaches mass levels in some villages, it lingers at low levels in others (Garip 2008). *The heterogeneity in migration patterns presents a puzzle that current accounts of migration cannot explain*.

We apply the theoretical framework developed in the first part of this paper to explain this heterogeneity. First, we contribute to cumulative causation theory by focusing on the interconnectedness of migration decisions of socially-related individuals.[17] Prior migrants, we contend, provide information and assistance that reduces the risks and increases the anticipated benefits of migration. We model migration as a diffusion process with network externalities, where the prior adoption of migration in a population increases the probability of first-time migration for remaining non-adopters. We further argue that potential migrants' choices are more likely to be influenced by prior migrants to whom they are socially connected. We identify household, village and regional ties as three broad categories of social relations that channel diffusion of migration.

Second, we investigate how social homophily moderates the impact of network externalities, and creates differential migration diffusion patterns across communities. Using time-varying information on characteristics of all individuals (within an age range)

---

[17] Van de Rijt's (2009) study of the diffusion of "assimilation" among migrants provides the only formal model of choice interdependence in the context of migration.

in the study villages, we produce measures of homogeneity that serve as proxies for homophily. (As described below, we base this decision on high correlations between overall homogeneity and ego-network homogeneity [in agricultural mutual-assistance networks] for the one year for which reliable measures of both variables are available.) We focus on how aggregate change in migration rates over time is associated with village-level homogeneity.

We test four hypotheses about the relationship between network effects, externalities, and migration, each subject to *ceteris paribus* conditions and modeled with extensive controls.

The first hypothesis is based on the expectation that migration in earlier time periods reduces risks and raises expected rewards of current migration. If this is the case, then:

> *Hypothesis 1: The greater the number of prior migrants, the higher the migration*
> *rate.*

In testing this hypothesis we replicate results of previous research demonstrating positive network effects on migration (Massey and Espinosa 1997).

The second hypothesis addresses the effect of social homogeneity (viewed as a proxy for network homophily) on migration rates. Previous research suggests that social homogeneity reduces the diversity of information entering a population, limiting members' ability to learn from prior experience (Garip 2008). Homogeneity limits the number of channels through which migrants are recruited, the diversity of positions they occupy in urban destinations, and the diversity of urban social ties to which their experiences gives other villagers access. Thus:

Hypothesis 2: The more homogeneous the population, the lower the rate of

migration.

Note that this hypothesis is consistent with results from the Internet-diffusion case, which

demonstrated that diffusion levels decline as social homophily increases.

The third hypothesis addresses the role of social homogeneity (again, proxying

homophily) in moderating the strength of network effects.   We anticipate that even if

(per hypothesis 2) homogeneity reduces migration rates by limiting access to available

information, homogeneous villages circulate whatever information prior migrants provide

more efficiently than more heterogeneous places. As Putnam (2007) has argued, social

heterogeneity may reduce social cohesion. This, in turn, may reduce the velocity and

scope of information flow, concentrating network effects on subsets of the population.

By contrast, homogeneous populations may experience benefits of prior migration more

quickly and broadly.[18] Moreover, persons in homogeneous populations may be more

likely to identify with prior migrants, and hence benefit more from the latter's

experiences.[19] If this is the case:

Hypothesis 3: The more homogeneous the population, the stronger the effect of

prior on current migration).

The fourth hypothesis applies to the village, rather than the individual, level of

analysis.   In the Internet case, we found that homophily exacerbated intergroup

---

[18] Alternately one might argue that homogeneity could interact negatively with the number of prior
migrants if each migrant brings back less useful information (as per Hypothesis 2), in that relatively
heterogeneous village networks would connect potential migrants to prior migrants who could provide in-
formation about a more diverse set of destinations or work options. We opt for Hypothesis 3 because we
expect heterogeneity's positive impact to be captured by its direct effect, rather than by its interaction with
prior migration.

[19] We thank an external reviewer for pointing this out.

inequality. By analogy, we anticipate that among Thai villages, social homogeneity will increase variance in outcomes by amplifying the effects of initial differences.

Hypothesis 4: At the village population level, variance in migration rates will be higher among homogeneous villages than among heterogeneous villages.

Data and Methods

Data come from the Nang Rong Survey of 22 migrant-sending villages between 1972 and 2000, when Thailand's economy shifted from agriculture to manufacturing, propelling migration from rural areas to Bangkok and other urban destinations.[20] Nang Rong, a relatively poor district in an historically poor region of Thailand, has been an important source of migrants to urban centers, primarily Bangkok.

Migration flows from this area once consisted mostly of seasonal migrants seeking alternative livelihoods during droughts that preceded the monsoon rains (Phongpaichit 1990). This seasonal character began to change in the mid-1980s. As growing manufacturing exports increased labor demand in Bangkok and its provinces (Bello, Cunningham, and Poh 1998),[21] rural migrants in their teens and early twenties, mostly from Northeastern Thailand (including Nang Rong), flocked to urban factory, construction and service jobs at unprecedented rates. Although most migration remained temporary, durations increased.

---

[20] Nang Rong Survey is a collaborative effort between the University of North Carolina and Mahidol University, Thailand. More information is available at: http://www.cpc.unc.edu/projects/nangrong.

[21] This growth, and its centralization in Bangkok, accompanied transitions in the Thai political system from an authoritarian monarchy to a semi-democratic bourgeois polity dominated by the entrepreneurial class (Somrudee 1993). Business's increasing political power exacerbated developmental inequality among regions through the 1980s. Dramatic economic growth in the 1980s and 1990s disproportionately benefited Bangkok, and the most dynamic sectors of the Thai economy remained concentrated around Bangkok, employing labor from the rural periphery (Mills 1999).

In 1984, the first year of the Nang Rong Survey and roughly the beginning of the period of dramatic growth in rural-urban flows, the 22 study villages looked similar. Ranging from 340 to 380 kilometers in distance from Bangkok, they shared common economic features. Rice cultivation was the primary economic activity: 87 percent of respondents listed farm work as their main occupation with little variation across villages. Only one village had a nearby factory to provide an alternative to agriculture work. None of the villages had phone lines and only seven had electricity. The villages were similar in size, the number of households ranging from 70 to 154 with a median of 124; and in ethnic composition (inferred from language spoken in household). In fifteen villages more than 90 percent of households primarily spoke Central Thai; in the others, the majority spoke either Lao or Cambodian (Khmer). Language homogeneity, measured on a scale from 0 to 1, averaged 0.90 across villages. Given initial similarities (and controlling statistically for observed differences), one would expect similar rural-urban migration patterns across villages in the following years.

Data. We observe migration patterns using three waves of data, collected in 1984, 1994, and 2000. The 1984 wave was a census of all village residents, including inform-ation on demographic characteristics, household assets and village institutions.[22] The 1994 wave followed all 1984 respondents still living in the original village (along with new village residents), and included a retrospective life-history component collecting in-formation about migration experiences of all individuals between aged 13 to 35 (the group at most risk for migration) from age 13 onward. The 1994 survey also interviewed migrants in the four major destinations (Bangkok, the Eastern Seaboard, Korat and

---

[22] Surveys were conducted in 51 villages, but we use data only from the 22 villages for which migrants were followed up in destinations in subsequent waves.

Buriram), reaching about 70 percent of eligible migrants (Rindfuss et al. 2007). Previous

research establishes that missing information from unlocatable migrants is unlikely to

bias results on migration patterns (Garip and Curran 2009). The 2000 wave followed the

1994 respondents and added new residents to the database. It also collected retrospective

life-history information about migration experiences starting at age thirteen from all

village residents aged 18 to 41 and from migrants in the four destinations.

Thus our results apply to individuals who were between the ages of 13 and 35 in

1994. For a 35-year-old migrant in 1994, for example, we observe annual migration

moves retrospectively from 1972, when the individual was 13, to year 2000, when he or

she turned 41. As a result, the age composition of the sample at risk of migrating varies

over time, containing only 13 year olds in 1972, 13 to 25 year olds in 1984, and 13 to 35

year olds in 1994. (We ran our statistical models with year dummies to take into account

the varying age structure over time. The results, available upon request, did not change.)

The retrospective life histories span the period 1972 to 2000, allowing us to model

the diffusion of individual migration over time.[23] Significant variability in these diffusion

patterns among villages provides a unique opportunity to discover the social mechanisms

that create heterogeneous migration outcomes among all villages.

Measures. Because our interest is in the diffusion of practices, we focus on *first*

migration moves and define migrant status in each year with a binary indicator that

equals 1 if the index person lived away from Nang Rong for more than two consecutive

---

[23] Left-censorship of retrospective data is not a problem as urban migration in this region took off only in the 1980s: as late as 1984, the percentage of migrants among all village residents ranged from 1 to 7 percent in the study villages (Korinek, Entwisle and Jampaklay 2005; Rindfuss et al. 2007).

months for the first time.[24] This criterion restricts migrant status to labor migrants, excluding persons who make trips of brief duration. We operationalize village-level homogeneity with respect to two dimensions, education and occupation, likely to shape individuals' migration opportunities, and the nature of information they can provide to household or village members once they migrate.[25] There are six occupational categories: student, farm work, factory, construction, service, and other. Homogeneity is measured as the complement of two diversity measures: the logarithm of variance for the ordinal attribute of education, and Shannon's entropy for the categorical attribute of occupation. In raw form, these diversity measures capture village-level heterogeneity rather than homogeneity, so, after normalizing them to [0,1] range, we compute homogeneity as 1-diversity.[26]

In our analysis, we use village-level measures as a proxy for homophily, because village homogeneity is available in each year, whereas homophily can be estimated reliably only for one year (2000, when data were collected on rice-harvesting networks among households). To assess the relationship between homogeneity and homophily (estimated based on homogeneity within ego-networks), we treated all individuals in connected households as also connected; measured educational and occupational homophily in each individual's network with the logarithm of variance and Shannon's

---

[24] The Nang Rong life history survey defines a "migrant" as someone who has had a spell of greater than two consecutive months living in an urban place.

[25] Education and occupation, when coded an ordinal scale (increasing in degree for the former, and in average earnings for the latter), are positively associated (polychoric correlation coefficient=0.31). Polychoric correlation uses the qualitative knowledge on the ordering of categories, and provides a more accurate measure of association in the case of ordinal variables compared with the Pearson's correlation.

[26] Results are robust to alternative diversity measures. Using an index of qualitative variation (Lieberson, 1969) instead of Shannon's entropy for the categorical occupation measure does not affect results, and using a Gini coefficient instead of log variance for the variable leaves the results unchanged.

entropy, respectively; took the mean ego-network homogeneity (with respect to education and occupation, respectively) for each village; and correlated these with village-level homogeneity on each dimension. Village-level homogeneity and and ego-network homogeneity were correlated at 0.82 for education and 0.55 for occupation, thus justifying the inference that effects of homogeneity may reflect mechanisms in which homophily plays a critical role.

To capture effects of network externalities, we define households, villages and the entire Nang Rong region as possible diffusion channels for migration.[27] For each individual, we count the number of prior migrants in the household (excluding the index individual) and village (excluding the index individual's household) through the previous year. We also compute a global count of prior migrants in all villages in Nang Rong (excluding the index individual's village). By comparing the effects of household and village migration rates, on the one hand, and global migration rates on the other, we assess the relative importance of specific and general network externalities for inducing migration flows.

To compute homogeneity and migration-experience indicators, we use annual information from the age group included in the life history surveys rather than data on the whole village population captured in the 1984, 1994 and 2000 censuses.[28] By focusing upon the population at risk of migrating, we implicitly assume that the characteristics of

---

[27] We also considered potential connections across villages. We grouped villages using various criteria: (i) sub-district membership, (ii) 5 km periphery, (iii) 10 km periphery, (iv) bus route connections, (v) worker-hiring relations, and (vi) temple-sharing relations, and assessed the impact of the number of migrants in ego's network from the village cluster (excluding those from ego's own village). In all cases, the effects of these village-group indicators were negligible and insignificant and the effects of household, village and Nang-Rong level networks remained robust (results available upon request). Therefore, we conclude that household, village and district ties delineate the main diffusion channels for migration.

[28] We use information from migrants in destinations, as well as residents in origin villages, to compute the village-level homogeneity measures to ensure that these measures are not endogenous to migration.

one's peers who are also at risk of migrating are the only ones that are consequential for

the migration decision.[29]

Finally, to rule out alternative explanations, we control for several characteristics

sometimes associated with social and economic incentives for migrating. For each

individual, we include indicators of age, sex, education and marital status. At the

household level, we measure household members' landholdings, both as a proxy for

household's wealth and as a direct influence on migration. Also included is the household

dependency ratio (ratio of those older than 64 or younger than 15 to those who are

neither). At the village level, we control for the presence of schools, newspaper reading

centers, factories, rice mills, the availability of electricity, and population.[30] (See Table

B1 in Appendix B for descriptive statistics.)

Analytic Approach: *Individual-level Analysis.* To model the individual-level

diffusion of migration, we use event-history methods, which allow us to account for the

timing of migration and censoring in data (i.e., not every individual migrates before the

end of the study period) (Strang 1991). An individual's decision to migrate is modeled as

a process occurring over time, where the numbers of prior migrants in household, village,

---

[29] Our fieldwork suggests that this assumption is reasonable. In focus group interviews that one of the authors conducted in the study villages in 2005, most migrants reported receiving help from their peers rather than their elders. Help from peers was more useful because peers were more likely to have worked in the service or factory jobs most available to new migrants. If we relax this assumption and measure experience and homophily for all village residents, albeit cross-sectionally based on the censuses in 1984, 1994 and 2000, we obtain very similar results, because measures based on the population at risk are highly correlated with those based on the whole population. The correlation between the number of migrants in the whole population and those in the at-risk population is 0.43 in 1984. The correlation between the education (occupation) homogeneity measures in the full versus the restricted sample is 0.78 (0.90) in 1984, 0.72 (0.86) in 1994 and 0.89 (0.71) in 2000.

[30] The life history information goes back to 1972, but household and village surveys start in 1984. We assume that the household and village characteristics (dependency ratio, land, village population, factories and rice mills) are at their 1984 values from 1972 to 1983. If we relax this assumption and restrict our analysis to data from the 1984-2000 period, our results remain unchanged. (Results available upon request).

or Nang Rong are time-varying factors affecting the hazard of adoption. We use the

proportional hazards model,

$$\log h_i(t) = \alpha + \beta x_i(t) + \gamma z_i \qquad (1)$$

in which the dependent variable is $h_i(t)$, the hazard of first migration at time t for

individual i, α is a constant, $x_i(t)$ is a vector of time-varying covariates characterizing

individual i, and z_i is a vector of time-constant covariates for individual i . β is a vector

of parameters describing effects of time-varying covariates on the hazard rate, and γ is a

vector of parameters describing effects of time-constant variables on that rate.[31]

From our theoretical perspective, this model's most important component is the

vector $x_i(t)$, which includes the number of adopters in an individual's household and

village (excluding the household), and in the Nang Rong region (excluding the village).

We use these indicators to test the first hypothesis, which posits positive effects of prior

on current migration. By comparing the magnitude and significance of the corresponding

β parameters for these indicators, we can also evaluate the relative importance for

migration of different foci of social relations (Feld 1981). If household ties are the sole

relations structuring the diffusion of migration, individual adoption rates will rise with

the number of prior adopters in the household. If village ties are important influences on

adoption, an individual's choice to migrate will also increase with the number of villagers

who have already migrated.

Village educational and occupational homogeneity are also of keen interest,

varying over time and also part of $x_i(t)$. To test the second and third hypotheses -- which,

respectively, predict lower migration rates and stronger network effects as village-level

---

[31] A discrete-time hazard model (with a logit transformation and year dummies) leads to similar results (results available upon request).

homogeneity increases -- we add the homogeneity measures as well as interaction terms between these measures and the number of prior migrants in the village. We also include the controls described above.

To evaluate alternative explanations, we employ several robustness checks (available upon request). First, we introduce village fixed effects to ensure that results are not driven by unmeasured village attributes. Second, we use bootstrapping (with re-sampling at the village-level) to test the sensitivity of results to the exclusion of specific villages. Third, we include year effects to control for unmeasured idiosyncratic shocks over time. None of these altered our results.

Analytic Approach: Village-level Analysis. To evaluate the fourth hypothesis, which posits more divergent migration rates among more homogenous villages, we use regression analysis to compare the variance in migration patterns among villages categorized as high, medium or low in educational and occupational homogeneity. We anticipate that the greater degree of homophily in more homogeneous villages will amplify small initial advantages, causing variance to grow.

Results: Migration as a Diffusion Process with Specific Network Externalities

We model individuals' first migration as a diffusion process, where prior adoption rates among social contacts influence future adoption; and explore the village-level implications of these network-driven processes over time. The empirical results both confirm expectations based on the agent-based model of Internet diffusion and suggest that network externalities and homophily may generate inequality (in this case, between villages) even in the absence of differences in initial endowments.

Trends in migration. We begin by observing the cumulative number of migration events between 1972 and 2000 in Nang Rong. The cumulative number of adopters (i.e. migrants) approaches an S-shaped curve over time (Fig. 10), while the frequency of new adopters per year approximates a normal, bell-shaped curve (Fig. 11). Early in the diffusion process, relatively few individuals adopt in each time period. Gradually, the adoption rate accelerates until many of the region's residents have migrated.

Figures 10, 11 and 12 here

The uniform path displayed in Figure 10 conceals variation across villages. This variation becomes apparent in the box plot in Figure 12, which demonstrates the distribution of cumulative migrants as a percent of village population across villages for each year between 1972 and 2000. Not only is there variation in the cumulative percentage of migrants across villages in each year, but this variation grows over time. (Similar patterns are observed for annual, rather than cumulative, percentages.) Due to this persistent and increasing between-village inequality in migration rates, prior work suggests, some villages in Nang Rong become integrated into the urban economy, whereas others remain isolated (Garip 2008).

If diffusion of migration were a population-level process operating within Nang Rong, we would expect similar adoption rates across villages, especially given the villages' economic and demographic similarity. The substantial variation in migration rates suggests that diffusion channels are specific to villages and not region-wide.

Network effects on migration. We investigate three possible diffusion channels. The first is membership in the same household. Because migration is a risky undertaking,

individuals are likely to rely on family ties to obtain trustworthy information about migration opportunities (Massey and Espinosa 1997). A second plausible channel of diffusion is village networks. Individuals are more likely to know migrants in their own village than outside their village. Reciprocity operates strongly within villages; individuals receive information or help from their fellow villagers with the expectation of reciprocating in the future (Portes and Landolt 2000).

A final channel of diffusion is Nang Rong region. Whereas the first two diffusion channels represent *specific* network externalities (identities of alters are known and significant to the individual), the third channel reflects *general* network externalities, where alters are likely to be anonymous. Specific network externalities accrue through information or help household or village members provide to potential migrants, which reduces the risks of migration. General network externalities, by contrast, work through regional institutions that develop over time to support migration flows. For instance, as one of the authors observed in her fieldwork, high migration flows from Nang Rong prompted urban employers to send recruiters, fostering more migration.

Table 4 displays results for four individual-level models of migration diffusion between 1973 and 2000. The baseline model includes exogenous variables that may influence an individual's hazard of migrating. Estimates indicate that migration rates decline with age, increase with education, and are higher for men than for women and for unmarried than for married individuals. Rates of first migration decline with the size of household landholdings and increase as the household dependency ratio rises. Village characteristics that increase migration rates include population, presence of a newspaper

reading center (where residents may mingle and share information about migration exper-

iences) and the availability of electricity (a proxy for economic development).

Table 4 here

Models 1, 2 and 3 successively add network parameters to the baseline model.

Model 1 presents the results for diffusion through the household channel. Models 2 and 3,

respectively, introduce prior migrants in the village and in other Nang Rong villages.

Likelihood ratio tests (Table 4, lower panel) show that each covariate improves model fit

significantly. Model 3, which includes all three covariates representing alternative

diffusion channels in mean-deviation form demonstrates that the migration rate

significantly increases with the number of migration events in the household, village, and

Nang Rong. Of these, the number of migrants in the household affects migration most

strongly, followed by the number of migrants in the village, and then, at a distance, the

number of migrants in Nang Rong. One additional migrant in the household has the same

effect on the migration rate as do about 200 more migrants in the village or about 800

more in the region. These results provide strong support for the first hypothesis, and

underscore the importance of specific, rather than general, network externalities for rural-

urban migration, with the effect of others' behavior on an individual's migration chances

increasing with social proximity.

Social homogeneity, network externalities and migration. Table 5 reports results

for a test of the second hypothesis, which posits that higher village homogeneity (a proxy

for homophily) depresses migration rates. Indicators of village educational and occupat-

ional homogeneity are highly correlated (Pearson's correlation = 0.74), so each is included in a different model, standardized to [0,1] range and centered around its mean.

Table 5 here

As per hypothesis 2, we expect status homogeneity (which, as we have seen is highly correlated with indicators of status homophily at the village level) to limit the diversity of information available to potential migrants, and thus to depress village migration rates. Model 1 supports this expectation: a unit increase above the mean level of educational homogeneity reduces the rate of first migration by 61.7 percent (0.383-1=-0.617). The model also controls for village-level mean education, which does *not* significantly affect migration. The effect of occupation homogeneity estimated in Model 3, is similar; a unit increase above the mean reduces the rate of first migration by 58 percent (0.420-1=-0.580).

The third hypothesis suggests that network effects increase with village status homogeneity due to increased cohesion and velocity of information flows between past and current migrants. To test this hypothesis, we interact the number of prior village migrants with education homogeneity in Model 2 and with occupation homogeneity in Model 4. Likelihood ratio tests show a significant improvement in fit compared to models without the interaction term. The positive coefficient estimates for the interaction terms, in both Models 2 and 4, provide strong support for the hypothesis. These results suggest that whereas homogeneity (in education or occupation) has a *negative direct effect* on the rate of first migration (consistent with Hypothesis 2), it has a *positive indirect effect* by strengthening network externalities (as posited by Hypothesis 3).

Explaining Heterogeneity in Migration Patterns across Villages. We now turn to Hypothesis 4: Can the model of network externalities and social homophily developed in this paper help to explain the puzzling divergence of in migration patterns across similar Thai villages?

The individual-level models presented above suggest that village social homogeneity reduces migration rates, but increases the impact of number of prior migrants on the rates of migration. By amplifying effects of prior migration, homogeneity (and, by implication, homophily) increases inequality among villages over time. This boosts the effects of prior migration for villages with many migrants and penalizes villages with relatively few. By contrast, where heterogeneity is high, migration is driven primarily by individual-level factors and household experience, generating less divergent village outcomes. In other words, because social homogeneity is associated with network homophily, which amplifies the effect of network externalities, we should observe a wider dispersion of migration patterns among homogeneous villages than among villages with medium or low homogeneity.

Table 6 here

We classify the 22 villages into high-, medium or low-homogeneity tertiles with respect to educational and occupational homogeneity, so that each group contains seven or eight villages each year. We compute the dispersion of the cumulative number of migrants across the villages in each group using the gini coefficient. We expect between-village dispersion to be higher in villages with high homogeneity. To test this

expectation, we regress dispersion on the binary indicators for the homogeneity classes. Observing three groups over 28 years yields 84 cases.

Estimates from an OLS model including year dummies as controls (Table 6) demonstrate that dispersion in migration outcomes is wider among villages with medium education homogeneity compared to those with low homogeneity (the reference category), and wider still among the most homogeneous villages. Results in the second column show similar patterns for classes based on occupational homogeneity: dispersion in migration outcomes is widest among villages with high occupation homogeneity, followed by villages with medium and low homogeneity, respectively. These results provide support for the hypothesis that social homogeneity, through its impact on network externalities, exacerbates inequality between villages in migration outcomes. Indeed it suggests that this mechanism can *generate* inequality based on small initial differences in ways that could not be predicted based on initial endowment. [32]

Conclusion

We argue that inequality emerges from the union of distinctive forms of interdependence, those that make courses of action more productive or less risky for actors whose friends and associates also pursue them; and those that lead people to form networks with alters of similar status. Inequality, we contend, is produced by the interaction of these two features of social structure: network externalities and homophily.

---

[32] See Salganik, Dodds and Watts (2006) for a similar finding in an experimental context.

We have presented two studies. The first was a computational model of changing intergroup inequality in access to the Internet over the course of that technology's diffusion. The second was an empirical study of trends in rural-urban migration in 22 Thai villages. The insight that interdependent choice processes under conditions of homophily generate inequality helped to illuminate each of these very different phenomena and yielded strikingly similar conclusions.

The first parallel was the network effects on individual-level choice, an assumption built into the Internet computational model and an empirical finding for Thai migration. The second result that emerged from the Internet model and was observed again in the migration study is that homophily (and the homogeneous networks it produces) tends to depress individual-level adoption. The third result, which again both emerged from the computational model and was observed in the Thai data, is that homophily stimulates positive network effects when adopters share networks. The fourth and most important result is that interdependent choices with network externalities under conditions of homophily *generate inequality.* In the Internet case, this mechanism augmented existing inequality between groups defined on the basis of income, education, and race. In the Thai case, it contributed to the emergence of inequality among initially similar villages.

These results demonstrate the utility and value of the approach we propose. Yet many questions remain about the scope of these mechanisms. Not all practices have positive externalities: consumption of many goods and services is competitive, with one person's enjoyment detracting from, rather than enhancing, another's. Moreover, only some network externalities are identity-specific rather than general: We want enough

people to like sushi for our favorite Japanese restaurant to keep its doors open and its servings fresh, or to contribute to our favorite charity that it can continue to promote causes we cherish, but we don't much care who those people are.   A challenge for further work is to estimate the contribution of the mechanisms described here to overall inequality and how that contribution has changed over time.

A second question is whether such mechanisms may under certain conditions *undermine* inequality. At the onset of the Internet, when young people seemed peculiarly able to master the technology, some suspected that the Internet would provide equal opportunity for young people from all backgrounds. That never happened. But there may be status-conferring or lucrative activities in which subaltern groups possess initial advantages that identity-specific externalities amplify. The advantages of African-American youth in hip-hop and basketball, or of young French-Canadian men in ice hockey, while posing little challenge to entrenched privilege, nonetheless suggest how externalities and homophily can combine to benefit non-elites.

A third question has to do with the potential of this mechanism to exacerbate intergroup differences in risk and harm. The risks of smoking cigarettes, for example, are greater if your network reinforces your behavior, discourages quitting, and teaches you to value the practice (Christakis and Fowler 2008). Unprotected sex is a greater risk in networks where members deride condom use than in milieus in which condom use is widespread (Tavory and Swidler 2009).   If externalities were entirely negative – if their only effect were to raise the cost or reduce the benefit of a behavior – then externalities would *discourage* adoption. But in both cases, in-groups offer salient short-term rewards (affiliation, identity, status, participation in shared rituals), alongside long-term,

discountable risks. To the extent that these networks are characterized by status homophily, such cases illustrate a potentially powerful source of cumulative disadvantage.

There is still much we do not understand about the ways that networks produce cumulative advantage by enhancing the value of new practices to their adopters. Nonetheless, the results reported in this paper establish the importance of network externalities in differentiated diffusion networks as a mechanism that must be taken into account if we are to understand the structural bases of social inequality.

References

Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *The Economic Journal* 99:116-31.

Baron, James N. and William T. Bielby. 1980. "Bringing the Firms Back in: Stratification, Segmentation, and the Organization." *American Sociological Review* 45: 737-765.

Bello, W., S. Cunningham, and L.K. Poh. 1998. *A Siamese Tragedy: Development and Disintegration in Modern Thailand*. London: Zed Books, Ltd.

Bonfadelli, Heinz. 2002. "The Internet and Knowledge Gaps: A Theoretical and Empirical Investigation." *European Journal of Communication* 17: 65-84.

Booth, Alan, John Edwards and David Johnson. 1991. "Social Integration and Divorce." *Social Forces* 70: 207-24.

Boudon, Raymond. 1973. *Education, Opportunity and Social Inequality: Changing Prospects in Western Society*. New York: Wiley.

Breen, Richard and John H. Goldthorpe. 1997. "Explaining Educational Differentials: Towards a Formal Rational Action Theory." *Rationality and Society* 9: 275-305.

Buhai, Sebastian and Marco van der Leij. 2008. "A Social Network Analysis of Occupational Segregation." Manuscript, Tinbergen Institute, Erasmus University, Rotterdam.

Buhler, Christoph and Ewa Fratczak. 2007. "Learning From Others and Receiving Support: The Impact of Personal Networks on Fertility Intentions in Poland." *European Societies* 9: 359-82.

Cairncross, Frances. 1997. *The Death of Distance: How the Communications Revolution Will Change Our Lives*. Boston: Harvard Business School Press.

Christakis, Nicholas and James H. Fowler. 2008. "The Collective Dynamics of Smoking in a Large Social Network," *New England Journal of Medicine* 358: 2249-2258.

Coleman, James S., Elihu Katz and Herbert Menzel. 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry* 20: 253-70.

Curran, Sara, Filiz Garip, Chang Chung, and Kanchana Tangchonlatip. 2005. *Social Forces* 84: 225-55.

Currarini, Sergio, Matthew O. Jackson and Paolo Pin. 2010. "Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation." *Proceedings of the National Academy of Science* 107: 4857-61.

DiMaggio, Paul and Bart Bonikowski. 2008. "Make Money Surfing the Web: The Impact of Internet Use on the Earnings of U.S. Workers." *American Sociological Review* 73: 227-52.

DiMaggio, Paul and Joseph Cohen. 2004. "Information Inequality and Network Externalities: A Comparative Study of the Diffusion of Television and the Internet." In

The Economic Sociology of Capitalism, edited by Victor Nee and Richard Swedberg. Princeton: Princeton University Press.

DiMaggio, Paul, Eszter Hargittai, Coral Celeste, and Steven Shafer. 2004. "From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality." In *Social Inequality*, edited by Kathryn Neckerman. New York: Russell Sage Foundation.

Dunlevy, J. A. 1991. "On the settlement patterns of recent Caribbean and Latin immigrants to the United States." *Growth Change* 22(1): 54-67.

Featherman, David and Robert Hauser. 1978. *Opportunity and Change*. New York: Academic Press.

Feld, Scott. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86: 1015-1035.

Fischer. Claude. S. 1992. *America Calling: A Social History of the Telephone to 1940*. Berkeley: University of California Press.

Fussell, E. and Douglas Massey. 2004. "The Limits to Cumulative Causation: International Migration from Mexican Urban Areas." *Demography* 41(1):151-171.

Garip, Filiz. 2008. "Social Capital and Migration: How Do Similar Resources Lead to Divergent Outcomes?" *Demography* 45(3): 591-617.

Garip, Filiz and Sara Curran. 2009. "Increasing Migration, Diverging Communities: Changing Character of Migrant Streams in Rural Thailand." Forthcoming in *Population Research and Policy Review*.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-1380.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83: 1420-43.

Hedstrom, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. New York: Cambridge University Press.

Jaeger, Mads Meier. 2007. "Economic and Social Returns to Educational Choices: Extending the Utility Function." *Rationality and Society* 19: 451-83.

Kao, Grace and Jennifer S. Thompson. 2003. "Racial and Ethnic Stratification in Educational Achievement and Attainment." *Annual Review of Sociology* 29: 417-42.

Katz, Michael and Carl Shapiro. 1985. "Network Externalities, Competition and Compatibility." *American Economic Review* 75: 424-40.

Klopfenstein, Kristin. 2004. "Advanced Placement: Do Minorities Have Equal Opportunity." *Economics of Education Review* 23: 115-31.

Korinek, Kim, Barbara Entwisle and Aree Jampaklay. 2005. "Through Thick and Thin: Layers of Social Ties and Urban Settlement among Thai Migrants." *American Sociological Review* 70(5): 779-800.

Lazarsfeld, Paul and Robert K. Merton. 1954. "Friendship as Social Process: A Substantive and Methodological Analysis." Pp. 18-66 in *Freedom and Control in Modern Society*, ed. Monroe Berger, Theodore Abel, and Charles Page. New York: Van Nostrand.

Leibenstein, Harvey. 1950. "Bandwagon, Snob and Veblen Effects in the Theory of Consumption." *Quarterly Journal of Economics* 64: 183-207.

Leigh, Andrew and Robert D. Atkinson. 2001. *Clear Thinking on the Digital Divide.* PPI Policy Report (June 26). Washington: Progressive Policy Institute. http://www.ndol.org/documents/digital_divide.pdf . (Last accessed March 23, 2008.)

Levitt, Peggy. 1998. "Social Remittances: Migration Driven Local-Level Forms of Cultural Diffusion." *International Migration Review* 32(4): 926-948.

Lieberson, S. 1969. "Measuring Population Diversity." *American Sociological Review* 34(6): 850-862.

Lin, Nan, Walter M . Ensel, and John C . Vaughn. 1981 . "Social Resources and Strength of Ties : Structural Factors in Occupational Status Attainment ." *American Sociological Review* 46(4):393-405,

Macy, Michael W. and Robert Willer. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28: 143-66.

Massey, Douglas. 1990. "Social Structure, Household Strategies, and the Cumulative Causation of Migration." *Population Index* 56(1): 3-26.

Massey, Douglas, L. Goldring and Jorge Durand. 1994. "Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities." *American Journal of Sociology* 99:1492-533.

Massey, Douglas and R. Zenteno. 1999. "The Dynamics of Mass Migration." *Proceedings of the National Academy of Sciences* 96(9): 5328-35.

Massey, Douglas and K. E. Espinosa. 1997. "What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis." *American Journal of Sociology* 102(4):939-99.

McPherson, J. Miller and Lynn Smith-Lovin. 1987. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52(3): 370-379.

McPherson, Miller, Lynn Smith-Lovin, and Matthew E. Brashears. 2006. "Social Isolation in America: Changes in Core Discussion Networks over Two Decades." *American Sociological Review* 71: 353-375.

McPherson, J. Miller., Lynn Smith-Lovin, and J.M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415-44.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159(3810): 56-63, January 5, 1968.

Mills, Mary Beth. 1999. *Thai Women in the Global Labor Force: Consuming Desires, Contested Selves.* New Brunswick, NJ: Rutgers University Press.

Offe, Claus and Helmut Wiesenthal. 1980. "Two Logics of Collective Action." *Political Power and Social Theory* 3: 67-115

Phongpaichit, Pasuk. 1980. "The Open Economy and Its Friends: The "Development" of Thailand." *Pacific Affairs* 53(3): 440-460.

Popielarz, Pamela and J. Miller McPherson. 1995. "On the Edge or In Between: Niche Position, Niche Overlap, and the Duration of Voluntary Association Membership." *American Journal of Sociology* 101: 698-720.

Portes, A. and P, Landolt. 2000. "Social Capital: Promise and Pitfalls of its Role in Development." *Journal of Latin American Studies* 32: 529-547.

Putnam, Robert. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies* 30:137-74.

Rindfuss, R., T. Kaneda, A. Chattopadhyay, and C. Sethaput. 2007. "Panel Studies and Migration." *Social Science Research* 36: 374–403.

Quillian, Lincoln. 2006. "Can Social Capital Explain Persistent Racial Poverty Gaps?" National Poverty Center Working Paper Series #06-12, Northwestern University, Evanston, Illinois.

Quillian, Lincoln and Mary E. Campbell. 2003. "Beyond Black and White: The Present and Future of Multiracial Friendship Segregation." *American Sociological Review* 68: 540-66.

Rogers, Everett M. 2003. *Diffusion of Innovations*, 5th edition. New York: Free Press.

Rossman, Gabriel, Ming Ming Chiu, and Joeri Mol. 2008. "Modeling Diffusion of Many Innovations Via Multilevel Diffusion Curves: Payola in Pop Music Radio." *Sociological Methodology* 38 (1): 201-230.

Salganik, Matthew J., Peter Sheridan Dodds and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311: 854-56.

Schelling, Thomas. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1: 143-86.

Schement, Jorge. 1995. "Beyond Universal Service: Characteristics of Americans without Telephones, 1980-1993." *Telecommunications Policy* 19: 477-85.

Shy, Oz. 2001. *The Economics of Network Industries*. 2001. New York: Cambridge University Press.

Shy, Oz. 2007. "Dynamic Models of Religious Conformity and Conversion: Theory and Calibration." *European Economic Review* 51: 1127-53.

Simon, Herbert. 1957. "The Compensation of Executives." *Sociometry* 20: 32-36.

Skvoretz, John. 1990. "Biased Net Theory: Approximations, Simulations, and Observations." *Social Networks* 12: 217–38.

Somrudee, Nicro. 1993. "Thailand's NIC Democracy: Studying From General Elections." *Pacific Affairs* 66(2): 167-182.

Stark, O. and E. Taylor. 1989. "Relative Deprivation and International Migration." *Demography* 26: 1-14.

Strang, D. 1991. "Adding Social Structure to Diffusion Models." *Sociological Methods and Research* 19(3): 324-353.

Tavory, Ido and Ann Swidler. 2009. "Condom Semantics: Meaning and Condom Use in Rural Malawi." *American Sociological Review* 74: 171-89.

Tilly, Charles. 2006. *Identities, Boundaries, and Social Ties.* Boulder, Colorado: Paradigm Publishers.

Turkle, Sherry. 1995. *Life on the Screen*. New York: Simon and Schuster.

Van de Rijt, Arnout. 2009. "Selection and Influence in the Assimilation Process of Immigrants." Working Paper, SUNY Department of Sociology.

Van den Bulte, Christophe and Gary Lilien. 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort." *American Journal of Sociology* 106: 1409-35.

Van den Bulte, Christophe and Stefan Stremersch. 2004. "Social Contagion and Income Heterogeneity in New Product Diffusion." *Marketing Science* 23: 530-44.

Watts, Duncan. 1999. "Networks, Dynamics and the Small-World Phenomenon." *American Journal of Sociology* 105: 493-527.

Western, Bruce, Deirdre Bloome, and Christine Percheski. 2008. "Inequality Among American Families with Children, 1975 to 2005." *American Sociological Review* 73: 903-20.

White, Harrison C. 1970. *Chains of Opportunity: System Models of Mobility in Organizations*. Cambridge: Harvard University Press.

Appendix B: Descriptive Statistics for Thai Village Data

Table B1. Descriptive Statistics for Thia Village Data (1973-2000)

Variable	Mean	Standard deviation	Minimum	Maximum
Individual level (N = 50,198 person-years)				
Age	21.87	6.70	14	41
Sex (Male=1)	0.44	0.50	0	1
Years of education	5.72	2.31	0	19
Married	0.39	0.49	0	1
Household land (rai)	27.75	25.85	0	209
Household dependency ratio	0.69	0.59	0	6
Village level (N = 616 village-years)				
Village population	641.14	147.71	347	1020
Factory w/in 5km of village?	0.22	0.41	0	1
Number of rice mills	2.97	1.67	0	7
School in village?	0.58	0.49	0	1
Newspaper reading center in village?	0.31	0.46	0	1
Years since village electrified	4.11	5.27	0	20
Cumulative no of village migrants in prior ye	85.57	78.49	0	328
Number of annual migrants	7.65	5.65	0	36
Education homophily in year [0,1]	0.23	0.13	0	1
Occupation homophily in year [0,1]	0.51	0.23	0	1

FIGURES

Fig. 1: Stylized diffusion curves for two pairs of populations
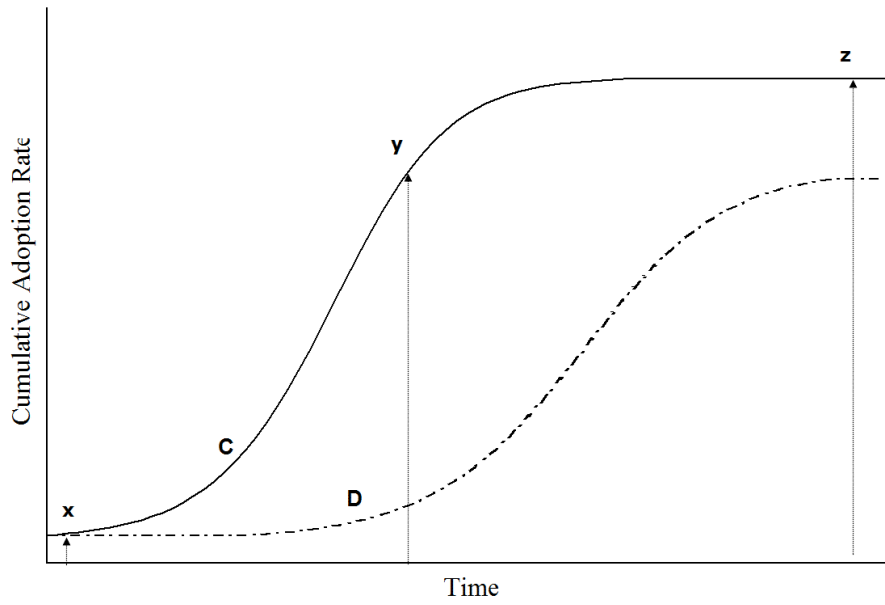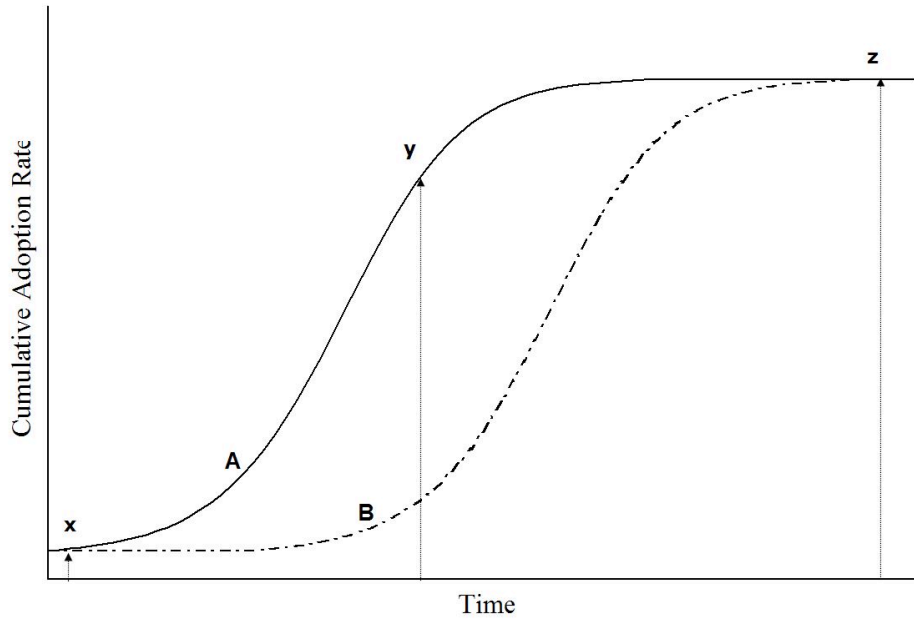
Odds Ratio of Internet Use at Home by Gender
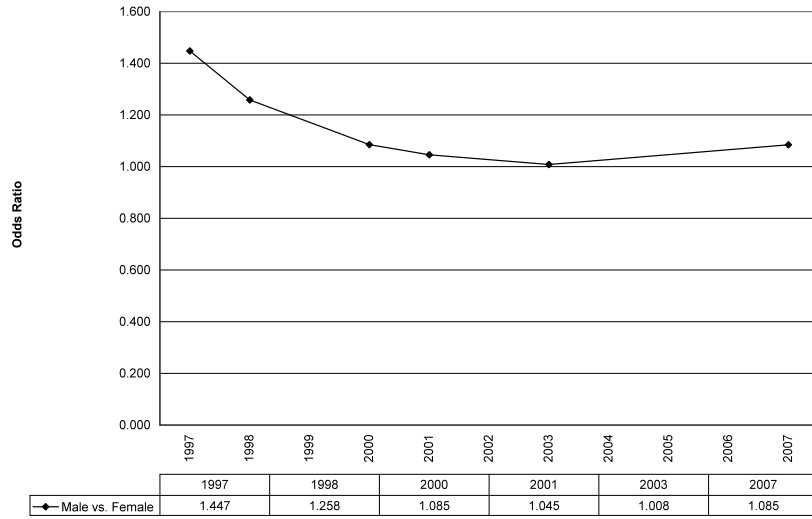
	1997	1998	2000	2001	2003	2007
Male vs. Female	1.447	1.258	1.085	1.045	1.008	1.085

Odds Ratio of Internet Use at Home by Race

	1997	1998	2000	2001	2003	2007
White vs. Black	2.994	3.091	2.594	2.469	2.293	2.264
White vs. Hispanic	2.901	2.955	2.995	2.931	2.858	2.490

Odds Ratio of Internet Use at Home by Age

	1997	1998	2000	2001	2003	2007
56+ vs. 36 - 55	0.235	0.277	0.295	0.308	0.350	0.432
56+ vs. 18 - 35	0.248	0.304	0.314	0.310	0.367	0.537
36 - 55 vs. 18 - 35	1.056	1.101	1.064	1.007	1.049	1.243
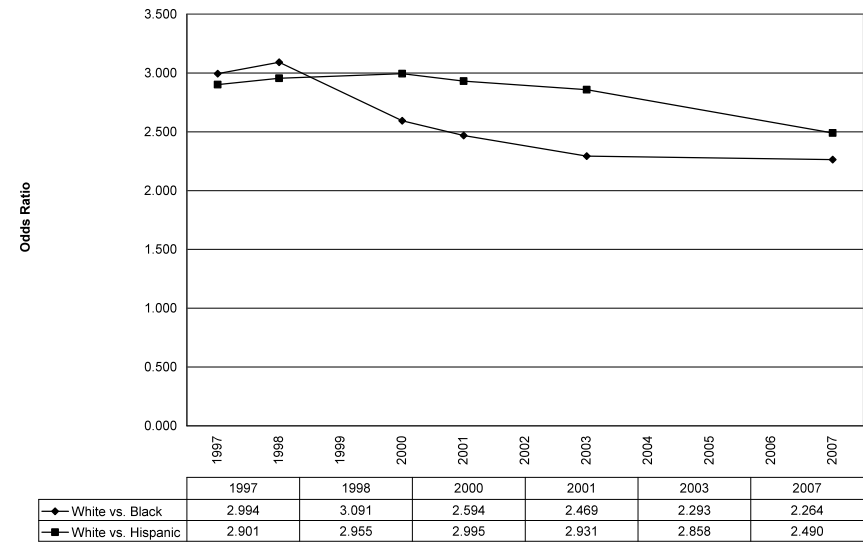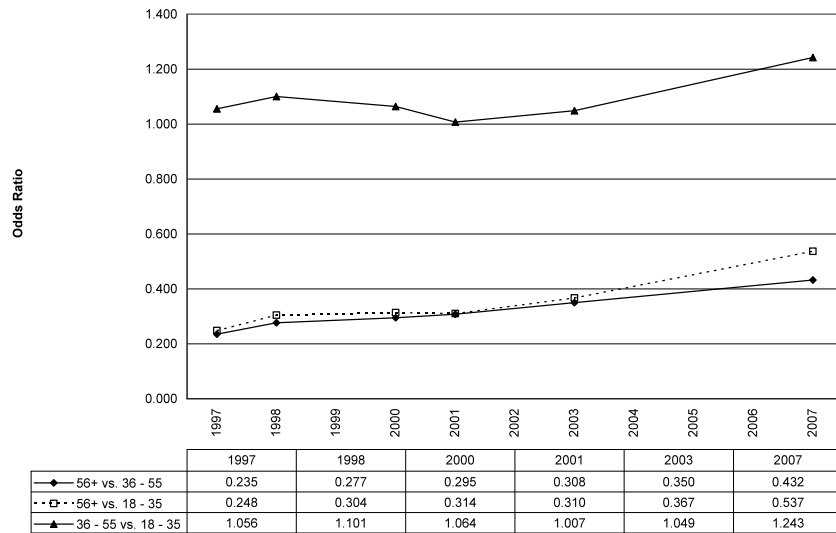
Odds Ratio of Internet Use at Home by Education

	1997	1998	2000	2001	2003	2007
College+ vs. HS	3.529	3.325	3.291	3.233	3.456	3.611
College+ vs. LHS	22.104	16.138	17.437	15.123	16.291	12.916
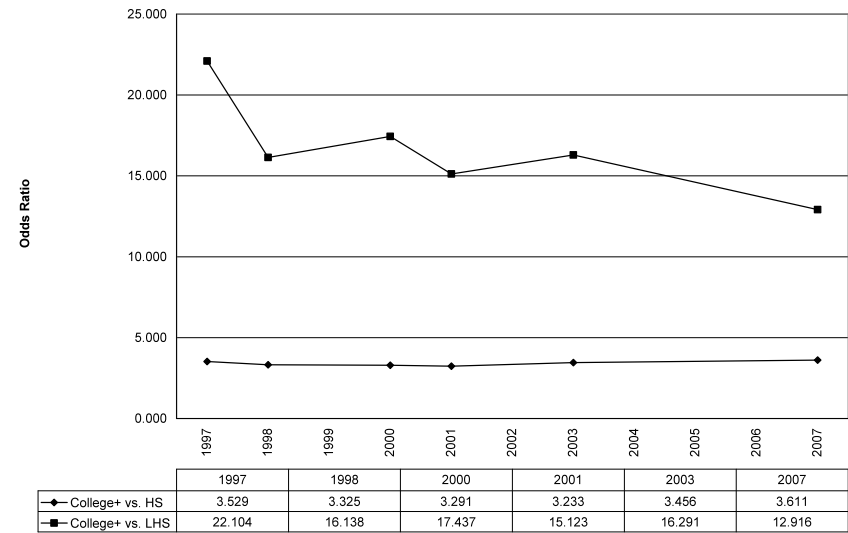
Figure 3: Diffusion Trajectories for Seven Conditions of Network Externalities and Homophily
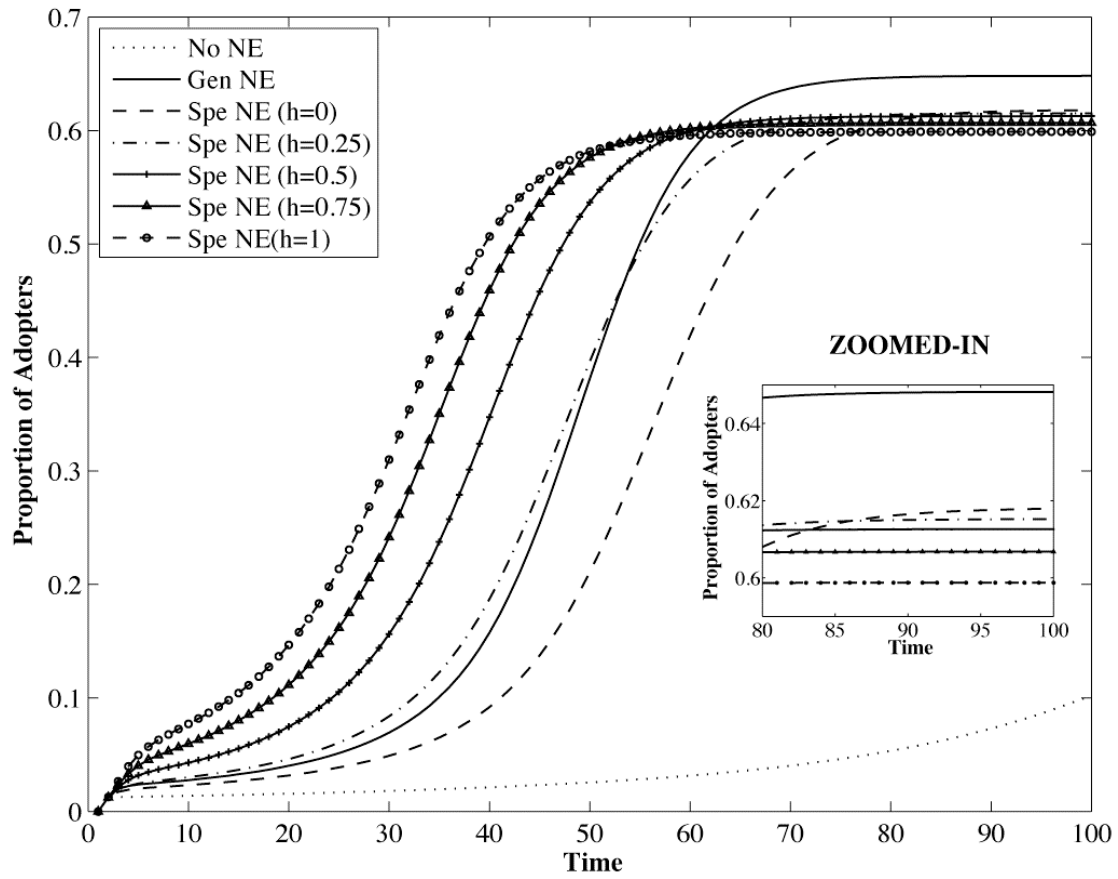
Figure 4: Odds Ratios of Diffusion Rates for Highest as compared to Lowest Income Classes in 6 Conditions of Externalities and Homophily
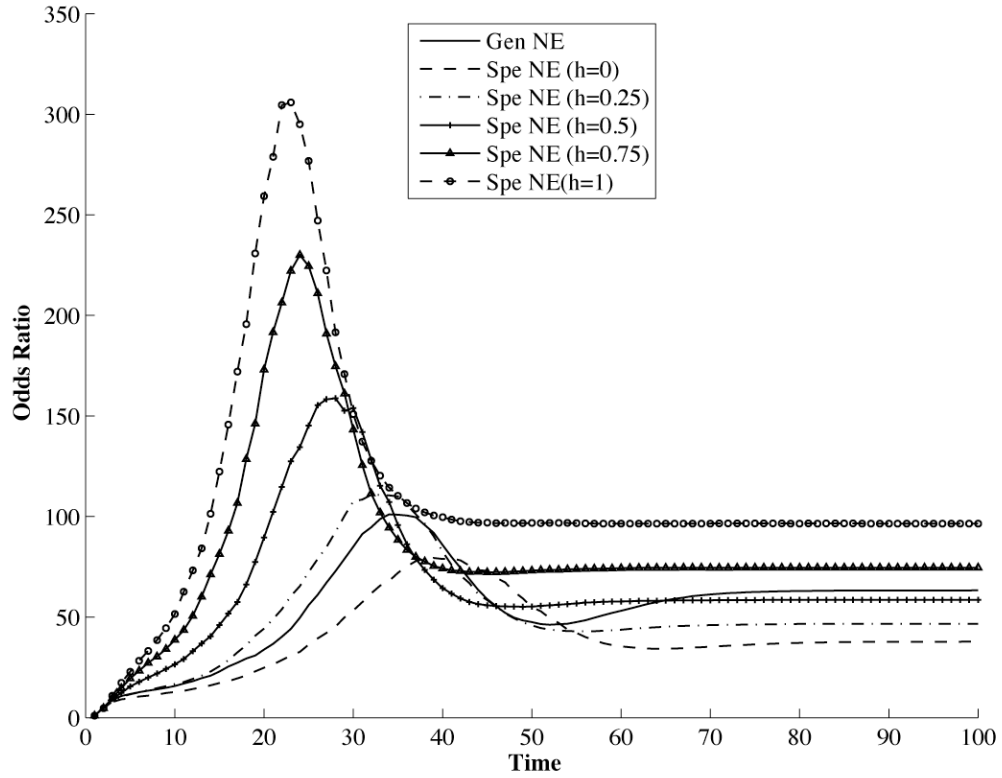
Figure 5: *Odds Ratios of Diffusion Rates for College Graduates as compared to Agents without High School Degrees in 6 Conditions of Externalities and Homophily*
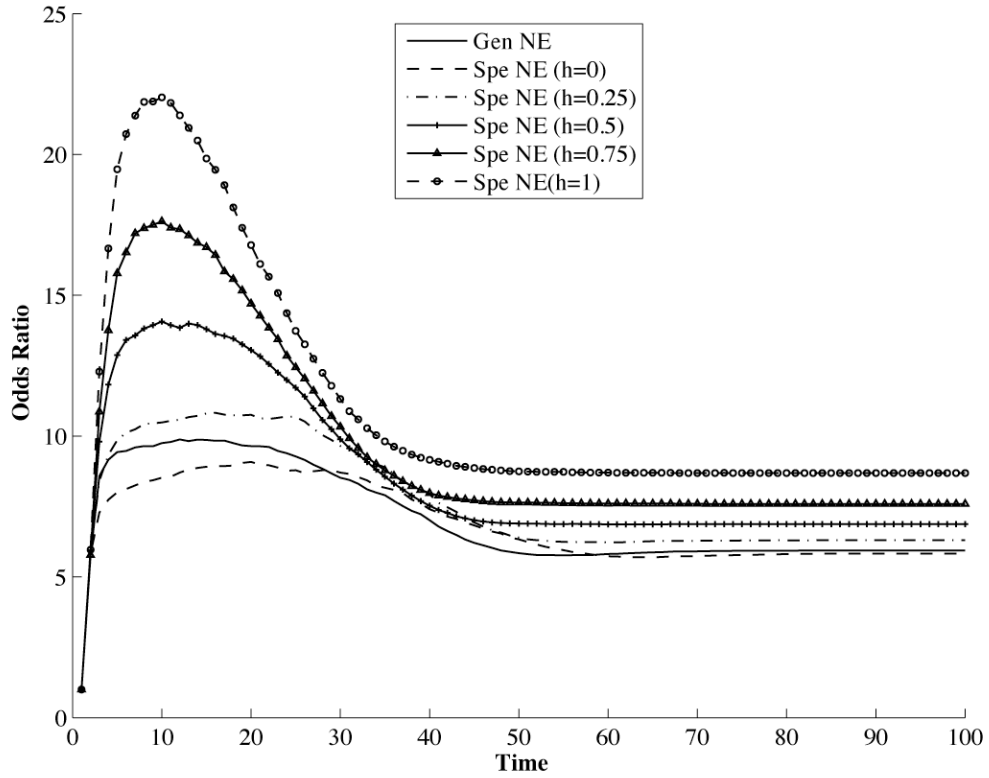
Figure 6: Odds Ratios of Diffusion Rates for Whites as compared to Blacks in 6 Conditions of Externalities and Homophily
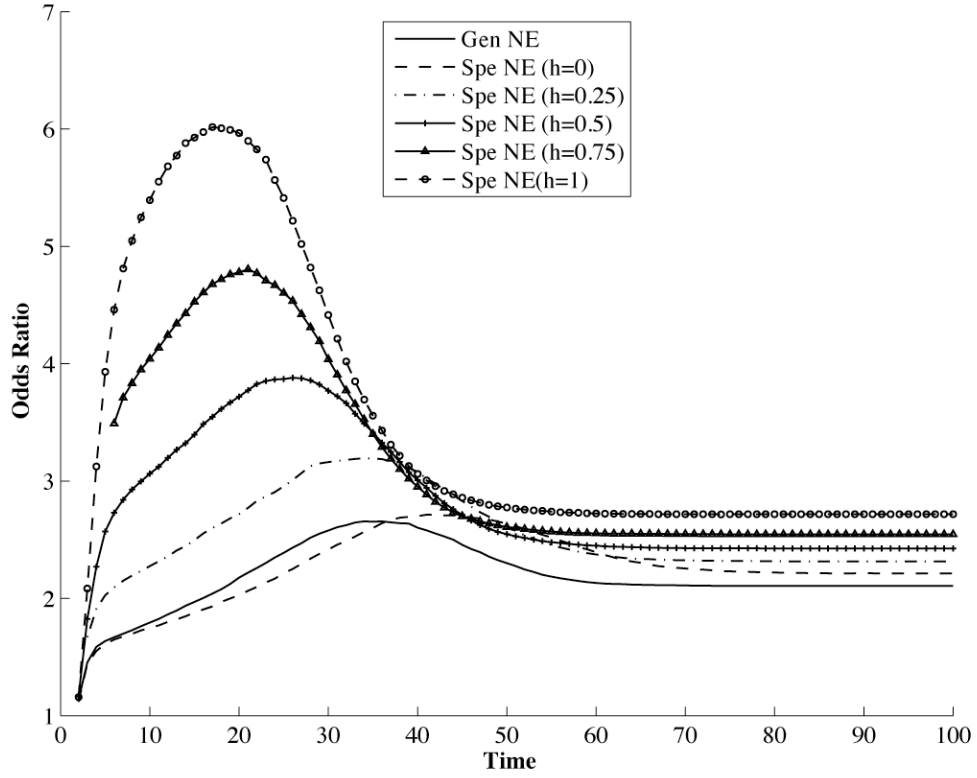
Figure 7: Changes in the Estimated Coefficient of Log (Income) Over Time for 6 Conditions of Externalities and Homophily
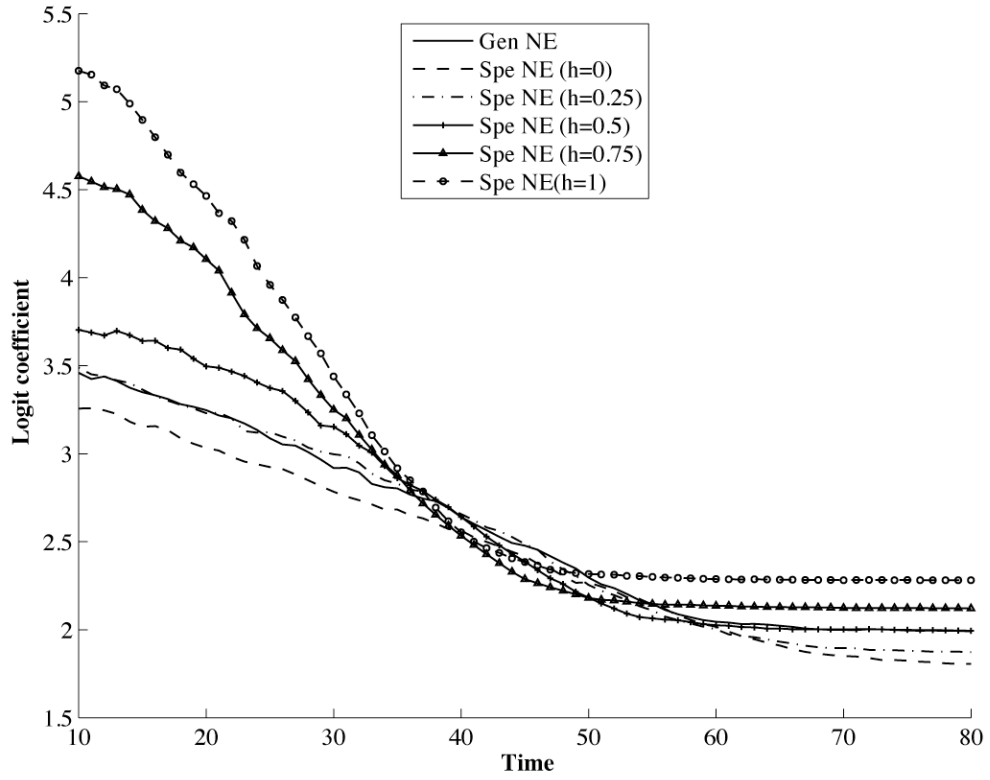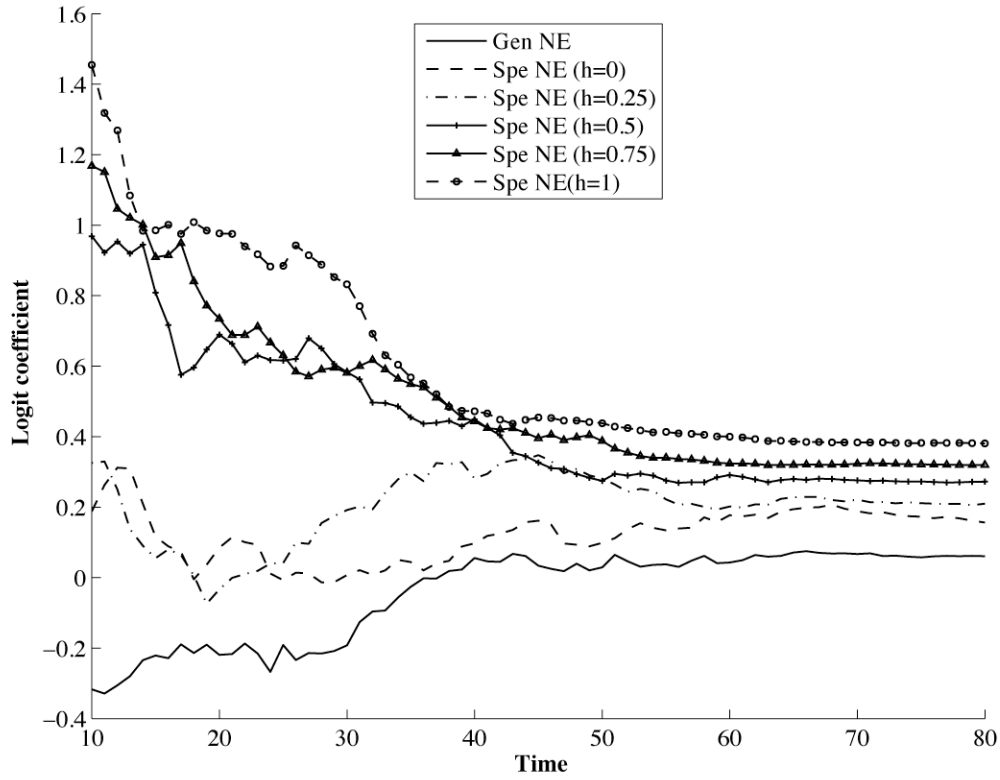
Figure 8: Changes in the Estimated Coefficient of Education over Time for 6 Conditions of Externalities and Homophily
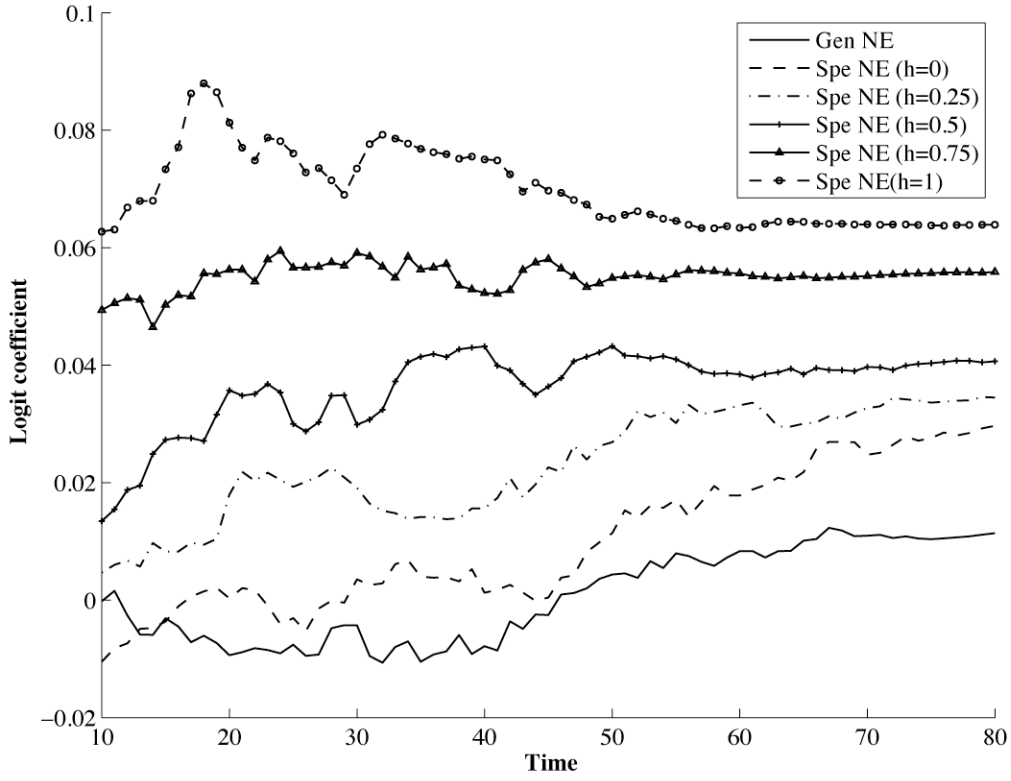
Figure 9: Changes in the Estimated Coefficient of Race (White, as compared to Black)
Over Time for 6 Conditions of Externalities and Homophily

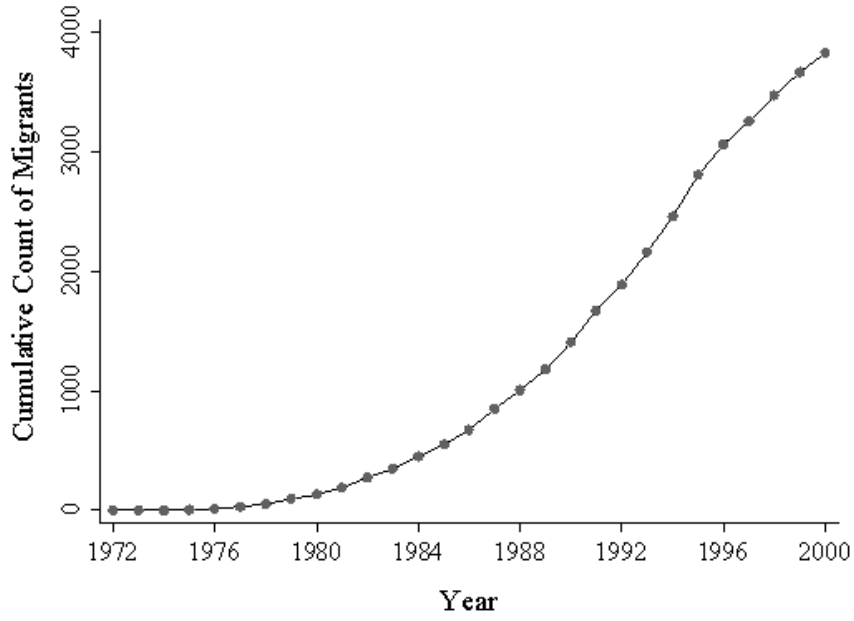## Figure 10: Diffusion of Migration in Nang Rong
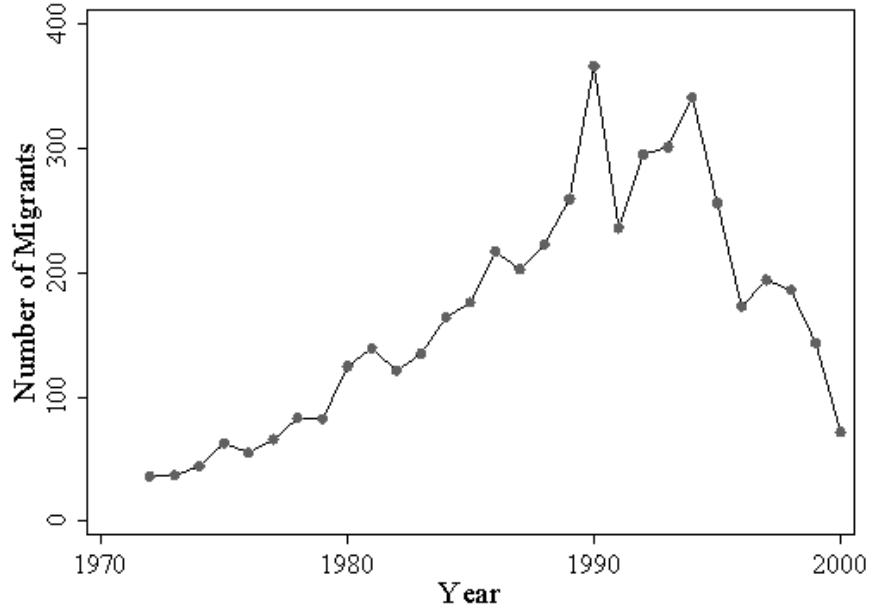
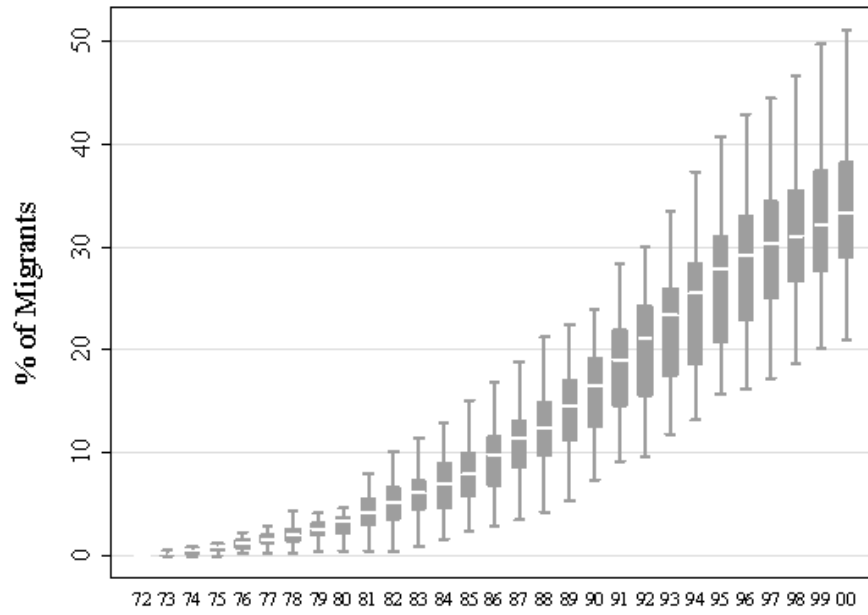*Figure 11: Number of New Migrants Each Year*

Figure 12: Dispersion of Cumulative Percentage of Migrants across Villages over Time

TABLES

Table 1. The Cases Compared

	Internet	Thai migration
Behavior	Adoption	Migration
Method	Agent-based model	Empirical analyses
Empirical focus	Intergroup inequality	Inter-village inequality
Homophily measured at level of	Ego network	Village
Key mechanism	Direct: Increased utility due to interactions with net	Indirect: Diminished risk due to information provided by net alters

Table 2. Linear Regression of Adoption Levels on Experimental Conditions

		RACE		INCOME		EDUCATION	
	All	Whites	Blacks	High income	Low income	BA education	Less than high school
(Reference: Homophily = 0)							
No network externalities	-0.516 **	-0.536 **	-0.399 **	-0.685 **	-0.238 **	-0.611 **	-0.351 **
General network externalities	0.030 **	0.028 **	0.043 **	0.032 **	0.017 **	0.023 **	0.030 **
Homophily = 0.25	-0.003 **	-0.001	-0.012 **	0.009 **	-0.014 **	0.005 **	-0.011 **
Homophily = 0.5	-0.005 **	-0.002 **	-0.024 **	0.017 **	-0.028 **	0.010 **	-0.024 **
Homophily = 0.75	-0.011 **	-0.006 **	-0.040 **	0.024 **	-0.046 **	0.012 **	-0.043 **
Homophily = 1	-0.019 **	-0.012 **	-0.061 **	0.029 **	-0.067 **	0.015 **	-0.068 **
Intercept	0.618 **	0.647 **	0.454 **	0.925 **	0.249 **	0.788 **	0.392 **
R^2	0.99	0.99	0.97	0.99	0.96	0.99	0.96
N	7000	7000	7000	7000	7000	7000	7000

**p<0.01, * p<0.05. All independent variables are binary. Both dependent and independent variables are measured on the final period of simulations (T=100).

Table 3. Linear Regression of Logged Odds Ratios on Experimental Conditions

	RACE	INCOME			EDUCATION		
	White/Black	High/Low income	High/Medium income	Medium/Low income	BA/Less than high school	BA/High school	Some college/ Less than high school
(Reference: Homophily = 0)							
No network externalities	-0.02 **	-0.13 **	0.02 *	-0.15 **	-0.07 **	-0.05 **	-0.13 **
General network externalities	-0.05 **	0.51 **	0.41 **	0.10 **	0.02 *	0.01	0.06 **
Homophily = 0.25	0.04 **	0.21 **	0.15 **	0.06 **	0.08 **	0.05 **	0.06 **
Homophily = 0.5	0.09 **	0.43 **	0.31 **	0.13 **	0.16 **	0.10 **	0.13 **
Homophily = 0.75	0.14 **	0.67 **	0.46 **	0.21 **	0.26 **	0.14 **	0.22 **
Homophily = 1	0.20 **	0.93 **	0.62 **	0.31 **	0.39 **	0.20 **	0.33 **
Intercept	0.79 **	3.62 **	1.87 **	1.76 **	1.75 **	1.01 **	1.54 **
R^2	0.28	0.66	0.57	0.29	0.41	0.35	0.27
N	7000	7000	7000	7000	7000	7000	7000

**p<0.01, * p<0.05. All independent variables are binary. Both dependent and independent variables are measured on the final period of simulations (T=100).

Table 4. Proportional Hazards Model of Individual Migration in 22 Nang Rong Villages (1973-2000)

Variable	(0) Baseline		Diffusion Channel (1) Household		(2) Village		(3) Nang Ro	
Age	0.941	**	0.940	**	0.935	**	0.932	**
	(0.011)		(0.011)		(0.012)		(0.012)	
Sex (Male=1)	1.132	**	1.143	**	1.134	**	1.133	**
	(0.043)		(0.043)		(0.043)		(0.043)	
Years of education	1.098	**	1.097	**	1.091	**	1.085	**
	(0.009)		(0.009)		(0.009)		(0.009)	
Married	0.619	**	0.623	**	0.618	**	0.618	**
	(0.038)		(0.038)		(0.038)		(0.038)	
Household land (in 10 rais)	0.973	**	0.968	**	0.970	**	0.972	**
	(0.008)		(0.008)		(0.008)		(0.008)	
Household dependency ratio	1.242	**	1.262	**	1.232	**	1.221	**
	(0.035)		(0.036)		(0.035)		(0.035)	
Village population (in 100s)	1.038	**	1.034	**	1.010		1.023	
	(0.013)		(0.013)		(0.013)		(0.014)	
Factory w/in 5km of village?	0.976		0.986		1.011		0.999	
	(0.049)		(0.050)		(0.051)		(0.050)	
Number of rice mills (in 10s)	1.061		1.072		1.085		1.073	
	(0.129)		(0.131)		(0.131)		(0.130)	
School in village?	0.992		1.004		0.977		0.982	
	(0.043)		(0.044)		(0.043)		(0.043)	
Newspaper reading center in village?	1.202	**	1.183	**	1.123	**	1.129	**
	(0.049)		(0.048)		(0.047)		(0.047)	
Years since village electrified	1.011	**	1.005		0.986	**	0.977	**
	(0.005)		(0.005)		(0.006)		(0.007)	
No of prior migrants in household			1.105	**	1.083	**	1.077	**
			(0.015)		(0.015)		(0.015)	
No of prior migrants in village (excl. hh, in 100s)					1.262	**	1.151	**
					(0.049)		(0.057)	
No of prior migrants in Nang Rong (excl. village, in 100s)							1.010	**
							(0.004)	
N (person-years at risk)	50,198		50,198		50,198		50,198	
Likelihood ratio χ² vs. prior model			50.17	**	35.23	**	8.53	**

**p<0.01, *p<0.05. Standard errors are given in parentheses. Results are presented in hazard ratios.

Table 5. Proportional Hazards Model of Individual Migration in 22 Nang Rong Villages (1973-2000)[a]

Variable	Education Homogeneity		Occupation Homogeneit	
	(1)	(2)	(3)	(4)
No of prior migrants in household	1.076 **	1.061 **	1.063 **	1.048 *
	(0.015)	(0.015)	(0.015)	(0.015)
No of prior migrants in village (excl. hh, in 100s)	1.132 *	1.254 **	1.193 **	1.467 *
	(0.060)	(0.068)	(0.071)	(0.091)
No of prior migrants in Nang Rong (excl. village, in 100s	1.013 **	1.002	1.014 **	1.001
	(0.004)	(0.004)	(0.004)	(0.005)
Mean education in village	0.972	1.037		
	(0.062)	(0.069)		
% working in a farm in village			0.990 *	0.988 *
			(0.004)	(0.004)
% working in a factory in village			0.996	1.013
			(0.007)	(0.008)
% working in construction in village			1.007	1.023
			(0.014)	(0.014)
% working in service in village			0.922 **	0.999
			(0.017)	(0.020)
Homogeneity in year [0,1]	0.383 *	1.857	0.420 *	1.757
	(0.159)	(0.857)	(0.178)	(0.805)
Homogeneity x No of prior migs in village		1.036 **		1.018 *
		(0.004)		(0.002)
N (person-years at risk)	50,198	50,198	50,198	50,198
Likelihood ratio χ² vs. model w/o interaction		102.89 **		96.63 *

**p<0.01, *p<0.05. Standard errors are given in parentheses. Results are presented in hazard ratios.

[a] Controls for age, sex, years of education, marital status, houshold land and dependency ratio, population, factories, rice mills, schools, newspaper reading center and electrification in the village are included. The variables used in the

Table 6. OLS Model of the Dispersion in Village-Level Migration in 22 Nang Rong Villages (1973-2000)

Variable	Education Homogeneity	Occupation Homogeneity
Medium homogeneity group (0/1)	0.07 **	0.06 *
	(0.01)	(0.02)
High homogeneity group (0/1)	0.13 **	0.16 **
	(0.01)	(0.02)
R^2	0.82	0.80
N	84	84

*p<0.05; **p<0.01. Std errors given in parentheses. Unit of analysis is 84 village-clusters based on tertiles of education or occupation homogeneity (3 groups x 28 years). The dependent variable is the dispersion of cumulative number of migrants across villages in group measured by the gini coefficient. Low homogeneity group is the reference category. Year dummies are included in both models.