Network diffusion under homophily and consolidation as a mechanism for social inequality

Linda Zhao
Harvard University
lzhao@fas.harvard.edu

Filiz Garip
Cornell University
fgarip@cornell.edu

Network diffusion under homophily and consolidation as a mechanism for social inequality

Abstract

DiMaggio and Garip (2011) define *network externalities* (where the value of a practice is a function of network alters that have already adopted the practice) as a mechanism exacerbating social inequality under the condition of *homophily* (where advantaged individuals poised to be primary adopters are socially connected to other advantaged individuals). The authors use an agent-based model of diffusion on a real-life population for empirical illustration, and thus, do not consider *consolidation* (correlation between traits), a population parameter that shapes network structure and diffusion (Blau and Schwartz 1984, Centola 2015). Using an agent-based model, this paper shows that prior findings linking homophily to segregated social ties and to differential diffusion outcomes are contingent on high levels of consolidation. Homophily, under low consolidation, is not sufficient to exacerbate existing differences in adoption probabilities across groups, and can even end up alleviating inter-group inequality by facilitating diffusion.

Sociologists are interested in how social networks shape the distribution of resources (Lin 1993; Portes 1998), attitudes (Small, Lamont and Harding 2010), and behaviors (Boyd 1989, Marsden and Gorman 2001, Smith and Christakis 2010) in society. One line of inquiry focuses on relating network formation processes to social stratification outcomes. This work suggests that *homophily* – the tendency for actors to associate with similar others – leads to segregated social networks, and accordingly, to inter-group inequality in outcomes for which network peers offer a positive influence (DiMaggio and Garip 2011, Manzo 2013, Montgomery 1991).

This idea has a strong basis in empirical findings that show the ubiquity of homophily in social networks (McPherson, Smith-Lovin and Cook 2001), and the prevalence of positive network effects in diffusion of various practices.[1] But, the idea is not tested as a unified causal chain, first, because it is difficult to find data that simultaneously captures how individuals form networks and how they adopt different practices, and second, because it is often hard to tell apart network selection from network diffusion even when data are available (Manski 1993).

These difficulties have led researchers to turn to formal analysis or agent-based models to demonstrate network-induced unequal distributions in particular domains. Montgomery (1991), for example, has developed a mathematical model to connect job referral networks to wage inequality. Manzo (2013) has proposed a computational model to link network-based educational choice to educational disparities in France. In both examples, the authors argued that higher degrees of homophily in social ties (by ability in the former study, and by socio-economic status in the latter) lead to higher degrees of inequality (in wages in the former, and in educational attainment in the latter).

In a recent article, DiMaggio and Garip (2011) – DG hereafter – generalized this argument to any good or practice that displays *network externalities*, that is, becomes more valuable (or less risky) to a person as more network peers adopt that good or practice. Network externalities, the authors argued, exacerbate inter-group inequality under the condition of homophily, that is, if advantaged individuals who are the likely primary adopters of a good or practice are socially linked to other advantaged individuals in the population. The authors illustrated this process with an agent-based model of Internet adoption in the United States, sampling the agents from the 2002 General Social Survey (GSS) to produce realistic distributions of income, educational attainment, and race. Similar to earlier work, the authors found, higher degrees of homophily (by income, education, and race) generate higher degrees of inequality in adoption.

By using real-life data, DG's empirical illustration presumed a particular state of the world as given, and did not fully take into account the context for homophilous tie formation. The fact that income, education, and race are highly correlated in the U.S. setting, for example, likely produced highly clustered and segregated networks even under low levels of homophily.

In this paper, we argue that prior results linking homophily to network-based inequality depend on presumed correlations between different characteristics in a population. We are inspired by Blau and Schwartz's (1984) seminal work which suggested *consolidation* – the correlation between traits in a population – as a key population parameter shaping social interactions. Centola (2015) recently used an agent-based model to investigate how homophily and consolidation jointly alter network structure and diffusion outcomes.

We follow a similar strategy, but modify Centola's model in two ways. Like DG, we first introduce status differences among individuals which affect adoption probability, and second compare group-specific adoption rates and equilibrium levels under different conditions of homophily and consolidation. We find that whether (and how much) homophily exacerbates inter-group inequality is contingent on the level of consolidation in the population. In what follows, we first situate our approach in the existing literature, then describe the novelty of our model, and finally discuss the implications of our results.

SOCIAL NETWORKS AS A MECHANISM FOR INEQUALITY

Cumulative advantage from social networks?

Sociologists have long regarded 'cumulative advantage' as a general mechanism for inequality that accrues greater increments of benefits to individuals that possess an initial advantage (Merton 1968; DiPrete and Eirich 2006). Empirical evidence has consistently found success leading to further success across a range of reward systems, most notably, in scholarly publications (Allison, Long and Krauze 1982) and recognition (Cole 1970, Reskin 1977), although the extent of support has varied by definition and method (van de Rijt et al. 2014, Salganik, Dodds and Watts 2006).

A large sociological literature has also shown that social networks provide access to various resources – such as information on jobs (Granovetter 1974) or migration opportunities (Massey and Espinosa 1997), normative pressures to assume healthy behaviors (Smith and Christakis 2008) or to improve school performance (Burke and Sass 2008, Sacerdote 2011) – that help individuals get ahead. Research has suggested that social networks can perpetuate inequality to the extent that such network-based resources are uneven in their distribution or benefits across groups (e.g., Garip 2008, Lin 2002, Smith 2005).

Scholars have only recently begun to connect these two lines of research, and consider social networks as a path toward cumulative advantage. In the area of labor markets, for example, scholars have argued that job referrals will increase wage inequality to the extent that workers of similar productivity are socially connected (Montgomery 1991, Arrow and Borzekowski 2004). In health research, studies have suggested that network effects in healthy behaviors will compound initial differences if high socio-economic status individuals associate with one another (Pampel et al. 2010). In the field of education, researchers have attributed a wider gap in academic achievement to tracking, which puts students of similar academic standing together (Gamoran 2011).

There are two common elements to these arguments. First, social networks are presumed to provide a positive influence in adopting a practice (in finding a job, taking up a healthy habit, or aiming for academic success). Second, social networks are assumed to be homophilous, that is, stratified by traits that are related to adoption probability (productivity, financial resources, or academic ability).

There is vast evidence for each of these assumptions. Reviews of research on labor markets (Marsden and Gorman 2001), health (Smith and Christakis 2008), and education (Epple and Romano 2011) all suggest strong positive network effects in various outcomes. And studies of social network composition reveal a strong tendency towards homophily in many settings (Kandel 1978; McPherson et al. 2001; Rivera et al. 2010). But there is only limited evidence for how network effects and homophily in social ties, in combination, compound initial advantages and widen existing disparities in the adoption of a practice.

Challenges in linking social networks to inequality

It is difficult to establish a unified causal chain from network formation to network diffusion and inequality. The first obstacle is the lack of data. Longitudinal data sets that simultaneously track individuals' network ties and behavior are rare, and when available, typically restricted to an institutional setting (e.g., schools in the AddHealth data). Such incomplete data, although useful for some questions, do not offer reliable measures of homophily or adoption levels in individuals' social networks.

The second obstacle is the notorious difficulties in the identification of network effects (also referred to as *network diffusion*, *peer effects*, or *endogenous interactions*). Individuals might adopt a practice because prior adopters in their network offer information, help, or influence that makes that practice less risky or more beneficial. Or, individuals might adopt a practice simply because they are subject to similar unobserved contextual factors as the prior adopters. To discard the latter possibility, researchers search for instances where individuals in the same environment vary in their exposure to network effects (e.g., Liu et al. 2010). But, even then, it is hard to address the so-called 'reflection problem', that is, to establish that each individual is truly responding to the group-level behavior (rather than the group-level behavior simply reflecting the sum of individual choices) (Manski 1993).

The third, and perhaps the thorniest, obstacle is the potential confounding of network formation and network diffusion. Individuals might self-select into a network in anticipation of engaging in a practice (for example, a student looking to improve academic performance might join a study group, and start working harder) (Elwert and Winship 2014). To address this issue of endogenous selection, researchers typically seek natural experiments assigning individuals to particular networks (e.g., Sacerdote 2001), or test the sensitivity of results to varying degrees of confounding (e.g., VanderWeele 2011).

Needless to say, these issues become even more intractable in combination, making it very difficult to provide evidence to existing claims in the literature that positive network effects in adoption and homophilous tie formation together increase inequality. A number of researchers, as a result, have turned to formal analysis and computational models to investigate the network dynamics underlying cumulative advantage.

Formal and computational models of social networks and inequality

Montgomery (1991) has proposed a social-learning model of a labor market, where employers can hire through referrals, and if so, pay higher wages. A mathematical model illustrated that, if workers are matched to their contacts in productivity, then the wage differences between high- and low-productivity workers will increase over time. Interestingly, if workers are matched to their contacts on characteristics unrelated to productivity (e.g., gender), and if there are initial differences in employment based on those characteristics (e.g., lower employment among women), wage inequality along those characteristics will also expand over time.

More recent models have built on this set-up, and reached similar conclusions on the role of networks on labor market inequality (Arrow and Borzekowski 2004, Calvó-Armengol and Jackson 2004). But, these formal analyses relied on simple models to remain tractable. For example, Calvó-Armengol and Jackson's (2004) work specified job search as a finite state Markov process, where transition in an agent's employment state is dependent on the states of its network ties. The authors took network configuration as given, not emergent, although they did consider potential drop-outs from the network. Unlike Montgomery (1991), the authors also did not take into account homophily, that is, the tendency for high-ability workers to be connected to high-ability job candidates.

Researchers have turned to agent-based models for more flexible specifications of network formation and diffusion. Agent-based models are synthetic worlds with computer-generated agents that follow rules for interacting with other agents and with their environment. The models simulate agents in interaction, and typically produce emergent macro-levels patterns that cannot be deduced from a simple aggregation of micro-level rules. Consequently, the models provide a critical tool for linking micro and macro level analysis, and for developing new theory (Bruch and Atwell 2015). Agent-based models do not solve the identification problems stated above, but they do allow researchers to create a complete (albeit synthetic) data set with no unobserved heterogeneity or endogenous selection, where the pure network effects can be isolated with counterfactual manipulations.

These advantages led Manzo (2013) to employ an agent-based model to investigate the sources of educational inequality in France. In his model, agents of different socio-economic groups make educational choices based on their ability, perceived pay-offs, and (in some cases) choices of other agents in their own group. The results showed that the empirical stratification in the French data can only be generated *in silico* if the model incorporates network influences from within one's own socio-economic group. In other words, network effects in educational choice, and homophily in social ties, together, provide a plausible explanation for the observed inequality in educational attainment in France.

DiMaggio and Garip (2011) – DG hereafter – followed a similar strategy, and used an agent-based model to understand the racial disparities in Internet adoption in the United States. The authors argued that initial differences in income and education between whites and African-Americans would translate into an enduring gap in adoption rates if one presumes (in line with the empirical evidence) that prior adopters in an ego's network encourage adoption, and that network alters are likely to resemble the ego in terms of income and education.

To fully grasp this idea, consider the first subscribers to home Internet service. These early adopters are likely to have sufficient financial and cultural resources (i.e., high income and education) to afford the new technology, and such individuals are disproportionately white in the U.S. context. Now consider the next round of subscribers. These individuals still need to have the requisite financial and cultural resources, but some also enjoy *network externalities*, that is, higher returns to the Internet as they can use the service to communicate with the earlier adopters. Now, if social ties were established at random, such externalities would be uniformly distributed in the population, and not change existing levels of inequality in adoption. But, as research shows, there is a high degree of homophily by education in personal networks in the United States, and even a higher degree of homophily by race (Marsden 1987, 1988). Introducing this pattern into the example, then, one can see that network externalities are bigger for the rich, better-educated and white individuals who are more likely to be connected to earlier adopters (who are also rich, better-educated and white). In this case, network externalities do not just perpetuate initial differences between the rich and the poor (or whites and African Americans), but externalities also make such differences larger than would be expected based on income or education differences alone.

Driven to explain a real-life puzzle, DG struck a fine balance between using an agent-based model to push theory on the one hand, and calibrating that model to fit the empirical case, on the other. Crucially, the authors did not create synthetic agents, but used the sample from the network module of the 2002 GSS to replicate the observed marginal and joint distributions of income, education, race, and network size in the United States.

Their algorithm first generates a network of connections among agents with a given degree of homophily.[2] Each agent has a reservation price – a price at which it will subscribe to Internet service. The reservation price is an increasing function income, education, and the share of network alters who have already adopted.[3] The price of Internet service is a declining function of overall adoption level to reflect economies of scale. This set-up implies that an agent can adopt because its reservation price has increased due to prior adopters in its network, or because the price of the service has dropped below its reservation price due to adoption in the population. At each time period, the algorithm computes each agent's adoption outcome by comparing its reservation price to the price of the Internet, updates reservation prices and the price of the service, and runs until adoption reaches an equilibrium level in the population.

The results showed that, as homophily increases in the network, the slope of the diffusion curve (which plots the cumulative proportion of adopters across time) becomes steeper. This is because individuals with an initial advantage (i.e., high income and education) become more likely to be connected to other advantaged individuals, compounding the network effects within this select group. But, this speed in adoption comes at a cost. As homophily rises in the network, the equilibrium adoption level in the population declines. While the practice diffuses quickly among the advantaged, it fails to spread to less advantaged groups.

Considering the context for homophily

Prior empirical, formal and computational analyses discussed above (and more comprehensively in DiMaggio and Garip's (2012) review) put forth a common argument: Social networks exacerbate inequality if social ties facilitate the adoption of beneficial practices, and if those ties exhibit homophily, that is, similarity in individual traits related to adoption.

But, interestingly, in both Montgomery (1991) and DG, one observes significant disparities in adoption by a trait when social ties display homophily in that trait – even when the trait itself is *unrelated* to adoption advantage. Montgomery's model implies surplus wage inequality by gender (which cannot be attributed to ability differences alone) if social networks exhibit homophily by gender (which itself does not affect employment outcomes, but only the size of the initial employed pool). Similarly, DG find inequality in Internet adoption by race (above and beyond what would be expected based on income and education differences) when social ties are homophilous with respect to race (a characteristic unrelated to individuals' adoption propensity).

These patterns should not surprise us given that homophily in a trait changes the network structure, and thus network diffusion, even if that trait does not directly affect individuals' adoption probability. Then, any structural factor that relates to homophily should also have implications for inequality.

One structural factor that moderates the implications of homophily is *consolidation* – the correlation among different traits in a society. Indeed, in their influential work, Blau and Schwartz (1984) declared homophily and consolidation as key parameters shaping social interactions. Low levels of homophily and consolidation, the authors predicted, ensure cross-cutting social ties, and as a result, social cohesion in a community.[4]

Centola (2015) tested this idea with an agent-based model, and reached unexpected conclusions. Low levels of homophily and consolidation lead to a random network structure, and fail to support the diffusion of a common norm (a proxy for social cohesion). Middle levels of homophily and consolidation, however, induce overlapping patterns of connections (or 'wide bridges') that are optimal for *complex contagion* – the process by which norms diffuse via reinforcement from multiple network alters (Centola and Macy 2007).

Centola also showed that homophily and consolidation interact in their effect on diffusion. When consolidation is relatively high, for example, only low levels of homophily can create the social structure necessary to support successful diffusion. Because various traits are highly correlated, even slight increases in homophily lead to a highly-balkanized network, and stop diffusion in its tracks. When consolidation is at middle levels, however, middle and high levels homophily can also support effective diffusion. Because the traits are not as correlated in this case, increases in homophily do not immediately translate into a segregated network.

Implications of homophily and consolidation for inequality

Centola's model is concerned only with network formation and diffusion, but his results could be used to understand how homophily and consolidation, in concert, can contribute to inter-group inequality. In this paper, we use Centola's model, but modify it to resemble the DG set-up, where there is status differentiation in adoption probability. In effect, then, we replicate DG's analysis with synthetic (rather than real-life) agents and generic (rather than calibrated) parameters, which allows us to vary not just homophily, but also consolidation (a factor the authors could not consider as they relied on the GSS sample with a fixed covariance structure).

We argue that the effect of homophily on inter-group inequality depends on the level of consolidation in a society. Consider the extreme case of full consolidation, where traits in a population are perfectly correlated. If we know a person's income, for example, we can perfectly predict his or her education, residential neighborhood, and so on. There is, effectively, a single axis of differentiation. Now consider the other extreme of no consolidation, where individuals are randomly scattered in the multi-dimensional trait space. If we know a person's income, in this case, we still have no idea what their education or neighborhood is. There are multiple axes of differentiation.

We expect that the contribution of homophily to inter-group differences in diffusion will vary greatly across these two cases. In the full consolidation case, even low levels of homophily will be sufficient to concentrate advantage, and generate differential diffusion across groups. Let's presume (like DG) that there exists a status dimension, defined to be positively related to adoption probability, for instance, income. In a full consolidation setting, if individuals have a slight preference for similar alters (in terms of multiple characteristics, say income and education), then high income individuals will be more exclusively connected because they are also highly educated. Due to consolidation (even with dimensions that are not status related), an individual's advantage in the status dimension will also be reflected in network alters, further compounding the differences in adoption between high- and low-income groups.

In the no consolidation case, by contrast, the power of homophily in connecting high-status individuals, and in concentrating advantage, will be diluted by the lack of relational concentration in the status characteristic. Even under high homophily, since high earners could have any level of education, the preference for similar alters will not generate well-defined and exclusively high-earning groups. The advantage in adoption that comes from the status attribute of income, in other words, will not be consolidated.

This logic leads us to qualify DG's findings. We argue that the high levels of inter-group inequality the authors observed, and connected to homophily, are in part due to the high level of consolidation in the sample from which the agents are drawn. Specifically, in the GSS sample, income, education and race are strongly correlated. Therefore, even a small degree of homophily (based on all three characteristics) is likely to generate a highly- balkanized network structure, and lead to inter-group differences in adoption. We hypothesize that homophily bias alone will not be sufficient to generate inter-group inequality under low levels of consolidation. Below, we describe our modeling strategy to test this hypothesis.

METHODS

We use an agent-based model to create artificial worlds of individuals with social identities and social ties. Similar to Centola's (2015) model, individuals are first assigned social identities, or sets of characteristics. An individual's identity defines his or her social distance to the other members of the population. Individuals then form connections based on social distance. Once ties are established, individuals have the opportunity to influence behaviors of their network alters. This set-up closely resembles DG's model, but differs in using synthetic (rather than real) identities for individuals.

Generating the population

To formalize, we assign individuals to a set of positions that can take on *H* possible values (heterogeneity) within *D* social dimensions (complexity). There are *G* individuals with each social position, and hence, *N* (= *H* x *G*) individuals in the population. DG consider a fixed number of social dimensions (income, education, and race) with a fixed degree of heterogeneity. There are three equal-sized groups based on income (high, medium, and low), three groups based on educational degree (bachelor's, high school, middle school or less), and two groups based on race (white and black). In our model, we make the number of dimensions (*D*), the number of social positions in each dimension (*H*), and the number of individuals in each position *(G*) flexible to be able to vary the extent of social differentiation.

Establishing network ties

DG compute the social distance between all pairs of individuals (defined as the Euclidean distance based on standardized values of income, education, and race), and then establish ties between individuals such that homophily bias occurs with a given probability. We follow a similar logic, but create a more complex architecture. In a nutshell, this architecture allows us not only to define the distance between individuals according to their social positions to control homophily ,but also to define the distance between positions across those dimensions, in order to control consolidation in a similar way.

[FIGURE 1 HERE]

Figure 1 illustrates the set-up that is developed originally by Watts, Dodds and Newman (2002), and refined subsequently by Centola (2015). The figure shows a branching tree of $L = 4$ levels with a branching ratio of $B = 2$. This tree offers a visual representation of a population with a single social dimension ($D = 1$), and a hierarchical structure of $H = 8$ social positions (circles) that include a group of $G = 6$ individuals (black dots) each. Social distance between individuals in the same group is defined as $x = 1$. Social distance between individuals in different groups equals one plus the number of steps it takes to reach the fork in the tree where the individuals share the closest common branch. (e.g., distance between individuals i and j in the figure is $x_{ij} = 3$). To see how this construction defines a hierarchy, imagine that the social dimension in the figure is occupation, where the left four positions (circles) represent management, and the right four positions represent workers. The first branching point in the tree, then, captures the social difference between management and workers – two classes of positions that are maximally distant ($x = 4$). The second branching point within management allows for finer-grained social distances, for instance, between executives and middle management.

Occupation is a single dimension of social life, and one can think of many other dimensions, such as education, income, or residential neighborhood. Similar to Centola (2015), we introduce multiple dimensions in our model by replicating the tree-like structure in Figure 1 for each dimension. We then apply the same scale used to measure social distance between individuals within a given dimension to capture the distance between a single individual's social positions across different dimensions. This, latter, distance equals the number of steps up the branching tree to find a common ancestor if the two trees were to be super-imposed on one another. For example, when an individual is located in the same position (e.g., the left-most circle) across two social dimensions (e.g., income and education), the distance between positions is 1. When an individual is located in positions that are at the opposite ends of the spectrum for the two dimensions (e.g., the left-most circle for income, and the right-most circle for education), the distance between positions is 4.

Using the distance between positions within a single dimension, and the distance between positions across dimensions, we control degrees of homophily and consolidation, respectively, as follows. For each of the *N* individuals, we start with a dimension, d_1, and place individuals in a position randomly. To determine positions in the other dimensions, we rely on the consolidation parameter (β). That is, for each subsequent dimension, *d*, we draw a random social distance (*y*) among positions for each individual with probability

$$P(y) = c \cdot e^{-\beta y} \qquad (1)$$

where β is the consolidation parameter and *c* is a normalizing constant. We assign social position at random among all positions in dimension *d* less than or equal to social distance *y* from the individual's position in d_1. For intuition, note that when consolidation is at its minimum ($\beta_{\min}$ = -1), the probability for the maximum social distance *y* is the highest. This can be interpreted as lack of consolidation as an individual is assigned a position at random, regardless of position in d_1. As consolidation increases, individual's positions become more correlated across dimensions ($\beta_{\max}$ = 3).

Similar to Watts et al. (2002) and Centola (2015), we set homophily based on the *shortest* social distance (i.e., one plus the number of steps up the tree to reach a common root) between a pair of individuals across *all social dimensions*. This captures the intuitive notion that closeness in one dimension (e.g., education) is sufficient to connote affiliation, for example, when geographically and ethnically distinct researchers collaborate on the same project given the same social position in the occupational dimension. A useful property of this metric is that it violates the triangle inequality, which states that if individuals i and j share a group in one dimension such that $x_{ij} = 1$, and similarly, if individuals j and k share a group in another dimension such that $x_{jk} = 1$, it is perfectly feasible to have person i and person k be socially distant such that $x_{ik} > x_{jk} + x_{ij}$.

After defining social distance between all individuals in the population, we introduce the homophily parameter (α) and start building social ties among individuals. Individuals start with no ties. We select an individual i randomly from among all individuals with available ties (that is, remaining individuals whose existing ties < degree size, Z). For each individual, we draw a random social distance x with probability

$$P(x) = c \cdot e^{-\alpha x} \qquad (2)$$

where α is the homophily parameter and c is a normalizing constant. We then choose a random individual j to establish a social tie from among all individuals at distance x or less from the individual i. Note that homophily parameter defines the maximum social distance an individual is able to tolerate when making a tie, and thus effectively sets an individual's "search radius" in selecting network alters. High values of the homophily parameter imply a small search radius, and a strong preference for similar others. For intuition, consider that under the maximum level of homophily (α_{max} = 3), individuals can only make social contacts with other individuals at a social distance of 1 (with constant *c* set accordingly). Decreasing the homophily parameter increases the search radius.

We continue the tie formation process until individuals in the population, on average, have the same pre-set degree size, *Z*. An important simplification (similar to DG) is to assume ties to be symmetrical. That is, if individual *i* establishes a tie with individual *j*, individual *j* is also considered to have established a tie to individual *i*.

Modeling diffusion and inequality

We use Centola's (2015) architecture outlined above to generate a population of particular characteristics, to establish network ties among individuals, and to model the diffusion of a beneficial practice. Similar to DG, however, our goal is to observe group-specific diffusion rates, and consider their implications for social inequality. This goal requires a modification in Centola's set-up, which assigns all individuals the same *adoption threshold*, defined as the number of adopters in ego's network necessary to induce the ego to adopt a practice. Centola's model considers practices that diffuse through *complex contagion*, that is, those with an adoption threshold of 2 or higher. DG's analysis also focuses on complex contagion, but allows the adoption threshold to vary in the population. Indeed, a particular feature of their model is to make the adoption threshold an inverse function of an individual's economic and cultural resources. In this model, each individual has a *reservation price* – a price at which he or she is willing to adopt the practice. This price is an increasing function of one's income, education, and the number of network alters who have already adopted. Then, all else equal, high income and high education individuals have the highest reservation prices (and hence the lower implied thresholds for adoption).

We introduce this important feature into Centola's set-up as follows. First, we let an arbitrarily chosen social dimension, *d*, indicate status for all individuals (similar to income or education in the DG model). We designate a given proportion (*P*) of the population as high status, an identical proportion as low status, and the remainder as medium status. Similar to Gondal (2014), we vary the proportion of high status individuals in order to test the sensitivity of our results to the size of the elite.

Second, we make the adoption threshold an inverse function of status. Specifically, we assign high status individuals an adoption threshold of 1, medium-status individuals a threshold of 2, and low-status individuals a threshold of 3. Note that an adoption threshold of 1 implies a *simple contagion* process among the elite (to use Centola and Macy's (2007) terminology), while thresholds of 2 and higher bring about a *complex contagion* process among the rest of the population. Because we set the proportion of high-status and low-status individuals to be equal, the average adoption process still resembles a complex rather than simple contagion, and hence, our results at the population level closely follow those of Centola (see Figure 2).

To initiate the diffusion process, like Centola (2015), we randomly seed one high-status individual and their network (that is, individual's first-order ties) as the first adopters. This approach follows from DG, where the reservation prices are highest for, and hence the initial adoption is most likely among, high-income and high-education individuals. We also randomly seed one low-status individual (but not their network) in order to avoid undefined odds ratios in subsequent analysis of inter-group inequality. This deliberate seeding of a low-status individual – absent in DG – renders our estimates of inequality relatively conservative.

Following the initial seeding, at each time period, all individuals who have not adopted the practice simultaneously make decisions on whether or not to adopt, and the algorithm outputs the cumulative and group-specific percentages of adopters by status. The process ends after $T = 50$ time periods, which, as the results indicate, provides ample time to reach equilibrium diffusion levels.

Parameters and robustness checks

In this study, we focus on the effects of homophily (α) and consolidation (β) on network diffusion and inequality; therefore, we keep the remaining parameters fixed. Below, we list all the parameters used in generating our results. Centola (2015) has investigated the robustness of the results to variations of all parameters, but one. Therefore, here, we report the sensitivity of our findings to P (the proportion of high-status individuals in the population), which is the only parameter not included in Centola's analysis, and refer the reader to the original paper for all other checks. We have run our models with P ranging from 1/16 to 5/16. We have not considered higher values as they are not meaningful substantively. We found qualitatively similar results (available upon request) for P between 2/16 and 5/16. We observed slightly different results for the case of $P = 1/16$ where the difference in adoption inequality for low and moderate levels of homophily is not statistically significant under high consolidation (whereas it consistently is for higher values of P), and where high levels of homophily impede diffusion under low consolidation (whereas it actually helps it for higher values P). Therefore, our results are consistent for cases where high-status adopters with an initial adoption advantage make at least 1/8th of the population.

Population parameters

- D = complexity (number of dimensions) = 10
- H = heterogeneity (number of social positions) in a dimension = 16
- G = group size (number of people in each position) = 50
- N = population size = $H \times G$ = 800
- P = proportion of high status individuals = 2/16

Network parameters

- α = level of homophily, ranging from -1 to 3
- β = level of consolidation, ranging from -1 to 3
- Z = average degree (number of ties for each person) = 5

Adoption parameters

- T = time intervals = 50

We code our algorithm in R (version 3.3.1). To account for the randomness in initial seeding, and in the social positions of individuals by status, we run the simulations 100 times for each set of parameters, repeating the entire process of population construction, network formation, and diffusion. These repeat runs give us means and standard deviations across all realizations of specific sets of social conditions. In what follows, we report not just the average diffusion outcomes (as both Centola (2015) and DG do), but also the distribution of diffusion outcomes across repetitions. As a methodological point, we note that conclusions based on average observations might be misleading for particular parameter combinations, and at particular time-points.

RESULTS

Overall diffusion under homophily and consolidation

Before considering group-specific diffusion rates, we investigate overall diffusion patterns, and confirm that our results are similar to Centola's (2015) findings. This is important because our model has implemented adoption thresholds that vary inversely with individuals' status, which might disrupt the diffusion dynamics in the original model based on fixed adoption thresholds. While we keep the average adoption threshold equal to that in Centola, it is still not obvious that diffusion patterns will remain similar with a given proportion (P = 2/16) of high- and low-status individuals (who require less and more social reinforcement to adopt, respectively). In Figure 2, we plot overall equilibrium cumulative adoption rates (z-axis) under status-based thresholds while varying consolidation (y-axis) and homophily (x-axis), keeping all other parameters at values listed in the preceding section.

[FIGURE 2 HERE]

Introducing status-based thresholds as in DG does not change the relationship between homophily, consolidation, and diffusion at the population level. Similar to Centola (2015), we observe that (i) homophily and consolidation interact in their effect on diffusion, and (ii) moderate levels of homophily and consolidation best support diffusion.

We now turn to the implications of our model for DG's conclusions on how homophily affects diffusion and inter-group inequality. While DG treat consolidation as fixed, our model allows us to vary it. For the sake of simplicity, in the remainder of this paper, instead of showing the full range of values for homophily and consolidation (as in Figure 2), we display results for (i) four levels of homophily, including random mixing (α = -1), low (α = 0), medium (α = 1), and high homophily (α = 2), and (ii) two levels of consolidation, low (β = 0) and high (β = 2). This exposition allows us to observe adoption rates under varying degrees of homophily and consolidation, and also over time, similar to DG.

We present not just average diffusion patterns, but also the 95 percent confidence bands around the average across multiple realizations. To clarify, the variance in simulations is not generated by different parameter values, but by randomly occurring differences in the social positions or networks of initial seeds under a given set of parameter values. It is important to consider the full distribution of simulations to ensure that our results are stable (and not driven by volatile realizations of the model), and that the differences we point to across multiple scenarios are statistically significant.

Panel B in Figure 3 shows the cumulative adoption (y-axis) over time (x-axis) by homophily in a high-consolidation world. Note that Figure 3 can be understood as a cross-section of Figure 2, but with the added dimension of time, and confidence bands (gray regions) around the average diffusion curve across 100 realizations per a given set of parameters.

We observe several patterns. First, low homophily (α = 0) leads to the highest diffusion level in equilibrium, where the practice consistently reaches full saturation in the population. Medium (α = 1) and high (α = 2) levels of homophily lead to much lower – and statistically indistinguishable – adoption levels in equilibrium. Second, low homophily (α = 0) leads to the slowest rate of diffusion initially, surpassed by the equally fast medium and high levels of homophily. In line with DG's observations, then, while homophily boosts adoption speed initially (as the practice can spread quickly among the advantaged individuals socially connected to one another), it hurts the equilibrium adoption level (as the practice fails to reach the less advantaged individuals).

[FIGURE 3 HERE]

Another pattern in the figure seems to defy DG's expectations. The no homophily case (α = -1) leads not only to the slowest diffusion rate, but also to low (and highly variable) overall adoption levels in equilibrium. Centola (2015) explains this finding with reference to complex contagion. When a practice requires reinforcement from multiple social ties to be adopted, networks need to have 'wide bridges', that is, overlapping social ties, to effectively spread that practice. Such wide bridges are only possible with some degree of homophily and consolidation. Under no homophily, even high levels of consolidation are not sufficient to produce the social structure needed to sustain diffusion. (Note that, in Figure 2, when homophily is less than -0.5, diffusion is flat at zero, regardless of the level of consolidation.) Why then, one can ask, did DG not observe this pattern in their analysis?

The answer, we believe, is simple. Because DG relied on the GSS data, and a fixed number of individuals (N=2,257) to establish a large number of social ties (an average of 28 ties per person, as given by the actual degree sizes reported by each individual), the resulting networks ended up with some degree of (incidental) homophily even when homophily bias was set to zero. After all, even random mixing can still bring about homophilous tie formation (especially if there are few network alters to choose from). (By comparison, in our model, there are 800 individuals to establish an average of 5 ties per person.) Then, DG's no homophily case is likely to resemble our low homophily condition, and there is no contradiction in the findings.

This significant overlap between DG and our results, however, disappears under low consolidation, as shown in panel A in Figure 3. In this case, medium (α = 1) and high (α = 2) levels of homophily, although not statistically distinguishable from one another, lead to much higher adoption rates and equilibrium levels compared to low (α = 0) and no homophily (α = -1) conditions. Under low consolidation, then, homophily actually helps both the speed and level of diffusion, directly opposing DG's conclusions based on a high-consolidation setting.

The explanation for this pattern, again, comes from the conditions for complex contagion. Under low consolidation, only higher degrees of homophily can create overlapping social ties that are key to successful diffusion. When consolidation and homophily are both low, the diffusion fails the take off (as in the inner-left corner of Figure 3).

Comparing panels A and B reveals another insight that is not obvious from Figure 2 (and, thus, Centola's original analysis). The low-consolidation condition (panel A) creates a lot more variance in diffusion than its high-consolidation counterpart (panel B). Because Figure 2 only displays the average outcome, it is hard to know whether, say, a diffusion level of 0.5 indicates that diffusion consistently reaches half the population, or that diffusion is at full saturation (100%) half the time, and fails completely in the other half. In Figure 3, we see that the former interpretation (of stable results) is reasonable for high-consolidation worlds, where the confidence bands around the outcome are rather tight. But, the latter interpretation (of highly volatile realizations) is more accurate for low-consolidation worlds, where the confidence bands cover a broad range of outcomes. For example, for medium homophily (α = 1), diffusion reaches 75% at equilibrium as many times as it lingers at 25%. The average diffusion curve settles at 50%, and hides this large variation. This observation is important for empirical work. Because in real-life many characteristics are clustered (e.g., income is highly correlated with education, neighborhood, wealth, and so on), low consolidation worlds are rare. Researchers not only get few chances to observe such worlds, but, given the inherent volatility of diffusion outcomes, they also are more likely to reach faulty conclusions based on their observations (compared to high-consolidation settings).

Group-specific diffusion under homophily and consolidation

We now turn to diffusion outcomes for different status groups, and inequality in those outcomes across groups. Figure 4 shows the log odds ratios of adoption rates between high-status and low-status adopters (1/8th of the population each) under varying levels of homophily, contrasting low consolidation (panel A) and high consolidation (panel B) worlds. [5] Lower log odds ratios indicate lower inequality.

[FIGURE 4 HERE]

First, it is important to note that the 95% confidence bands (gray region) around the inequality curves are much wider when consolidation is low. This is consistent with the earlier observation that low consolidation condition generates a lot more variance in diffusion outcomes (figure 3, panel A) compared to its high consolidation counterpart.

Second, we observe that, when homophily is at moderate to high levels, log odds ratios are larger when consolidation is high. For example, log odds ratios are greater than 4.5 for high (α = 2) and medium (α = 1) homophily cases when consolidation is high, but they linger around 1 when consolidation is low. This observation is perhaps not surprising. When individuals strongly prefer to associate with similar others, less correlation between different dimensions of social life generates less social reinforcement for exclusive groups based on status dimension. Thus, one finds less difference between high- and low-status individuals' outcomes when different dimensions of social life are not strongly correlated (i.e., under low consolidation).

Third, and conversely, when homophily is low (α = 0), log odds ratios are smaller (0 vs. 2) at equilibrium when consolidation is high (β = 2) rather than low (β = 0). And the reason is simple. Under low homophily, a practice reaches full saturation (and hence, zero inequality) only under high consolidation (see α = 0 condition in Figure 3, panel B). The practice remains moderately diffused (<10%) under low consolidation (see the same condition in panel A), and presents some inequality by status.

Fourth, when individuals establish ties randomly (α = -1), there is no difference between high- and low-consolidation worlds in terms of equilibrium inequality. The practice remains minimally diffused in both cases: only slightly above 0% under low consolidation (Figure 3, panel A), and around 20% but with a large variance under high consolidation (panel B). In both cases, the differences are likely driven by the status imbalance in initial seeding, that is, by the fact that there are more initial adopters that are high rather than low status.

Another way of thinking about these results is that, save for the α = -1 case, the relative implications of homophily levels on inequality in diffusion are flipped for high and low consolidation worlds. The conditions that generate the most to least inequality are α = 0, 1, 2 versus α = 2, 1, 0 for low and high consolidation worlds, respectively. The reasons why homophily exacerbates inequality in the high consolidation world is intuitive and consistent with DG. The counterintuitive finding in low consolidation worlds can be explained by the following intuition. In low consolidation worlds, homophily helps reduce inequality by supporting diffusion. High homophily creates the bridge width that is necessary for successful diffusion. And inequality is always the lowest when diffusion reaches the whole network. Thus, in low consolidation worlds, high homophily decreases inequality by supporting overall diffusion. Although there is an initial exacerbating effect of homophily on inequality, the benefit to diffusion of high homophily for high status adopters is eventually tempered by its later support for overall diffusion (this starts to occur at T = 15). By contrast in the low consolidation world where homophily is also low, diffusion is incomplete, so inequality stays at the same moderate initial level.

DISCUSSION

DG's work unequivocally argues that homophily exacerbates inequality in adoption if a practice is subject to network externalities, and if the practice is initially more likely to be adopted by high-status individuals. Our findings confirm some of their findings, while qualifying or extending others.

Confirming DG, we find that homophily exacerbates adoption inequality if consolidation is moderate to high. When social dimensions are strongly correlated, increasing homophily makes networks more segregated by status, and locks the practice into network regions where high-status actors cluster.

Qualifying DG, we show that homophily reduces adoption inequality if consolidation is low. When social dimensions are largely independent, increasing homophily is not sufficient to segregate social networks, but it is helpful to create overlapping social ties, or 'wide bridges', across which behaviors can spread. Regardless of level of consolidation, homophily creates an advantage for high status initial adopters because it makes it more likely for other high-status adopters to be closely connected to the initial adopters. When consolidation is low, this advantage disappears if homophily is high enough to generate the overlapping social ties for diffusion to reach lower status individuals who require more social reinforcement for adoption. In this case, eventual successful diffusion by low status individuals overcomes inequality.

Extending DG, we observe that some homophily is needed for effective diffusion. Under no homophily (a case DG could not test effectively due to the constraints of their data), a network does not have the sufficient structure to coordinate behavior, regardless of the level of consolidation in the population.

We also make a few finer points. First, low consolidation worlds, all else equal, create more variance in diffusion and inequality outcomes compared to high consolidation cases. Second, the diffusion and inequality trajectories shift over time. For example, while diffusion is quite fast in a high homophily condition, its equilibrium level is low compared to a low homophily condition. Or, while inequality increases steeply at first in a high homophily setting, it can end up at a lower equilibrium compared to that in a low homophily case. Put differently, one can settle on the wrong conclusions if one observes only average values, or collects data mid-process (that is, prior to equilibrium).

CONCLUSION

A major insight in sociology suggests that social networks can provide access to useful resources or positive influences that help individuals succeed (Portes 1998). Another key insight indicates that early advantages can lead to benefits that pre-dispose individuals to obtain more advantages over time (Merton 1986, DiPrete and Eirich 2006). These two ideas go well together in that social networks can be a mechanism for generating cumulative advantage.

Recent formal and computational analyses have made this connection explicit by studying homophily. Studies have linked, for example, network effects in job search to wage inequality (Montgomery 1991, Calvó-Armengol and Jackson 2004) or the peer effects in educational choice to disparities in educational attainment between groups (Manzo 2013). DiMaggio and Garip (2011) have offered arguably the most general theoretical statement, and identified three necessary conditions for network effects to exacerbate inequality in the adoption of a beneficial practice. First, the authors have argued, adoption should be more likely among the more advantaged individuals (for example, high earners or highly educated). Second, adoption should be more likely if one's peers have adopted. That is, network effects should be positive. And, third, and most importantly, networks should exhibit homophily (tendency for ties among similar individuals) with respect to traits related to adoption.

In this paper, we extend DiMaggio and Garip's (2011) analysis, and discover a fourth, and crucial, condition neglected in prior work. For social networks to exacerbate inequality, we argue, characteristics in a population need to consolidated (that is, highly correlated). Without consolidation, homophily in any given characteristic is not be sufficient to segregate social networks, and isolate adoption to particular segments of the population.

Earlier work has not considered consolidation as a factor in network formation either because it has relied on formal analysis where networks were considered given (e.g., Calvó-Armengol and Jackson 2004) or because it has used computational models calibrated to a particular real-life setting (e.g., DiMaggio and Garip 2011, Manzo 2013).

We draw inspiration from Blau and Schwartz's work (1984, p.12), which declared consolidation to be of "prime significance for intergroup relations." Recently, Centola (2015) confirmed this insight with an agent-based model, and demonstrated that homophily and consolidation together shape network structure and diffusion outcomes.

In this paper, we start with Centola's generic model, and use it to first replicate, and then to extend, DiMaggio and Garip's findings. We generate a sample of synthetic agents, and introduce status differences between them. We then give high-status individuals a small advantage in adoption. We vary levels of homophily and consolidation, and observe whether initial advantages by status are compounded via network effects.

We report several findings. First, similar to DiMaggio and Garip, we find that homophily exacerbates adoption inequality, but only if consolidation is relatively high. Second, different from DiMaggio and Garip, we show that homophily actually alleviates adoption inequality if consolidation is low.

These patterns all owe to an important insight gained from Centola's (2015) analysis: Low to moderate levels of homophily and consolidation generate social networks with overlapping ties, or 'wide bridges', that can support the diffusion of a practice through reinforcement from multiple network alters (or, what Centola and Macy (2007) call 'complex contagion'). Higher levels of homophily or consolidation, however, lead to segregated or balkanized social ties, and can stop diffusion in its tracks.

In a high consolidation world, increasing homophily brings about the latter, detrimental, structure for diffusion, and also makes the status-based divide in adoption deeper. But, in a low consolidation world, rising homophily leads to the former, favorable, structure for contagion, and alleviates status-based differences in adoption.

Our analysis allows us to make a number of methodological points. First, prior work using agent-based models routinely presents the average diffusion curve (computed over multiple simulation runs), but not the confidence bands around the average. We call future research to be more attentive to, and report, the variation across multiple realizations. In our case, the interpretation of our results would be quite different had we just relied on just the average patterns. Second, while we recognize that agent-based models can be a useful tool when calibrated to real-life cases (as in DiMaggio and Garip's work), we also argue that such applications might not always lead to fully generalizable conclusions. Researchers should pay more attention to how much of their results are driven by the model parameters, and how much is due to the particular setting the parameters or data are set to capture.

We believe our work opens to doors to studying many other factors that we did not consider in our analysis. For example, recent work considers the interactions between different kinds of homophily (e.g., that based on individual choice and that induced by population compositions in institutional settings) in producing particular network structure and diffusion patterns (Kossinets and Watts 2006). Recent work also shows how adoption of a practice can pave the way to establishing new dimensions for status differentiation, and deepen inequality (Gondal 2014). Future work, therefore, can incorporate patterns of induced homophily or path-dependent diffusion trajectories for multiple practices into our set-up.

Recent work also considers how population processes (such as in- and out-migration) can affect population composition, network structures, as well as diffusion outcomes (Garip and Zhao, forthcoming), suggesting another possible direction of extension. Finally, qualitative work finds that individuals' perceptions of social norms (and their respective adherence to them) may depend not just on their position in the network, but on the particular patterns of social interactions, and in particular, to degrees of exposure to network peers (Shepherd 2017). Therefore, future attention could also be given to incorporating social interaction and exposure into modeling frameworks with differential impacts on contagion.

REFERENCES

Allison, Paul, J. Scott Long and Tad K. Krauze. 1982. “Cumulative Advantage and Inequality in Science.” *American Sociological Review* 47: 615-25.

Arrow, Kenneth J., and Ron Borzekowski. 2004. “Limited network connections and the distribution of wages.” Finance Economics Discussion Series, 2004-41, Federal Reserve, http://econpapers.repec.org/paper/fipfedgfe/2004-41.htm

Blau, Peter M., and Joseph E. Schwartz. 1984. *Cross-cutting social circles: Testing a macrosociological theory of intergroup relations*. Orlando: Academic (1984).

Boyd, Monica. 1989. “Family and personal networks in international migration: Recent developments and new agendas.” *International Migration Review* 23: 638-70.

Burke, Mary A., and Tim R. Sass. 2013. "Classroom peer effects and student achievement." *Journal of Labor Economics* 31 (1): 51-82.

Bruch, Elizabeth and Jon Atwell. 2015. “Agent-based models in empirical social research.” *Sociological methods & research* 44(2): 186-221.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood Choice and Neighborhood Change. *American Journal of sociology*, *112*(3), 667-709.

Centola, Damon. 2015. “The Social Origins of Networks and Diffusion.” *American Journal of Sociology*, 120(5): 1295-1338.

Centola, Damon and Michael Macy. 2007. “Complex contagions and the weakness of long ties.” *American Journal of Sociology* 113(3): 702-734.

Calvó-Armengol, Antoni, and Matthew O. Jackson. 2004. "The effects of social networks on employment and inequality." *American economic review* 94 (3): 426-454.

Cole, Stephen. 1970. “Professional Standing and the Reception of Scientific Discoveries.” *American Journal of Sociology* 76(2): 286-306.

Epple, Dennis, and Richard E. Romano. 2011. "Peer effects in education: A survey of the theory and evidence." In *Handbook of social economics*, Eds. J. Benhabib, A. Bisin, M.O. Jackson, vol. 1, pp. 1053-1163. New York: Elsevier

DiPrete, Thomas A. and Gregory M. Eirich. 2006. “Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments.” *Annual Review of Sociology* 32:271-97.

Elwert, Felix, and Christopher Winship. 2014. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual Review of Sociology* 40: 31-53.

Fischer, Claude S. 1977. *Networks and Places: Social Relations in the Urban Setting*. New York: Free Press.

Gamoran, Adam. 2011. "Designing instruction and grouping students to enhance the learning of all: New hope or false promise?" *Frontiers in Sociology of Education* 1(1): 111-126.

Garip, Filiz. 2008. "Social capital and migration: How do similar resources lead to divergent outcomes?" *Demography* 45(3):591–617.

Garip, Filiz and Linda Zhao. Forthcoming. "Mixing It Up? Social Cohesion Under Population Mobility." In *The Architecture of the Social*, edited by Peter Hedstrom and Petri Ylikoski.

Gondal, Neha. 2014. "Inequality preservation through uneven diffusion of cultural materials across stratified groups." *Social Forces* 93(3):1109-1137.

Granovetter, Mark. 1974. *Getting a Job*. Cambridge, MA: Harvard University Press.

Granovetter, Mark. 1978. "Threshold models of collective behavior." *American Journal of Sociology* 83: 1420–43.

Kandel, Denise B. 1978. "Homophily, selection, and socialization in adolescent friendships." *American Journal of Sociology* 84(2): 427-436.

Kossinets, Gueorgi, and Duncan J. Watts. 2006. "Empirical analysis of an evolving social network." *Science* 311(5757): 88-90.

Lin, Nan. 2002. *Social capital: A theory of social structure and action*. Cambridge, UK: Cambridge University Press.

Liu, Ka-Yuet, Marissa King, and Peter S. Bearman. 2010. "Social influence and the autism epidemic." *American Journal of Sociology* 115 (5): 1387-1434.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies* 60(3): 531-542.

Manzo, Gianluca. 2013. "Educational choices and social interactions: A formal model and a computational test." *Class and Stratification Analysis*: 47-100.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52:122-31.

_____. 1988. “Homogeneity in confiding relations.” *Social Networks* 10: 57-76.

Marsden, Peter V. and Elizabeth H. Gorman. 2001. “Social networks, job changes, and recruitment.” In *Sourcebook of Labor Markets: Evolving Structure and Processes,* ed. I. Berg, A.L. Kalleberg, pp. 467–502. New York: Kluwer Acad./Plenum.

Massey, Douglas S. and Kristin E. Espinosa. 1997. “What’s driving Mexico-US migration? A theoretical, empirical, and policy analysis.” *American Journal of Sociology* 102: 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 James M. Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27(1): 415-444.

Merton, Robert K. 1968. “The Matthew Effect in Science.” *Science* 159:56-63.

Pampel, Fred C., Patrick M. Krueger, and Justin T. Denney. 2010. "Socioeconomic disparities in health behaviors." *Annual Review of Sociology* 36: 349-370.

Portes, Alejandro. 1998. "Social capital: Its origins and applications in modern sociology." *Annual Review of Sociology* 24: 1-24.

Reskin, Barbara. 1977. “Scientific Productivity and the Reward Structure of Science.” *American Sociological Review* 42(3): 491-504.

Reskin, Barbara F., Debra B. McBrier, and Julie A. Kmec. 1999. “The determinants and consequences of workplace sex and race composition.” *Annual Review of Sociology* 25(1): 335-361.

Rivera, Mark T., Sara B. Soderstrom, and Brian Uzzi. 2010. "Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms." *Annual Review of Sociology* 36: 91-115.

Rossman, Gabriel, Ming Ming Chiu, and Joeri M. Mol. 2008. "Modeling diffusion of multiple innovations via multilevel diffusion curves: Payola in pop music radio." *Sociological Methodology* 38(1): 201-230.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly journal of economics* 116 (2): 681-704.

Sacerdote, Bruce. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" *Handbook of the Economics of Education* 3: 249-277.

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.

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

Small, Mario Luis, David J. Harding, and Michèle Lamont. 2010. "Reconsidering culture and poverty." *The Annals of the American Academy of Political and Social Science* 629: 6-27.

Smith, Kirsten P., and Nicholas A. Christakis. 2008. "Social networks and health." *Annual Review of Sociology* 34: 405-429.

Smith, Sandra Susan. 2005. "`Don’t put my name on it’: Social capital activation and job-finding assistance among the black urban poor." *American Journal of sociology* 111(1): 1-57.

Van de Rijt, Arnout, Soong Moon Kang, Michael Restivo, and Akshay Patil. 2014. “Field experiments of success-breeds-success dynamics.” *Proceedings of the National Academy of Sciences* 111(19): 6934-6939.

VanderWeele, Tyler J. 2011. “Sensitivity analysis for contagion effects in social networks.” *Sociological Methods & Research* 40(2):240–55

Watts, Duncan J., Peter S. Dodds and Mark E. Newman. 2002. “Identity and search in social networks.” *Science* 296(5571): 1302-1305.

ENDNOTES

---

[1] Network effects are present when the probability that an actor will adopt a practice is an increasing function of the network alters who have already adopted that practice. There are useful reviews of networks effects on several outcomes including migration (Boyd 1989), education (Sacerdote 2011, Epple and Romano 2011, Marsden and Gorman 2001), and on health (Pampel, Krueger and Denney 2010, Smith and Christakis 2008).

[2] The authors used Skvoretz's (1990) definition of homophily (or 'tau bias') which is the probability that an ego will select a similar alter above and beyond the probability of such a pairing under random choice.

[3] DG defined *network externalities* broadly as applicable to any practice whose value to an individual is an increasing function of the prior adopters in that individual's network. The authors operationalized this mechanism with a 'network term' included in the reservation price. This term implied that an individual's willingness to adopt a practice (i.e., the price he or she is accepting to pay for it) increases linearly with the *share* of adopters in his or her network. In a subsequent review article, DiMaggio and Garip (2012) offered a more specific definition for *network externalities*, and differentiated it from other mechanisms for network effects such as *social facilitation* or *normative influence*. (See Rossman, Chiu and Mol (2008) for a similar analytic framework.) According to the updated definition, the authors argued, network externalities need to be operationalized as a function of the *number* (not share) of adopters in an individual's network. In this paper, we use the updated definition, and operationalize

network externalities with an adoption threshold that depends on the *number* of adopters. This choice carries an implicit assumption that an individual only cares about the adopters in his or her network and is indifferent to the non-adopters (Granovetter 1978, Centola and Macy 2007). See the methods section for details.

[4] Another structural factor shaping homophily is the institutional setting. McPherson et al. (2001) distinguish homophily based on individual *choice* from homophily *induced* by institutions such as schools (Fischer 1977), voluntary associations (McPherson and Smith-Lovin 1987), or work environments (Reskin et al. 1999). Such institutions might include particular group compositions, and thus, exhibit induced homophily in ties even when individuals are selecting alters randomly. Induced homophily is a crucial concept, but it is hard to incorporate into a formal or computational modeling framework in abstract form (that is, without first defining relevant institutions and deciding on their recruitment criteria). Thus, we do not focus on it in this paper.

[5] The log odds ratios between high-status and medium-status adopters are not reported because they are qualitatively similar to the log odds ratios between high-status and low-status adopters. The only difference is that they are, unsurprisingly, slightly smaller.