JAZZ NETWORKS: USING RESPONDENT-DRIVEN SAMPLING TO STUDY STRATIFICATION IN TWO JAZZ MUSICIAN COMMUNITIES*

Douglas D. Heckathorn, Cornell University
douglas.heckathorn@cornell.edu

and

Joan Jeffri, Columbia University Teachers College
jj64@columbia.edu

(April 13, 2003)

* Paper to be presented at the annual meeting of the American Sociological Association, Atlanta, GA, August 2003. This research was made possible by grants from the National Endowment for the Arts, the Centers for Disease Control and Prevention, and the National Institute on Drug Abuse.

Jazz Networks: Using Respondent-Driven Sampling to Study Stratification in Two Jazz Musician Communities

Abstract

The literature on jazz has focused on biographies of major figures, the history of styles, and the influence of jazz on society. Only two studies, one in the Netherlands and one in France, have included obscure musicians as well as the famous, and both were limited by the use of nonprobability sampling methods. This study is based on 564 interviews with jazz musicians in the greater New York metropolitan area and the San Francisco Bay area. It employs respondent-driven sampling, a method that permits representative samples to be drawn from hard-to-reach populations. The analysis extends that method by introducing a means for studying the structure of large social networks, by identifying combinations of in-group affiliation bias (homophily) and out-group affiliation bias (heterophily). The results indicate that cohesion among jazz musicians is based, in part, on race and ethnicity, gender, and age (the latter two of which can also be the basis for stratification), and on style of performance, but the most important determinants are professional contacts, involvement in the jazz community, and primary instrument. Inequality by age and gender are greater in San Francisco, and racial and ethnic boundaries are stronger in New York. Drawing on concepts from the literature on small-world networks, we examine spatial networks, and San Francisco is found to have stronger small-world properties than New York.

Jazz Networks: Using Respondent-Driven Sampling to Study Stratification in Two Jazz Musician Communities

Introduction

The literature on jazz is vast, focusing on the lives of influential musicians (Palmer 2002), the history of jazz at the local or national level (Shipton 2002), the emergence of new styles (Yanow 2001), the venues in which jazz is played (Bjorn and Gallert 2001), the influence of jazz on literature and the arts (Appel 2002), and the social organization of musicians and the larger system of which they are a part (Becker 1951, 1982). These works focus overwhelmingly on the extraordinary—the people, events, and places that have shaped the emergence and evolution of jazz. In contrast, very few studies have examined the jazz musician community in ways that include not only the famous but also the obscure, not only the leaders but also the followers.

A challenge to such a study is the lack of consistent institutional identifiers for jazz musicians—that is, the absence of a sampling frame from which representative samples can be drawn. Because jazz education programs are relatively recent and most jazz musicians learned the craft through informal apprenticeships, there is no consistent educational credentialing. There are also no licensing requirements, such as those required for the practice of medicine or law. Musicians' unions generally do not record the types of music played by their members, so there are no institutional affiliations that serve to define the jazz musician community, and in any case, most jazz musicians do not join such unions. Major venues attract prominent musicians but exclude the less experienced. A further problem is that jazz musicians are such a small group—less than one-third of 1% of the general population—that using a general population survey to construct the sampling frame would be impractical.

Because of those problems, studies of jazz musicians have used nonprobability sampling methods. For example, both IJdens (1999), who studied the jazz musician community in the Netherlands, and Coulangeon (2000), who conducted a similar study in France, emphasized the methodological limitations of conclusions based on a mix of nonsystematic institutional samples and ethnographic interviews. Similar problems arise in studies of other artists. A number of investigators have undertaken surveys based on information culled from membership lists and other organizational documents and have gathered important and valuable information. But they have also clearly articulated the limitations of such approaches to identifying samples and have sought ways to learn more about the networks to which artists belong and the interaction of those networks with one another (Jeffri 1989, 1998; Jeffri et al. 1990a,b,c; Throsby 1983; Throsby and Mills 1989).

Though the boundaries of the jazz musician community are not defined in official records, they are clearly defined by its members. Known for their collegiality and willingness to play with peers, jazz musicians have a clear sense for who is and who is not a member of that community. Because of the way jazz artists work—often freelance, sometimes affiliated, with varying degrees of formal education and great strengths in self-teaching and apprenticeships—the linkages binding the community are largely informal, consisting of relationships that emerge when musicians play together, share information about potential jobs, educate one another, and in other ways function as members of an integrated and cohesive community (Heckathorn and Jeffri 2001). The result is that researchers must in some manner enlist their efforts in the sampling enterprise.

In this article we report on a study of jazz musician communities in the greater New York City (NYC) metropolitan area, including Connecticut and New Jersey, and the San Francisco (SF) Bay area. One aim of the study was to draw a statistically valid sample of jazz musicians from each metropolitan area. We employed a network sampling method developed by the first author (Heckathorn 1997, 2002a), termed respondent-driven sampling (RDS). RDS is part of a new class of sampling methods termed adaptive/link-tracing designs (Thompson and Frank 2000). Whereas in traditional approaches, the sampling plan is fixed before sampling begins, in adaptive sampling, the plan changes as information accumulates during the sampling process. These approaches are more computationally demanding than traditional approaches, but they are also generally more efficient, especially for sampling clustered populations. RDS is now commonly employed in public health studies of hard-to-reach populations (Semaan, Lauby, and Liebman 2002) because it provides not only information on network structure but also estimates of the sizes of population subgroups, including estimates of the variability of those estimates (i.e., standard errors). It is therefore useful for tasks such as estimating HIV seroprevalence among injection drug users and assessing the effectiveness of HIV-prevention interventions (Broadhead et al. 1998; Heckathorn et al. 2002). It has also been used to study Vietnam War–era draft resisters (Hagan 2001). This is the first application of RDS to study a population of artists. Results are based on

264 face-to-face interviews with jazz musicians from the NYC area, and 300 from the SF area conducted from March to November 2001.

A second aim was to study the network structure of both jazz musician communities. Research on social networks generally employs one of two approaches. The first, analysis of global networks, requires a complete enumeration of individuals and the ties by which they are linked. This approach is impractical in studies of large networks, however. For example, the estimates for the numbers of jazz musicians in NYC and SF generated as part of this study are 33,000 and 18,000, respectively; so complete enumeration would not be practical. The second, the egocentric approach, is based on identifying a set of focal individuals and asking them to report on their network partners. This approach is well suited to studying large networks, but because reports by respondents are restricted to public knowledge, the egocentric approach is ill suited for studying matters that are private or matters where knowledge by peers is imprecise. For example, many musicians are secretive regarding their incomes and cannot report accurately on the incomes of their partners, so detailed reports that differentiate among household income, personal income, and income from music would be impossible. RDS constitutes a third approach, for like other link-tracing designs (Klovdahl 1989), it provides not only a sample but also behavioral links between respondents (Heckathorn and Rosenstein 2002). Therefore, affiliation patterns can be studied based on any item in the survey. For example, respondents report only on their own personal incomes, household incomes, and income from music, and the extent to which each type of income affects network structure is then determined by examining the behavioral links among respondents. This portion of our study focused on determining the extent to which affiliations are stratified along demographic lines, including race, ethnicity, gender, and income, versus factors specific to the jazz musician community, including contacts within that community, extent of professional activity, and musical style.

Part I summarizes the RDS method, showing how a reliable and valid sample can be drawn from a population for which no sampling frame is available, thereby permitting estimates to be derived of the composition of the jazz musician communities by race, ethnicity, gender, income, and style of performance. The results show that, like other artist groups, jazz musicians have a strongly positively skewed income distribution: a few have very high incomes, but many have incomes so low it limits their access to essential services, such as health insurance.

Part II shows how RDS can be used to analyze the structure of large social networks based on an index of homophily that measures group-based clustering (Heckathorn 1997, 2002a). This index quantifies the degree to which in-group ties are favored over out-group ties (i.e., positive homophily) or out-group ties are favored over in-group ties (i.e., negative homophily, or heterophily). Types of structures are identified based on the combinations of homophily or heterophily including core-periphery structures and cohesive subgroups. Structures of affiliation are examined with respect to gender, race and ethnicity, style of performance, and principal instrument.

Part III extends the analysis of network structure by introducing a method, termed breakpoint analysis, to analyze the effects of continuous variables, such as network size, age, income, and number of groups with which the respondent performs. In both NYC and SF, affiliation by network size corresponds to a core-periphery structure. An elite of stars has entourages of lesser stars, who in turn have entourages of more minor stars; less well connected musicians thus interact primarily through those who are better connected. The results also reveal substantial differences in the network structure of the two communities. For example, affiliation by age in NYC reflects a cohort structure, with relatively independent music scenes for each age group, whereas in SF affiliation by age reflects a core-periphery structure, with older musicians forming the densely networked core, and younger musicians interacting primarily through those who are older.

Finally, the conclusion examines whether jazz musician networks have small-world properties, a question relevant for assessing both the strength of integration of these communities, and also the validity of the sampling method employed. The conclusion also discusses the implications of the findings for understanding artist communities.

I: RDS as a Method for Sampling Hard-to-Reach Populations

The attributes that make jazz musicians difficult to sample establish them as a "hidden population." These are populations with three characteristics. First, no sampling frame (i.e., list of population members) is available, so the size and boundaries of the population are unknown. Standard probability sampling methods cannot be employed unless prior formative research succeeds in creating a sampling frame. Second, the population is a small proportion of the general population, so general population-screening methods are impractical for assembling the sampling frame. For example, based on population estimates, in SF more than 350 individuals would have to be contacted to locate each jazz musician, and in NYC more than 550 would have to be contacted. Third, privacy concerns limit access by outsiders to the population. This may result from stigmatized behavior, as in the case of hidden populations that are relevant to public health, such as injection drug users; in the case of artist groups, it results from the tight but informal networks, which outsiders find hard to penetrate.

Great attention has been devoted recently to the problems involved in sampling hidden populations because of two events, the AIDS epidemic (Laumann et al. 1989) and decreases in the accuracy of the U.S. Census (Brown et al. 1999). Efforts to address both problems have focused attention on problems in sampling hidden populations (Semaan et al. 2002), particularly injection drug users, men who have sex with men, and the homeless. Though these groups differ profoundly in social and legal status from artist groups, methods developed for studying stigmatized hidden populations can be adapted to study nonstigmatized hidden populations. That is the premise of the study described in this article.

Three basic methods have dominated research on hidden populations. One is location sampling, which involves identifying venues where members of the population can be found, and then deploying interviewers. This is suitable for sampling geographically concentrated hidden populations, and it is most practical for large public venues. Although prominent jazz musicians perform in jazz clubs and festivals, those who are less prominent perform in smaller, more private gatherings, such as weddings. Because venue size is associated with professional recognition, sampling from large public venues would overrepresent the most prominent musicians and make it impossible to study the career path by which novice musicians ultimately grow into masters.

The second frequently employed method for sampling hidden populations involves institutional sampling. The problem is that the institution does the sampling. Some institutions, such as jails and prisons, tend to oversample the dispossessed. Other institutions, such as voluntary associations, tend to oversample the privileged. An example is the American Federation of Musicians, whose membership extends to only an estimated 22% of jazz musicians in NYC and 8% in SF. These union members have substantially higher incomes and levels of professional activity, so sampling only from this institution would produce a severely biased sample of jazz musicians.

The third approach to sampling hidden populations is link-tracing designs, also known as a chain-referral samples. This method is suitable for studying populations linked by a contact pattern. AIDS-prevention research has focused on contact patterns created by drug sharing and sexual contact (Heckathorn 1997). In the case of jazz musicians, contact patterns are created by playing together, through informal mentoring, through seeking opportunities to perform, and through bonds created by appreciation of the music. This sampling method thus takes advantage of the best information regarding who is a member of the jazz musician community—that possessed by the jazz musicians themselves.

A problem that has limited the utility of this method is ensuring that the sample is representative of the population from which it was drawn. In a now-classic article, Erickson (1979) argued that the sample begins with a bias because the choice of initial subjects in a hidden population cannot be random, and further biases of an unknown nature are added as the sample expands during subsequent referrals, or recruitment waves. For this reason, chain-referral samples have been considered a form of convenience sampling. Subsequent to Erickson's analysis, additional biases have been identified, so sources of bias in chain-referral samples now number four:

- the choice of initial subjects—the "seeds"—which cannot be selected randomly;
- those with large personal networks tend to be oversampled because more recruitment paths lead to them;
- some groups recruit more effectively than others, so the recruitment patterns of the best recruiters tend to be overrepresented in the sample; and
- groups with a stronger self-affiliation bias (i.e., homophily) tend to be oversampled because the recruitment chain becomes trapped in such groups before moving on to other groups.

Because of those problems, chain-referral samples have traditionally been seen as suitable only for pilot studies and formative research.

Nevertheless, there has been a resurgence of interest in chain-referral methods because of their unique ability to reach those who would be missed by other methods, including those who shun public gatherings and institutional affiliations (Thompson and Frank 2000; Frank and Snijders 1994;

Heckathorn 1997). Research on the "small-world problem" suggests that any two people in the country are connected by no more than six network links—the now-famous "six degrees of separation." The implication is that everyone can be reached by a maximally expansive chain-referral sample after only a handful of waves.

Respondent-driven sampling was developed to overcome the biases traditionally associated with chain-referral sampling (Heckathorn 1997, 2002a). RDS has served as the recruitment mechanism for an HIV-prevention intervention (Heckathorn et al. 1999) that is based on a formal theory of collective action (Heckathorn 1990, 1993, 1996, 2002b). The design principle is straightforward. If the biases associated with chain-referral methods are understood, it is possible to redesign the sampling process to eliminate those biases that are not inherent in the method, and through constructing a statistical theory of the sampling process to quantify and control those that are inherent in the method, through poststratification. Therefore, RDS includes both a specific method for structuring the chain-referral process to reduce one set of biases, and analytic procedures to weight the sample, poststratifying it to compensate for those that remain. In this way, RDS yields unbiased population estimates. More specifically, the population estimate has been shown to be the ratio of Horvitz-Thompson estimators (Salganik and Heckathorn 2003), a form of estimator that is known to be unbiased. The ratio of two such estimators is asymptotically unbiased, which means that bias is on the order of 1/(Sample Size), so bias is negligible even in modest samples.

FIGURE 1 ABOUT HERE

Figure 1 depicts the largest referral chain from the NYC study, including the race, ethnicity, gender, and primary instrument for each respondent. It began (wave 0) with a black female bass player who recruited (wave 1) a white female keyboard player, a white female singer, and a female alto saxophone player of "other" race or ethnicity. Over the course of 10 waves, the sample expanded from the single seed to include more than 100 respondents. An examination of recruitment patterns shows that, as emphasized by Erickson, referrals reflect respondent characteristics. For example, inspection of the figure reveals several recruitment patterns, including a tendency for women to recruit women, whites to recruit whites, and singers to recruit singers. In such cases, recruitment reflects homophily (McPherson and Smith-Lovin 1987).

TABLE I ABOUT HERE

Table 1A depicts recruitment by gender in all the referral chains in the NYC sample. Males recruited, on average, 71% other males even though males made up an estimated 58% of the population, and females recruited 61% other females even though they made up 42% of the population. The pattern for race and ethnicity in NYC is similar: whites recruited, on average, 69% other whites even though whites made up an estimated 58% of the population, and blacks recruited 46% blacks even though they made up only 33% of the population.

However, an opposite pattern is possible. Consider, for example, recruitment by gender in SF (see Table IB). Females recruited 9% females even though they made up an estimated 14% of the population, so out-group recruitment was favored. Similarly, non-Hispanic whites in SF recruited other whites 55% of the time even though they made up an estimated 64% of the population, so again out-group recruitment was favored. Thus, recruitment can reflect either positive or negative homophily.

As we show in the following subsections, RDS solves these problems with chain-referral sampling, through restructuring the sampling process to ensure reliability, and post-stratification to produce unbiased population estimates.

Reliability of Respondent-Driven Sampling

Can chain-referral sampling be a reliable method even though seeds from a hidden population cannot be selected randomly and referrals are biased by homophily? It might seem that homophily would make chain-referral samples irrevocably biased—that a group overrepresented among the seeds with which recruitment began would then become increasingly prominent as the sample expanded, such that the initial biases might seem to be multiplied. The manner in which homophily affects recruitment as the chain-referral sample expands from wave to wave can be identified by modeling the process as a form of stochastic model known as a Markov chain (Kemeny and Snell 1960). A Markov chain consists of a set of two or more states (e.g., subject characteristics, such as gender or ethnicity), and transition probabilities from state to state (i.e., probabilities that a subject with a given set of characteristics will recruit a subject with each other possible set of characteristics).

FIGURE 2 ABOUT HERE

For an illustration of a Markov chain, see Figure 2, which depicts Table IC's data on recruitment by race and ethnicity in NYC. The four states correspond to the recruiter's race and ethnicity (i.e., non-Hispanic white, non-Hispanic black, Hispanic, or other), and arrows depict the transition probabilities across states. Recruitment is a stochastic process that can be visualized as a point whose location corresponds to the state of the most recent recruit; cross-state recruitment then moves the point to a different state by following the straight arrows (i.e., the recruit differs from the recruiter), and within-state recruitment keeps the point at the same location by following a looped arrow (i.e., the recruit is similar to the recruiter).

FIGURE 3 ABOUT HERE

An advantage of modeling the recruitment process using Markov chain theory is that a large body of mathematical theory becomes applicable to understanding the sampling process. Modeling the recruitment process as a Markov chain produces a surprising result—that biases introduced by the selection of initial respondents are not multiplied but instead are progressively *weakened* with each recruitment wave and ultimately eliminated (Heckathorn 1997). The manner in which this occurs is illustrated in Figure 3, which depicts the results of two simulations showing how the composition of each wave would have changed had recruitment begun from either one or more black musicians (Figure 3A) or one or more white musicians (Figure 3B), based on projections from Figure 2's recruitment patterns. The vertical axis represents the percentage of musicians of each type, and the horizontal axis represents recruitment waves, where wave 0 refers to the seed or seeds, which in this simulation were assumed to be ethnically homogeneous. Wave 1 refers to the seeds' recruits; wave 2 refers to the recruits' recruits, and so forth. Had recruitment begun with only black seeds, the percentage of blacks in each wave decreases from the initial value of 100%, to 46% in wave 1, 34% in wave 2, eventually stabilizing at 30%. Once this stable point, or *equilibrium,* is reached, the composition of that and each additional wave is 61% non-Hispanic whites, 30% non-Hispanic blacks, 2% Hispanics, and 6% others.

In contrast, when the simulated recruitment begins with only white seeds (Figure 3B), the percentage of blacks in each wave increases from the initial value of 0% to 23% in wave 1, 29% in wave 2, and stabilizes at 30% in waves 3 and subsequent waves. Note that after equilibrium is attained, the composition by wave in Figure 3B is the same as in Figure 3A. This convergence reflects an important characteristic of RDS: if sampling is allowed to proceed through a minimum number of waves, it will attain an equilibrium that is *independent of the characteristics of the respondents from which sampling began*. This conclusion derives from the law of large numbers for regular Markov chains (Kemeny and Snell 1960), which has been shown to apply to chain-referral samples (Heckathorn 1997). Thus, it does not matter whether all seeds were drawn from the same group or from any mix of groups—the ultimate composition of the sample will be the same.

A further useful conclusion from Markov chain theory is that this equilibrium is reached rapidly, at a geometric rather than an arithmetic rate (Kemeny and Snell 1960). The implication for chain-referral samples is that the number of waves required to eliminate the bias introduced by the choice of seeds is modest—in previous applications of RDS, usually no more than three to six (Heckathorn 2002a). Whatever bias was introduced by the selection of initial respondents is eliminated if sampling is continued through enough waves. Therefore, the first design principle of RDS is an effort to produce long referral chains.

To ensure that referral chains will be lengthy, respondents in RDS receive modest financial rewards for their recruiting efforts. Rewards are kept modest—$15 for each recruit—so that that recruitment does not become coercive. Although every application of RDS has asked respondents whether their recruiter used excessive pressure, none has been reported (Heckathorn et al. 1999). A quota on recruitment rights—generally a limit of three recruits per respondent per six months in public health applications, partially relaxed in this study—also helps lengthen referral chains. Quotas on recruitment are implemented using a coupon system, in which potential recruiters are given dollar-bill-sized coupons that they give to their recruits. The coupon includes the study name, a phone number to call to make an appointment for an interview, and a map to the interview site. The coupon also includes a serial number with which to determine who recruited whom. This information is used both to calculate rewards for recruiters, and as part of the link-tracing design, to document connections among respondents.

After the desired number of subjects is recruited, computations can be performed (see Heckathorn 1997:186) to confirm that the composition of the sample converged with the equilibrium sample composition. For example, in a two-category system with groups A and B, where S_{ab} is the transition probability from A to B (i.e., when As recruit the proportion of Bs who are selected), S_{ba} is the transition probability from B to A. The proportion of As in equilibrium E_a is given by the expression (Heckathorn 1997):

$$(1) \qquad E_a = \frac{S_{ba}}{S_{ba} + S_{ab}}$$

For example, for recruitment by gender in NYC (Table IA), where As are males and Bs are females, the equilibrium proportion of males is .3846/(.3846+.2865) = .5731, which differs from the proportion of males in the sample, .6255, by 5%, only slightly more than the standard error. In SF the convergence is closer (Table IB): the equilibrium proportion of males is .8595, and the proportion in the sample is .8597, for a difference of less than 1%. For recruitment by race and ethnicity, the equilibrium differs from the sample composition by no more than about 2% (Table IIC and IID). This suggests that RDS can be a valid sampling process, in the sense that sample composition is not affected by the nonrandom selection of seeds. However, it leaves open the question of validity.

Validity of Respondent-Driven Sampling

Reliability is a necessary but not sufficient condition for a sampling method to be useful. A useful method must also be valid—that is, nonbiased. *Can a chain-referral method be valid despite the effect of variations in network sizes, recruitment effectiveness, and strength of self-affiliation?* Given these sources of bias, the sample composition may not reflect the population from which the sample was drawn, so a valid population estimator must take all three sources of potential bias into account. A paper extending the RDS sampling method (Heckathorn 2002a) showed how to calculate such a population estimator through a two-step process. First, data from the sample are used to calculate network indicators; then the network indicators are used to calculate population estimators, including the proportional size of subgroups in the population and the proportion of ties that cross rather than remain within groups.

The RDS population estimator is based on the *reciprocity model*. It draws on two findings regarding recruitment dynamics. First, respondents recruit overwhelmingly from those with whom they have a preexisting relationship—generally acquaintances, friends, or those closer than friends, such as relatives (Heckathorn 2002a). Second, in aggregate, respondents recruit as though they are sampling randomly from their personal networks, so recruitment patterns provide information on network composition (Heckathorn et al. 2002). Ties within a personal network are reciprocal because a link from any individual A to B implies that a link also exists from B to A. Hence for two such groups, the number of links from A to B (T_{ab}) will equal the number from B to A (T_{ba}). That is,

(2) $T_{ab} = T_{ba}$

Furthermore, the number of ties from A to B is the product of four terms: the population size for the system (PS), the proportional size of the group (P_a), the group's mean network size (N_a), and the proportion of ties from that go from A to B (S_{ab}). That is,

(3) $T_{ab} = PS\ P_a N_a S_{ab}$

Hence, for two groups, A and B,

(4) $PS\ P_a N_a S_{ab} = PS\ P_b N_b S_{ba}$

When group size is expressed as a proportion, so 1-P_a can be substituted for P_b, this expression can be solved for group A's proportional size, P_a as follows:

(5) $$P_a = \frac{S_{ba} N_b}{S_{ba} N_b + S_{ab} N_a}$$

This is the estimate of group size based on the reciprocity model, and it provides the means for controlling for three sources of bias—those due to differentials in network size, recruitment effort, and homophily (Heckathorn 2002).

For example, consider the case of gender in NYC. From Table 1A, where males are group A and females are group B, males recruit on average S_{ab} = .2865 females, and females recruited on average S_{ba} = .3846 males, so females had greater cross-group recruitment. Furthermore, the mean network sizes were 222.99 for males, and 232.23 for females, so females had larger networks. Substituting these values into equation 5 yields the estimated proportion of males in the population:

(6) $$\frac{.3846 \quad 232.23}{.3846 \quad 232.23 \quad + \quad .2865 \quad 222.99} \quad = \quad .583$$

Note that the estimated proportion of males, .583, is less than that in the sample, .6255. This population estimate balances the effects of several potential sources of bias. Males' slightly smaller networks would lead to undersampling. But males were more effective recruiters, carrying out 73% (178/243) of the recruitments even though they made up only 63% (152/243) of the sample; thus the male recruitment pattern would lead to oversampling. Finally (see below), males' slightly weaker homophily would again lead to undersampling. Thus the three potential sources of bias operate in inconsistent directions, and the reciprocity model produces a population estimate that balances these effects. That the population estimate exceeds the sample proportion suggests that on balance males were oversampled, so the second of the three factors was more decisive.

The RDS population estimate generalizes to systems with more than two groups. For example, the solution for a system with four groups requires solving a system of seven linear equations:

$$\begin{aligned} 1 &= P_a + P_b + P_c + P_d \\ P_a N_a S_{ab} &= P_b N_b S_{ba} \\ P_a N_a S_{ac} &= P_c N_c S_{ca} \\ P_a N_a S_{ad} &= P_d N_d S_{da} \\ P_b N_b S_{bc} &= P_c N_c S_{cb} \\ P_b N_b S_{bd} &= P_d N_d S_{db} \\ P_c N_c S_{cd} &= P_d N_d S_{dc} \end{aligned} \tag{7}$$

This equation system is overdetermined because there are more equations than unknowns. Given that it may not be possible to find a set of unknowns that satisfies all equations, the solution should satisfy all equations as closely as possible, where "close" is defined in the least squares sense. A procedure for deriving such a solution is Gauss-Jordan elimination. This is the procedure used to derive the population estimates for four-category systems (see Table IC and ID).

The compositions of the NYC and SF jazz communities, as revealed in Table I, are quite different. In NYC, 42% of jazz musicians are female, compared with only 14% in SF. In contrast, composition by race is similar at least in an ordinal sense, with 58% versus 64% non-Hispanic whites, 33% versus 23% non-Hispanic blacks, 7% versus 8% others, and 1% versus 4% Hispanics in NYC and SF, respectively.

FIGURE 4 ABOUT HERE

Figure 4 depicts the distribution of income from performing music, gross individual income, and gross household income for jazz musicians in NYC and SF. The results are consistent with studies of other artist groups, in which income distributions are positively skewed with a low mean. Like other artists, the typical jazz musician earns rather little, but some have very large salaries. A notable difference between the two income distributions is that income from music is greater in NYC than in SF. This may reflect greater performance opportunities in NYC as well as a stronger musicians' union. In both areas, income limitations impede access to essential services; for example, in NYC, 45% of jazz musicians lack health insurance, and in SF the figure is 34%.

FIGURE 5 ABOUT HERE

Figure 5 depicts the distribution of musical styles for both cities, where the list of styles was identified by a focus group of jazz musicians organized by the Research Center for Arts and Culture at Columbia University's Teachers College. Two styles were excluded from our analysis, "regional" and "other," because these both refer to heterogeneous categories of styles. The remaining 19 styles are ordered based on popularity for each city.

Choices of styles are quite divergent. In SF the two most popular styles are bop, a style of jazz that emerged after World War II in the works of Dizzy Gillespie, Charlie Parker, Max Roach, Kenny Clarke, Bud Powell, and Thelonious Monk; and hard bop, a style that emerged in the 1950s and early 1960s with the works of Art Blakey, Horace Silver, Hank Mobley, Wayne Shorter, Johnny Griffin and Branford Marsalis, Donald Byrd, Woody Shaw, Wynton Marsalis, and Lee Morgan. However, in NYC bop and hard bop are ranked only fifth and ninth. More popular styles include traditional Dixieland jazz, which began in New Orleans and whose development was complete in the 1910s in NYC; free jazz, or "New Thing" as it was called when it emerged in the late 1950s from saxophonist Ornette Coleman and pianist Cecil Taylor; and avant-garde, which emerged in the works of Lennie Tristano, Jimmy Giuffre, Gunther Schuller, Paul Bley, Andrew Hill, Anthony Braxton, Sam Rivers, Sunny Murray, and Andrew Cyrille.

In NYC, only three styles are performed by more than 25% of the respondents. As described by one NYC musician, "In jazz we have all sorts of different interpretations, approaches and styles, but the mainstream [media] show only one of many versions of the music. So what you see in the dominant media is only a fraction of what is really there." In contrast, in SF fully 11 styles are performed by a quarter of the respondents. This suggests that musicians in NYC are more specialized: they perform on average in only 2.3 styles, compared with 7.09 styles in SF.

II: Using RDS to Study the Network Structure of Jazz Musician Communities

Studies of affiliation patterns have focused on either of two measures of affiliation (McPherson, Smith-Lovin, and Cook 2001). One focuses on crosscutting ties—that is, the number of social network linkages to out-group members relative to the total number of linkages to in-group and out-group members. This is the approach embraced by Blau (1977, 1994) and others operating in the macrostructural tradition. This is also the focus of Putnam's (2000) studies of social capital, in which within-group ties are viewed as the source of social cohesion (i.e., *bonding social capital*), and cross-group ties promote cross-group integration (i.e., *bridging social capital*).

An issue faced by this approach is that cross-group ties reflect not merely social cohesion or solidarity but also relative group size. Consider, for example, a term devoid of social meaning: whether one was born in an odd- or an even-numbered month. Given that about the same number of people are born each month, and that birth month is socially irrelevant, one can expect that the odd and even groups will have about 50% in-group ties and 50% out-group ties. In Putnam's terms, bridging and bonding social capital are equal. If one then considers only those born in December, the results are different, with 1/12th in-group ties and 11/12ths out-group ties. Here, bridging social capital is greater than bonding social capital, so the December group appears atomized. In contrast, the non-December group has mostly in-group ties, so it could appear xenophobic. Obviously, such conclusions are nonsense because the characteristic is devoid of social meaning, and the number of in-group versus out-group ties depends merely on relative group size. In recognition of this issue, Blau (1994) and his associates have properly focused much of their attention on the association between group size and crosscutting ties. For example, the principal hypothesis of Blau's macrostructural analysis is that as group size decreases, the proportion of crosscutting ties will increase.

Formal network analysts have adopted a different approach to analyzing affiliation patterns (Fararo and Sunshine 1964; Rapoport 1979; Fararo and Skvoretz 1984), using as a baseline the notion of a randomly connected network. Structure is thereby seen as existing within a network only when affiliation patterns depart from those that would be produced by random mixing. The result is a means for measuring the strength of in-group affiliation that compensates for the effects of group size. Fararo and Sunshine (1964) defined one such index. Its value is 1 if all ties are to the in-group and 0 if ties are formed by random mixing; intermediate values have a straightforward interpretation. For example, a value of .55 indicates that the actor forms ties as though 55% of the time a tie is formed to the in-group, and 45% of the time a tie is formed irrespective of group membership. As thus defined, the index provides a measure of the strength of in-group affiliation bias that has the benefit of compensating for the effects of group size. Subsequently, this index was generalized to accommodate affiliation based on complementarity, in which ties are preferentially formed not to the in-group but to the out-group (Heckathorn 2002a). The range of index values was thereby extended to negative values, where an index value of –1 indicates that all ties are formed to the out-group, and intermediate values have an interpretation consistent with that for positive values. For example, a value of –.33 indicates that networks are structured as though 33% of the time a link is formed to out-group members, and the other 67% of the time a link is formed independent of group identity.

The equations for computing homophily (Heckathorn 2002a) are as follows, where H_a is the homophily of group A, P_a is the proportional size of the group, S_{aa} is the proportion of A's ties that are in-group ties, and homophily is positive if $S_{aa} > P_a$ and negative if $S_{aa} < P_a$:

$$
\begin{aligned}
H_a &= \frac{P_a - S_{aa}}{P_a - 1} \quad if \quad S_{aa} \geq P_a \\
H_a &= \frac{S_{aa}}{P_a} - 1 \quad if \quad S_{aa} \prec P_a
\end{aligned}
\tag{8}
$$

This index of network clustering is employed in our study for two reasons. First, it has been shown that if all groups have equal homophily, the RDS population estimate and the equilibrium sample composition will be equal (for the proof, see Heckathorn 1997). Therefore, this measure of homophily is the index that specifies the extent to which network structure can bias a link-tracing sampling process. Second, RDS provides estimates of both terms required to calculate the homophily index—the proportions of in-group and out-group ties, and the estimates of size for each group in the sample—so homophily can be calculated directly from RDS data. That is, substituting equation 5 into equation 8 above and substituting ($1-S_{ab}$) for S_{aa} and ($1-S_{ba}$) for S_{bb} yield the network-indicator-based estimate of homophily:

$$(9) \quad \begin{aligned} H_a &= \frac{N_a - S_{ba}N_b - S_{ab}N_a}{N_a} \qquad if\, S_{aa} \geq P_a \\ H_a &= S_{ab}\,\frac{N_a - S_{ba}N_b - S_{ab}N_a}{S_{ba}N_b} \quad if\, S_{aa} < P_a \end{aligned}$$

For example (see Table IA), in NYC male homophily is (223 – .385 * 232 – .287 * 223)/(223) = .313; and similarly, homophily for females is H_b = .34. Such substantial homophily for both males and females suggests the presence of relatively independent male and female jazz musician groups. This is consistent with the observation that in New York women have formed separate organizations, like International Women in Jazz, specifically to network and help each other. In contrast, in SF, males have a neutral homophily of –1%, and female jazz musicians have a substantial negative homophily of –34%, so females interact indirectly via males. This suggests that female jazz musicians have a substantially lower status in SF than in NYC, occupying a network position that Burt (1976) terms "sycophant." The difference may reflect differences in population sizes. Whereas in SF females make up only 14% of the musician community, in NYC they are a very substantial 42%. NYC may therefore have enough female jazz musicians to trigger the emergence of relatively independent female jazz musician groups. Scale may also be relevant, because the estimated number of jazz musicians in NYC is nearly double that in SF.

Homophily and Network Structure

Social structure has a precise meaning when defined quantitatively (Blau 1977; Rapoport 1979). A system is said to lack structure if social relationships are formed randomly. Thus, homophily and heterophily are the elements out of which social structures are built. As conceived in this way, four distinct structures can be identified:

- Cohesive subgroups, in which ties are formed among persons who are similar: each group has positive homophily. This fits, for example, affiliation by gender in NYC.
- Core-periphery structures, in which those of highest status affiliate differentially with one another, and persons of lower status interact primarily through those of higher status: the lower-status group has negative homophily and the higher group has positive homophily.

- Bipartite systems. in which interaction is based not on similarity but on complementarity, as in sexual contacts among heterosexuals: both groups have negative homophily because crosscutting ties are favored over in-group ties.
- Unstructured systems, in which affiliation patterns are consistent with random mixing: groups have zero homophily.

Homophily by Race and Ethnicity

Among NYC jazz musicians, whites, blacks, and “others” all exhibited some degree of positive or negative homophily, because their in-group affiliations, as reflected in their recruitment patterns, differed from their proportional group sizes (see Table IC). Homophily was strongest among whites (27%), weaker among blacks (20%), and negative among “others” (–17%). In contrast, in SF homophily was strongest among blacks (26%), and “others” (7%), and negative among whites (–13%).

TABLE II ABOUT HERE

Homophily is an index only for in-group affiliation. Yet degrees of affiliation can also vary across groups—that is, some groups may be closer to others. To provide a means for quantifying such differences, the homophily index was generalized as an index indicating the strength of affiliation not only toward one’s own group but to other groups as well (Heckathorn 2002a). Affiliations by race in NYC and SF are depicted in Table II. Note that each group’s self-affiliation is homophily, as reported in Table I. Affiliations with other groups are highly variable. For example, in NYC whites and blacks are most socially distant. Whites have a negative affiliation toward blacks of –25%, indicating that networks are formed as though 25% of the time a tie is formed to a nonblack, and the other 75% of the time, a tie is formed independent of race. Similarly, the affiliation from blacks to whites is –32%. Overall, whites orient positively only toward other whites, blacks orient toward other blacks and are near neutral (1%) toward those in the “other” group. “Others” have weak affiliations that are mildly positive toward whites (12%), and mildly negative toward blacks (–10%).

Those findings are in sharp contrast to the pattern in SF, where whites have near- neutral affiliations toward all other groups, with affiliation indices ranging from 1% to only 8%. Given that this group also has negative self-affiliation, it illustrates well a case in which ethnic boundaries have dissolved. In contrast, SF blacks exhibit substantial negative affiliation toward whites (–35%), with neutral affiliations toward the “other” group (2%). Finally, “others” have neutral affiliations toward everyone else. In both communities the tiny Hispanic group had near zero homophily, except when zero entries in the recruitment matrix produced values of –1.

This comparison of affiliation by race in NYC and SF leads to several conclusions. Simmel (1955) suggested that groups that strengthen their boundaries will elicit similar responses in other groups, so the strength of group boundaries will be positively associated. In contrast, these findings suggest that boundaries can be highly variable in strength, including the case in which some groups have dissolved their boundaries while other groups maintain them. For example, note that blacks are more homophilous in SF, where whites have dissolved their boundaries, than in NYC, where whites have maintained their boundaries.

What the data seem to indicate is that, among jazz musicians, boundaries of race are crossed more in San Francisco than in New York. What is also notable is that the boundaries that are kept or dissolved are different in each city and for different racial groups. This has implications for arts policies and programs in metro areas as well as for artists, especially after a decade when “multiculturalism” was a central issue. For students who choose to enter the professional jazz scene, the black networks are distinctly separate from the white ones. This was confirmed by one of our interviewees, a NYC black jazz musician who was advised by his companion not to tell the interviewer about another “scene” for developing jazz, unknown to the media, outside the mainstream, because those who attend keep it secret, private, mostly black, not open to just anyone.

To more precisely compare the strength of racial and ethnic boundaries in the two areas, it is useful to introduce a new measure, termed the boundary strength index (BSI), defined as the proportion of network ties that are governed by homophily. Note that this is not the same as the proportion of in-group ties, since even random mixing would produce in-group ties in proportion to each group’s size. The homophily index provides a measure of the proportion of ties governed by homophily for each group (in the NYC analysis, .2701 for whites and .1961 for blacks). The BSI index combines these proportions in a way that weights for differences in both the sizes of groups and their network sizes. Where P_i is the size of group i, N_i is its mean network size, and IGB_i its in-group bias (homophily if that

$$(10) \quad BSI = \frac{\sum_{i=1}^{n} P_i N_i IGB_i}{\sum_{i=1}^{n} P_i N_i} \quad where\ if\ H_i \succ 0, IGB_i = H_i\ else\ IGB_i = 0$$

term is positive, and zero otherwise), the expression for BSI is

In this expression, the denominator is a population-proportion-based measure of the total number of ties in the system (i.e., the sum across groups of each group's size multiplied by its mean network size), and the numerator is the number of ties formed based on in-group affiliation (i.e., the sum across groups of the number of ties multiplied by the proportion affected by positive homophily). Therefore, the expression indicates the proportion of ties formed based on in-group affiliation. This index provides a means for quantifying the extent to which racial and ethnic boundaries are stronger in NYC, where BSI = .2313, than in SF, where BSI = .0964. Thus, whereas in NYC almost one-quarter of ties are affected by racial and ethnic boundaries, in SF the figure is slightly less than one-tenth. This index, it should be noted, differs in concept from the well-known measures of network clustering based on triad closure (e.g., Fararo and Sunshine 1964; Watts 1999), which are unaffected by whether closure occurs within or across group boundaries. In contrast, the boundary strength index indicates the degree to which ties form within rather than across groups, and thereby indicates the extent to which group identity shapes the creation of ties.

Homophily by Performance Style

Analyzing homophily by musical style (see Figure 6) provides a means for determining the extent to which style affects affiliation patterns. A previous study of a jazz musician community suggests that this factor may be strong. In his analysis of jazz in the Netherlands, IJdens (1999) found the musicians organized into three mutually antagonistic clusters. Those who played contemporary improvisational jazz viewed themselves as leading in the creation of new musical forms; in their view, those who played traditional jazz were recycling old music for audiences incapable of appreciating contemporary music, and those who played fusion were sugar-coating jazz by blending it with rock and other forms to make it palatable to unsophisticated audiences. Those who played traditional jazz saw themselves as upholding a venerable tradition; in their view, those who played contemporary improvisational music were elitists incapable of attracting large audiences, and they shared the improvisationalists' view of those who played fusion. Finally, those who played fusion shared the traditionalists' and the improvisationalists' views of one another.

FIGURE 6 ABOUT HERE

Figure 6 depicts the analysis of homophily by musical style. The analysis of homophily for each style is based on dividing the respondents into those who report playing in the style, versus all other respondents. One style, ragtime/stride piano, was excluded because too few respondents played in this style for homophily to be computed. Two observations are notable. First, in most cases, those who play in a style are more homophilous than those who do not play in that style, indicating that each style provides a basis for affiliation. In contrast, the "other styles" category is heterogeneous because it contains musicians who play in all other styles and have little basis for affiliation. Second, compared with what would have been expected from the Dutch study, musical style is a less powerful determinant of affiliation, peaking in NYC at only 25% for the mainstream style, and peaking in SF at 38% for the bop style.

Given the close association among certain styles, such as free jazz and avant-garde, it might seem that affiliation would be most affected not by individual styles but by clusters of styles. Factor analysis provides a means for testing this possibility (see Table III). The results confirm that U.S. musicians are generalists: the component that explains by far the most variance is a generalist style, positively loaded for all styles, a result that is unaffected by rotation. In the NYC analysis, the generalist component is most heavily loaded for rhythm-and-blues, blues, and funk. In contrast, the second component is oriented toward dance-oriented styles, bop, hard bop, and swing. The third is oriented toward ragtime and scat; the fourth toward avant-garde and free jazz; the fifth toward mainstream; and finally the sixth toward traditional. Therefore, each component identifies a distinct cluster of compatible styles.

In SF, the analysis identified four clusters of styles. The first, the generalist component, is most heavily loaded for hard bop, bop, free jazz, and cool; the second is oriented toward acid jazz, fusion, world music, and funk; the third toward boogie-woogie and ragtime; and finally the fourth toward traditional. Similarities between these clusters of styles include associations between avant-garde and free jazz, and between bop and hard bop. This affiliation to the generalist cluster of styles may be directly tied to the marketplace, where jazz musicians must be flexible to earn money from their music. Resources to encourage both the growth and depth of networks and the professional contacts that help sustain them should be a concern for arts policy and funding agencies.

FIGURE 7 ABOUT HERE

Figure 7 shows homophily based on whether respondents have a positive or negative factor score for each set of styles. The first component is labeled “generalist” and other components are named for the single most heavily loaded style. In NYC homophily peaks at 28% for those who are positive toward the mainstream component, followed by 23% for those who are negative toward avant-garde. In SF homophily peaks at 29% for those who are positive toward the generalist component, followed by 21% for those positive toward the traditional component. Thus, though some styles are more compatible than others, as indicated by the patterns revealed in Table III’s analysis, affiliation is more powerfully governed by other factors.

FIGURE 8 ABOUT HERE

If affiliations were based solely on playing together, one would expect a bipartite structure for those who play instruments that seldom occur together in the same ensemble, such as drums, bass, and keyboard; and one would expect cohesive subgroups for those whose instruments frequently occur together, such as singers. Of course, such a clear-cut pattern cannot be expected because playing in the same group is only one basis for affiliation. Nonetheless, a rough approximation of this pattern can be discerned in the analysis of homophily by primary instrument. Only the five most common primary instruments are listed. In both cities, drummers, bassists, and keyboard players are mildly or strongly heterophilous, which is consistent with a bipartite structure. However, unlike the pure bipartite structure, nondrummers, nonbass players, and nonkeyboard players are not heterophilous, perhaps because of these groups’ heterogeneity. For example, nondrummers include those who play the 24 instruments other than drums, and this provides conflicting bases for affiliation. It is notable that both singers and nonsingers are homophilous in NYC, perhaps reflecting a boundary between vocal and instrumental ensembles. The mild heterophily of singers and the homophily of nonsingers in SF is harder to interpret. It may be an artifact of its link with gender, because in SF voice was selected as the primary instrument by 38% of female respondents but only 4% of males, and 67% of those choosing voice were female. A multivariate analysis to clarify this issue would exceed the scope of this paper, but what is clear is that primary instrument can provide a substantial basis of affiliation.

III: Breakpoint Analysis: Quantifying the Effects of Continuous Variables on Network Structure

Many continuous variables, including social status, income, and age, are known to affect affiliation patterns. Here we introduce a new means for studying how network structure changes as a function of the value of a continuous variable that is applicable to RDS data. The procedure involves a sequence of analyses:

(1) A lower breakpoint for the value of a continuous analysis is defined, thereby dividing respondents into a lower and a higher group. For example, for an analysis of network size in NYC, if the breakpoint is set at 50, there are 24 respondents in the group with self-reported personal network sizes of 50 or fewer and 180 in the group with network sizes of more than 50. Of course, the lower breakpoint should be high enough to ensure the lower group is not empty.

(2) The mean network sizes for the lower and higher groups are calculated. For example, the mean network sizes are 36.875 and 252.123 for the lower and higher groups, respectively.

(3) The proportion of out-group recruitment for the lower and higher groups is calculated. For example, out-group recruitment is .875 (=3/24) and .083 (=15/180) for the lower and higher groups, respectively.

(4) Homophily is calculated using equation 9 for both groups based on mean network sizes and proportional out-group recruitment. For example, homophily for the lower and higher groups is –.683 and .789, respectively.

(5) The breakpoint is incremented, and steps 1 to 4 are repeated. This continues until too few respondents remain in the higher group for a meaningful analysis, as indicated by the absence of either cross-group or within-group recruitment for either the higher or the lower group.

Homophily and Personal Network Size

FIGURE 9 ABOUT HERE

Figure 9 depicts the breakpoint analysis of homophily by network size for both NYC jazz musicians (A) and SF jazz musicians (B). Network sizes were substantially higher in NYC than in SF, so a higher set of breakpoints was employed for the NYC analysis (range = 50 – 475) than in the SF analysis (range = 10 – 190). Note that in both cases, when the breakpoint is low, the lower group is strongly heterophilous and the higher group is strongly homophilous. This very strong core-periphery structure indicates that the lower group shuns its own members and both groups gravitate toward the higher group. For example, NYC musicians with networks of 50 or fewer form networks as though 68% of the time they form a tie to the higher group consisting of musicians with a network greater than 50, and the other 32% of the time they form a tie irrespective of network size. In contrast, musicians with networks greater than 50 gravitate toward one another, forming networks as though 79% of the time they form a tie to a musician with a network greater than 50, and the other 21% of the time they form a tie irrespective of network size.

As the breakpoint for network size sweeps upward, the core-periphery structure weakens, homophily converges toward zero, and the system becomes unstructured. For example, NYC musicians with networks greater than 475 and those with networks of 475 or fewer have homophily within 10% of the zero point. Thus, above a threshold, differences in network sizes lose significance.

The pattern shown in Figure 8 is characteristic of a stratified system in which sociometric stars have entourages of lesser stars, these lesser stars have entourages of still lesser stars as well as upward ties to more prominent stars, and this structure is repeated downward to musicians of ever-lower status. A San Francisco jazz musician confirms this by saying, “The tier structure appears to be in full force with the folks who are visibly ‘out there’ and getting press at the higher rung.” In New York, some of the minor stars have circles that interconnect with other constellations, which serve as connections to larger stars. The remarkable similarity of the pattern displayed in the two figures suggests that despite the difference in mean network size between the two cities, jazz musicians are stratified by relative network size in a similar manner.

The contrast between homophily by network size and by race or gender is also notable. As measured by the magnitude of homophily, network size is a more important determinant of affiliation. This suggests that jazz musicians form an integrated community, in which prominence in the community, as measured by network size, is more significant in determining affiliation than demographic factors, such as race or gender.

Homophily and Age

FIGURE 10 ABOUT HERE

Figure 10 depicts a breakpoint analysis of homophily by age in NYC and SF. The patterns are quite different from the structure for network size. Consider first Figure 9A, which depicts homophily by age in NYC. Observe that except for the highest breakpoints, homophily is positive for both the younger and the older group. Therefore, this is a form of cohesive subgroup structure. However, homophilous affiliation is based not on a nominal characteristic, such as ethnicity, but on similarity with respect to a continuous variable—in this case, similar age. Therefore, this is a cohort structure. That affiliation by age for NYC jazz musicians peaks at 58% for the older group and 31% for the younger group suggests that age is a substantial but not overwhelmingly important determinant of affiliation.

The older group exhibits consistently higher homophily in both NYC and SF. This reflects the higher status of older persons in the jazz musician community. The difference is less pronounced in NYC because the higher and lower groups both have more similar levels of positive homophily. The difference in status is much greater in SF. Indeed, in SF the younger group exhibits consistently negative homophily, so this structure can best be characterized not as a cohort structure but as a core-periphery structure. The implication is that younger musicians, like female musicians, have substantially lower status in SF than in NYC. As one SF musician stated, "Young jazz musicians are constantly factionalized by other musicians and the industry." Most SF venues are in restaurants, not jazz clubs that feature the music as a primary activity, and younger musicians may have difficulty competing against older musicians for these scarce jobs.

A final feature of Figure 9 warrants comment. In both NYC and SF, homophily declines as the breakpoint increases. This suggests that the significance of differences in age decreases over the life course. For example, the significance of a difference of 5 years in age is greater for persons who are 25 years old than for those who are 55. This principle is well known, but the fact that it can be derived from this analysis of homophily by age helps confirm the validity of this form of analysis.

Homophily and Sources of Income

FIGURE 11 ABOUT HERE

Figure 11 depicts homophily by all three types of income for both cities. The categories differ slightly because, for example, too few NYC respondents had personal and household incomes of less than $500, and too few SF respondents had music incomes of more than $50,001.

The difference between NYC and SF is substantial. In NYC, income from music has a stronger effect on affiliation than do other forms of income, as indicated by the higher-income-from-music group's having the greatest homophily for all but the highest income category. The lower-income-from-music group has minimal homophily. This structure contains features of both a core-periphery structure, because only the higher group has positive homophily, and a cohort structure, because homophily is nonnegative for the lower group. This reflects the greater status in the jazz musician community of those who perform for money.

Income from music exhibits a more clear-cut structure in SF. Those who earn more than $5,000 from music are substantially homophilous, 41%, while those who earn that amount or less are substantially heterophilous, –46%. As one SF musician said, "The musicians who are doing financially well generally hang together and form exclusive cliques." Income other than that from music, including personal and household income, in general has smaller effects on homophily in both cities. This reflects the strength of the jazz musician community's focus on commercial performance.

Homophily by Participation in Ensembles

FIGURE 12 ABOUT HERE

When jazz musicians perform, it is almost always as part of an ensemble, and many musicians perform with multiple groups. Figure 12 depicts the effects on homophily of the number of groups with which the respondent performs. The patterns for SF (12B) correspond to a core-periphery structure, with strong positive homophily for those who perform with one group or more, and substantial negative homophily for those who perform with no groups. In this respect, it closely resembles the analysis of homophily by income from music in SF (Figure 11B), suggesting that performing with greater numbers of groups increases one's status, and that the greatest increment in status comes from playing with one rather than zero groups. The patterns for NYC also resemble its pattern for homophily by income for music, with a rather weak cohort structure. Both the higher and the lower groups have positive homophily, but the magnitude is lower, peaking at 32% for the higher group's second breakpoint. This suggests that there is a weak tendency for those who play with a similar number of groups to affiliate.

In summation, the conclusion from this analysis is that U.S. jazz musicians are less balkanized than their peers in the Netherlands (Ijdens 1999). Instead, they form a more integrated community that is organized less by musical style than by levels of professional contacts and activity, as reflected in such factors as income from music and the number of groups in which the respondent is involved, and also by professional contacts, as reflected in network size. In addition, demographic factors, such as race and gender, also affect affiliations. The application of breakpoint analysis revealed numerous instances of both core-periphery structures and cohort structures. This finding reinforces the image of U.S. jazz musician communities as rather highly integrated systems, despite a NYC respondent's description of jazz artists as lone wolves, "self-contained and self-assured and unwilling to be in a group."

The comparison of the NYC and SF jazz musician communities also shows diversity across U.S. cities, with greater stratification by age, gender, and income from music in SF than in NYC. This difference may result from the greater numbers of highly prominent musicians and performance opportunities in NYC, such that competition for access is less severe than in SF. This situation may parallel stratification in U.S. high schools, where status derives primarily from participation in extracurricular activities such as varsity sports, cheerleading, and student government. Because the number of students who can participate in these activities tends to be independent of school size, competition is greater in larger schools, which therefore tend to be more highly stratified (Simmons and Blyth 1987). Similarly, greater competition for access to elite individuals and elite venues in SF may produce a steeper stratification system than that found in NYC.

Conclusion: Is the Jazz Musician Community a Small World?

This study documented the continuities and differences between jazz artists and other artists, and other professionals. Every profession includes some form of professional socialization, stratification system, or institutionalized means for moving up or down. Respondent-driven sampling may provide a richer measure of artists' situations than simple demographics because it can assess patterns of association and subcommunities. This provides insights into the needs of jazz musicians as highly singular individuals and lets us hear directly from the artists themselves.

Our study also has practical significance. It shows that the boundaries of the jazz musician community can be precisely determined, and therefore programs designed to foster the development of this art form can be precisely targeted and their effects documented. Reliable and consistent data about jazz musicians and their needs provide a useful guide to appropriate amounts of funding for agents, advocates, and the artists themselves.

By way of conclusion, let us consider whether the NYC and SF jazz musician networks have small-world properties. Since Milgram's (1967) articulation of the principle popularized as "six degrees of separation," a sizable literature has emerged on the small-world phenomenon, in which any two points in a large network are connected by a remarkably small number of links (Watts 1999). Many networks have been shown to have small-world properties, including the collaboration graph of film actors, the power grid of the western United States, and the neural network of the worm *Caeonorhabditis elegans* (Watts and Strogatz 1998). The question of whether jazz musician networks are small worlds is significant for two reasons. First, it provides a way of assessing the degree of integration of each jazz musician community. Second, it bears on the validity of the RDS sampling method upon which this study is based, for RDS works most effectively in small-world systems. To see why, consider how sampling would function in a large world—that is, a system where all ties are local, so path lengths are long. If, for example, individuals formed ties only to those living within one block, it would take 10 waves for the sample to extend only 10 blocks from the starting point, so a prohibitively large number of waves would be required to sample metropolitan areas as large as NYC and SF. In contrast, if in addition to local ties there are also long-distance ties, the sample can expand more rapidly and even large areas can be reached in a reasonable number of waves. The distinctive insight of Watts's (1999) analysis of the small-world phenomenon is that even modest numbers of long-distance ties can endow a network with the small-world property of short path length, so a considerable amount of local clustering is not incompatible with this property. The implication in the context of this paper is that RDS recruitment can expand quickly enough to cover a sizable metropolitan area, even in the presence of local clustering, as long as there are also long-distance ties. Examination of recruitment patterns shows that recruitment networks expand geographically rather rapidly. For example, in the network depicted in figure 1, the seed lives near Times Square, wave one recruits are separated by as much as 3.5 miles, this increases to 40 miles for wave two, and 55 miles for wave three, so rather distant parts of the metropolitan area are reached quite quickly. However, to more reliably determine whether jazz networks have small-world properties, a more systematic approach that takes into account all recruitment networks is useful.

Standard indices for measuring small-world properties, such as characteristic path length and the clustering coefficient (Watts 1999), require that the network be completely specified, so they cannot be calculated using RDS data. However, respondents' addresses provide sufficient information for a meaningful analysis. In a large world, each recruiter would tend to live closer to his or her own recruits than to randomly selected respondents. In contrast, in a system with random geographic mixing, each recruiter would tend to live about the same distance from his or her recruits as from randomly selected respondents. Comparing the two numbers thus provides a means for determining whether the jazz musician communities correspond to large- or small-world systems.

The most precise means for measuring distances between recruiters and recruits would rely on exact addresses, but zip codes offer an efficient approximation. The distance between two zip codes is assumed to be the distance between their center points, such that respondents from the same zip code have a distance of zero, and respondents from different zip codes have the same distance irrespective of where they live within the zip code area. The sizes of zip code areas vary inversely with population density. In high-density areas, such as midtown Manhattan, each zip code area is approximately one half square mile. This procedure provides a reasonable approximation of distance for areas with populations large enough to encompass many zip codes, such as NYC, where respondents were drawn from 96 zip codes, and SF, with respondents from 83.

In NYC, the mean distance between recruiter and recruit was 8.03 miles. There was also considerable variation in recruitment distances, with a standard deviation of 12.46 miles and an inter-quartile range of 2.55 to 13.25 miles. The maximum distance of 98.2 miles was attained when a respondent from Southold, NY, near Sag Harbor, recruited from South Orange, NJ. There was also evidence of local clustering: 23.29% of the recruiter-recruit pairs came from the same zip code. To estimate recruitment distance that would result from geographic random mixing, recruiters were paired with randomly selected recruits, the mean distance between recruiter and recruit was calculated, this procedure was repeated 10 times, and the estimate was the mean of the means. This procedure produced an estimated randomization-based distance of 9.71 miles. What may be termed a small-world index (SMI) can therefore be defined, where MD is the mean distance from recruiter to recruit, and RMD is randomized mean distance based on geographic random mixing,

$$(11) \quad SMI = \frac{MD}{RMD}$$

This index is zero if all recruiter-recruit pairs come from the same zip codes (each recruitment chain remains within the area in which it began); one if the mean distance equals the distance based on random mixing; and greater than one if respondents have an antipropinquity orientation (ties to more distant persons are preferred over ties to closer persons). For NYC, the index value is .83 (8.03/9.71), so random mixing is approximated to a very considerable extent of 83%.

In SF, the mean distance between recruiter and recruit was 14.22 ± 13.99 miles, with an inter-quartile range of 4.63 to 24.98 miles. The maximum distance of 49.6 miles occurred when a resident of southeastern San Francisco recruited someone from San Jose. The randomization-based mean distance was 14.22, which yields an SMI of .94 (14.22/15.19). Therefore, recruitment is SF approximates geographic random mixing-based recruitment by 94%. We conclude, first, that both the NYC and the SF jazz musician communities qualify as small-world systems: both have far more than the rather modest number of long-distance ties now recognized as required to endow a network with small-world properties. Our second conclusion is that geographic integration is greater in SF, making San Francisco the smaller world.

Table I. Recruitment by Gender and by Race and Ethnicity

Table IA: Gender, NYC jazz musicians (recruitment count; selection proportion, S) Gender of person who recruited		Gender of recruit	
	Male	Female	Total
Male	127	51	178
	0.7135	0.2865	1
Female	25	40	65
	0.3846	0.6154	1
Total recruits,	152	91	243
sample distribution, SD	0.6255	0.3745	1
Equilibrium, E	0.5731	0.4269	1
Mean network size, N	222.9885	232.2344	
Homophily, H	0.3129	0.3403	
Population estimate, P	0.5830	0.4170	
Standard error of P	0.0451	0.0451	

Table IB: Gender, SF jazz musicians (recruitment count; selection proportion, S) Gender of person who recruited		Gender of recruit	
	Male	Female	Total
Male	161	28	189
	0.8519	0.1481	1
Female	29	3	32
	0.9063	0.0938	1
Total recruits,	190	31	221
sample distribution, SD	0.8597	0.1403	1
Equilibrium, E	0.8595	0.1405	1
Mean network size, N	66.0429	65.6905	
Homophily, H	-0.0081	-0.3358	
Population estimate, P	0.8588	0.1412	
Standard error of P	0.0227	0.0227	

Table IC: Race and ethnicity, NYC jazz musicians (recruitment count; selection proportion, S) Race and ethnicity of person who recruited	Race and ethnicity of recruit				
	White	Black	Hispanic	Other	Total
Non-Hispanic white	86	28	2	8	124
	.6935	.2258	.0161	.0645	1
Non-Hispanic black	30	32	3	4	69
	.4348	.4638	.0435	.0580	1
Hispanic	6	1	0	0	7
	.8571	.1429	0	0	1
Other	11	7	0	2	20
	.55	.35	0	.1	1
Total distribution of recruits,	133	68	5	14	220
sample distribution, SD	.6045	.3091	.0227	.0636	1
Equilibrium, E	0.6095	0.3041	0.0231	0.0633	1
Mean network size, N	233.56	211.22	234.38	200.43	
Homophily, H	0.2701	0.1961	-1.0000	0.0284	
Population estimate, P (Linear least squares)	0.5801	0.3329	0.0133	0.0737	
Standard error of P	0.0405	0.0394	0.0099	0.0170	

Table ID: Race and ethnicity, SF jazz musicians (Recruitment count; selection proportion, S) Race and ethnicity of person who recruited	Race and ethnicity of recruit				
	White	Black	Hispanic	Other	Total
Non-Hispanic white	65	29	9	15	118
	.5508	.2458	.0763	.1271	1
Non-Hispanic black	19	20	2	5	46
	.413	.4348	.0435	.1087	1
Hispanic	3	3	0	1	7
	.4286	.4286	0	.1429	1
Other	15	4	1	4	24
	.625	.1667	.0417	.1667	1
Total distribution of recruits,	102	56	12	25	195
Sample distribution, SD	.5231	.2872	.0615	.1282	1
Equilibrium, E	.5114	.3036	.0575	.1275	1
Mean network size, N	52.84	85.26	97.78	88.03	
Homophily, H	-.1394	.2611	-1	.0901	

Population estimate, P (linear least squares)	0.6412	0.2340	0.0409	0.0840
Standard error of P	0.0374	0.0395	0.0157	0.0260

Table II: Affiliation by Race and Ethnicity

Table IIA: Affiliation index, NYC jazz musicians	Recipient of tie			
Source of tie	White	Black	Hispanic	Other
Non-Hispanic white	0.2701	-0.3217	0.0029	-0.1246
Non-Hispanic black	-0.2505	0.1961	0.0306	-0.2134
Hispanic	0.6598	-0.5709	-1	-1
Other	-0.0519	0.0256	-1	0.0284

Table IIB: Affiliation index, SF jazz musicians	Recipient of tie			
Source of tie	White	Black	Hispanic	Other
Non-Hispanic white	-0.1409	0.0154	0.0369	0.0471
Non-Hispanic black	-0.3558	0.2621	0.0027	0.027
Hispanic	-0.3316	0.254	-1	0.0643
Other	-0.0252	-0.2877	0.0008	0.0903

Table III: Factor Analysis by Style of Performance

Note that for both cities, the first component is positively loaded for all styles, and it explains more than four times more variance than any other component, so the dominant orientation is toward a generalist style. The other components identify clusters of styles. For example, component 2 in NYC is most strongly oriented toward bop, hard bop, and swing; and component 3 is oriented toward ragtime, scat, and boogie-woogie.

Table IIIA: Rotated component matrix, NYC jazz musicians

	Component					
	1	**2**	**3**	**4**	**5**	**6**
Acid jazz	0.567	0.221	0.400	0.116	0.308	0.151
Avant-garde	0.059	0.002	0.066	0.803	0.052	0.032
Blues	0.722	0.117	0.083	0.027	0.056	0.173
Boogie-woogie/honky-tonk	0.465	0.277	0.501	0.201	0.315	-0.056
Bop	0.050	0.809	0.057	0.025	0.094	0.067
Contemporary	0.272	0.423	0.352	0.054	-0.111	-0.105
Cool	0.554	0.343	0.384	0.087	0.365	-0.035
Free jazz	0.072	0.091	0.041	0.802	0.039	0.016
Funk	0.680	0.197	0.061	0.034	0.253	-0.019
Fusion	0.558	0.262	0.088	0.104	-0.207	-0.068
Hard bop	0.241	0.724	-0.073	0.090	0.257	0.004
Latin	0.507	0.308	0.304	-0.093	-0.153	0.101
Mainstream	0.092	0.072	0.056	0.059	0.782	-0.018
Ragtime/stride piano	0.041	0.133	0.840	0.086	-0.055	0.136
Rhythm-and-blues	0.733	0.099	0.172	0.050	0.144	0.097
Scat	0.400	-0.046	0.679	0.000	0.175	-0.094
Swing	0.238	0.618	0.196	0.005	-0.074	0.045
Traditional	0.142	0.050	0.058	0.057	-0.033	0.922
World music	0.492	-0.029	0.117	0.299	-0.237	-0.292
Percentage of variance explained	**30.948**	**7.392**	**6.809**	**5.648**	**5.509**	**5.307**

Extraction method: Principal component analysis
Rotation method: Varimax with Kaiser normalization

Table IIIB: Rotated component matrix, SF jazz musicians				
	Component			
	1	**2**	**3**	**4**
Acid jazz	0.062	0.746	0.090	0.012
Avant-garde	0.554	0.524	0.083	-0.115
Blues	0.311	0.212	0.509	0.243
Boogie-woogie/honky-tonk	0.114	0.161	0.709	-0.028
Bop	0.759	0.083	0.148	0.123
Contemporary	0.492	0.168	0.062	0.451
Cool	0.654	0.200	0.098	0.338
Free jazz	0.658	0.387	0.184	-0.132
Funk	0.272	0.625	0.299	0.214
Fusion	0.366	0.654	0.131	0.229
Hard bop	0.773	0.157	0.041	0.174
Latin	0.446	0.215	0.141	0.442
Mainstream	0.299	-0.004	0.263	0.475
Ragtime/stride piano	0.059	-0.045	0.668	0.065
Rhythm-and-blues	0.005	0.345	0.489	0.452
Scat	0.030	0.218	0.483	0.068
Swing	0.386	-0.020	0.417	0.384
Traditional	0.020	0.155	0.006	0.739
World music	0.195	0.628	0.125	0.314
Percentage of variance explained	**33.153**	**7.846**	**6.763**	**5.5**
Extraction method: Principal component analysis				
Rotation method: Varimax with Kaiser normalization				

References

Appel, Alfred, Jr. 2002. *Jazz Modernism: From Ellington and Armstrong to Matisse and Joyce.* New York: Knopf.

Becker, Howard S. 1951. "The Professional Dance Musician and His Audience." *American Journal of Sociology*, 57: 136-144.

———. 1982. *Art Worlds and Collective Activity*. Berkeley, CA: University of California Press.

Bjorn, Lars, with Jim Gallert. 2001. *Before Motown: A History of Jazz in Detroit, 1920-60.* Ann Arbor, MI: University of Michigan Press.

Blau, Peter M. 1977. *Inequality and Heterogeneity.* New York: Free Press.

———. 1994. *Structural Context of Opportunities.* Chicago, IL: University of Chicago Press.

Broadhead, R.S., D.D. Heckathorn, D.L. Weakliem, D.L. Anthony, H. Madray, R.J. Mills, and J. Hughes. 1998. "Harnessing peer networks as an instrument for AIDS prevention: Results from a peer-driven intervention." *Public Health Reports* 113(Suppl. 1):42-57.

Brown, Lawrence D., Morris L. Eaton, David A. Freedman, Stephen P. Klein, Richard A. Olshen, Kenneth W. Wachter, Martin T. Wells, and Donald Ylvisaker. 1999. "Statistical Controversies in Census 2000." Technical Report 537, Department of Statistics, University of California–Berkeley.

Burt, Ron S. 1976. "Positions in Networks." *Social Forces* 55:93-122.

Coulangeon, Philippe. 2000. *Les musiciens de jazz en France à l'heure de la réhabilitation.* Montréal, Canada: L'Harmattan.

Erickson, Bonnie H. 1979. "Some Problems of Inference from Chain Data." *Sociological Methodology* 10:276-302.

Fararo, Thomas J., and John Skvoretz. 1984. "Biased Networks and Social Structure Theorems: Part II." *Social Networks* 6:223–58.

Fararo, Thomas J., and Morris H. Sunshine. 1964. *A Study of a Biased Friendship Net.* New York: Syracuse University Press.

Frank, Ove, and Tom Snijders. 1994. Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics* 10:53-67.

Hagan, John. 2001. *Northern Passage: American Vietnam War Resisters in Canada.* Cambridge, MA: Harvard University Press.

Heckathorn, Douglas D. 1990. "Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control." *American Sociological Review* 55(June):366-384.

———. 1993. "Collective Action and Group Heterogeneity: Voluntary Provision Versus Selective Incentives." *American Sociological Review* 58:329-50.

———. 1996. "Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61:250-77.

———. 1997. "Respondent Driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44:174-99.

———. 2002a. "Respondent-Driven Sampling II: Deriving Statistically Valid Population Estimates from Samples of Hidden Populations." *Social Problems* 49:11-34.

———. 2002b. "Development of a Theory of Collective Action: From the Emergence of Norms to AIDS Prevention and the Analysis of Social Structure," pp. 79-108 in *New Directions in Sociological Theory: Growth of Contemporary Theories* (Joseph Berger and Morris Zelditch, editors). Rowman and Littlefield, .

Heckathorn, Douglas D., Robert S. Broadhead, Denise L. Anthony, and David L. Weakliem. 1999. "AIDS and Social Networks: Prevention through Network Mobilization." *Sociological Focus* 32(2):159-179.

Heckathorn, Douglas D., and Joan Jeffri. 2001. "Finding the Beat: Using Respondent-Driven Sampling to Study Jazz Musicians." *Poetics* 28:307-329.

Heckathorn, Douglas D., and Judith E. Rosenstein. 2002. "Group Solidarity as the Product of Collective Action: Creation of Solidarity in a Population of Injection Drug Users." *Group Cohesion, Trust, and Solidarity* 19:37-66.

Heckathorn, Douglas D., Salaam Semaan, Robert S. Broadhead, and James J. Hughes. 2002. "Extensions of Respondent-Driven Sampling: A New Approach to the Study of Injection Drug Users Aged 18-25." *AIDS and Behavior* 6:55-68.

IJdens, Teunis. 1999. *Schots en scheef. Artistiek werk tussen markt en organisatie.* Rotterdam: Erasmus Research Center for Art and Culture.

Jeffri, Joan. 1989. *Information on Artists.* New York: Research Center for Arts and Culture, Columbia University.

———. 1998. *Information on Artists II.* New York: Research Center for Arts and Culture, Columbia University

Jeffri, Joan, et al. 1990a. *The Artists Training and Career Project: Actors.* New York:
Research Center for Arts and Culture, Columbia University.

__________. 1990b. *The Artists Training and Career Project: Craftspeople.* New York: Research Center for Arts and Culture, Columbia University.

___________. 1990c. T*he Artists Training and Career Project: Painters.* New York: Research Center for Arts and Culture, Columbia University.

Kemeny, John G., and J. Laurie Snell. 1960 Finite Markov Chains. Princeton, NJ: Van Nostrand.

Klovdahl, Alden. S. 1989. Urban Social Networks: Some Methodological Problems and Possibilities. Pp. 176-210 in *The Small World,* edited by M. Kochen. Norwood, NJ: Ablex Publishing.

Laumann, Edward O., John H. Gagnon, Stuard Michaels, Robert T. Michael, and James S. Coleman. 1989. "Monitoring the AIDS epidemic in the United States: A Network Approach." *Science* 244:1186-1189.

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* 52:370-379.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-444.

Milgram, Stanley. 1967. "The Small World Problem." *Psychology Today* 2:60-67.

Palmer, Richard. 2002. *A Jazz Odyssey: The Life of Oscar Peterson.* New York, NY: Continuum.

Putnam, Robert D. 2000. Bowling Alone : The Collapse and Revival of American Community. New York, NY : Simon & Schuster.

Rapoport, Anatol. 1979. "A Probabilistic Approach to Networks." *Social Networks* 2:1–18.

Salganik, Matthew J. and Douglas D. Heckathorn. 2003. "Making Unbiased Estimates from Hidden Populations Using Respondent-Driven Sampling." Paper presented at the *International Social Network Conference*, Cancun, Mexico

Semaan, Salaam, Jennifer Lauby,. and Jon Liebman 2002. "Street and Network Sampling in Evaluation Studies of HIV Risk-Reduction Interventions." *AIDS Review* 2002:4:213-223.

Shipton, Alyn. 2002 *A New History of Jazz.* London: Reed Elsevier, Inc.

Simmel, Georg. 1955. *Conflict: The Web of Group Affiliations.* Trans. Kurt H. Wolff and Reinhard Bendix. New York: Free Press.

Simmons, Roberta G., and Dale A. Blyth. 1987. *Moving into Adolescence.* New York: Aldine de Gruyter.

Thompson, Steven K., and Ove Frank. 2000. Model-based estimation with link-tracing sampling designs. *Survey Methodology* 26:87-98.

Throsby, David. 1983. "The Artist in Australia Today. Report of Individual Artists' Inquiry." Sydney: Australia Council.

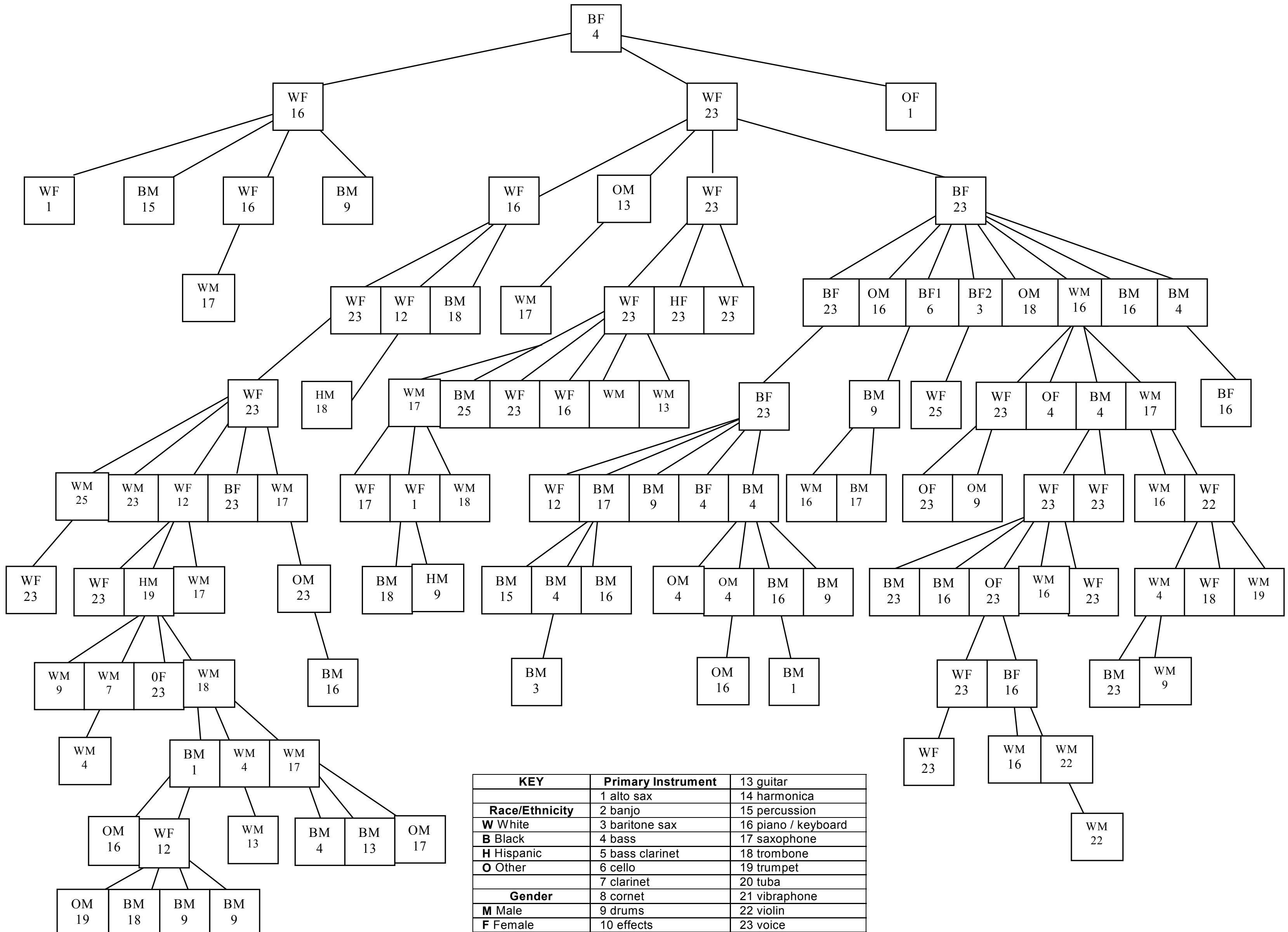
Throsby, David, and Devon Mills. 1989. "When Are You Going to Get a Real Job? An Economic Study of Australian Artists." Sydney: Australia Council.

Watts, Duncan J. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness.* Princeton, NJ: Princeton University Press.

Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393:440-442.

Yanow, Scott. 2001. *Trumpet Kings: The Players Who Shaped the Sound of Jazz Trumpet.* San Francisco, CA: Backbeat: Music Player Network.

Figure 1: Recruitment Network in a Respondent-Driven Sample: Jazz musicians in NYC, beginning with a single seed.



KEY	Primary Instrument	13 guitar
	1 alto sax	14 harmonica
Race/Ethnicity	2 banjo	15 percussion
W White	3 baritone sax	16 piano / keyboard
B Black	4 bass	17 saxophone
H Hispanic	5 bass clarinet	18 trombone
O Other	6 cello	19 trumpet
	7 clarinet	20 tuba
Gender	8 cornet	21 vibraphone
M Male	9 drums	22 violin
F Female	10 effects	23 voice
	11 flugelhorn	24 xylophone
	12 flute	25 other

Figure 2: Recruitment as a Markov Process. Note that every node can be reached from every other node, and no nodes are absorbing states. Therefore, the network is ergodic.

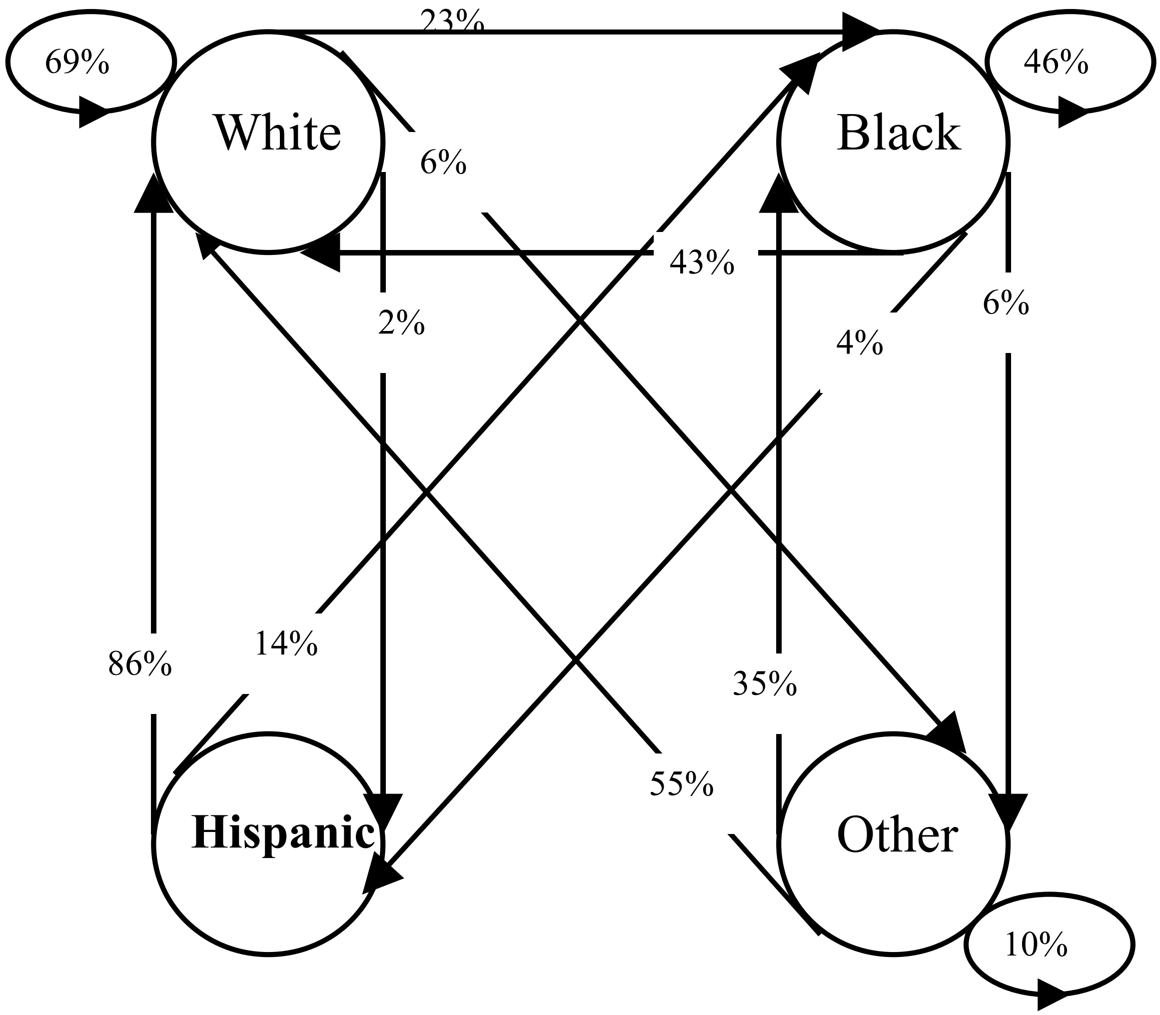
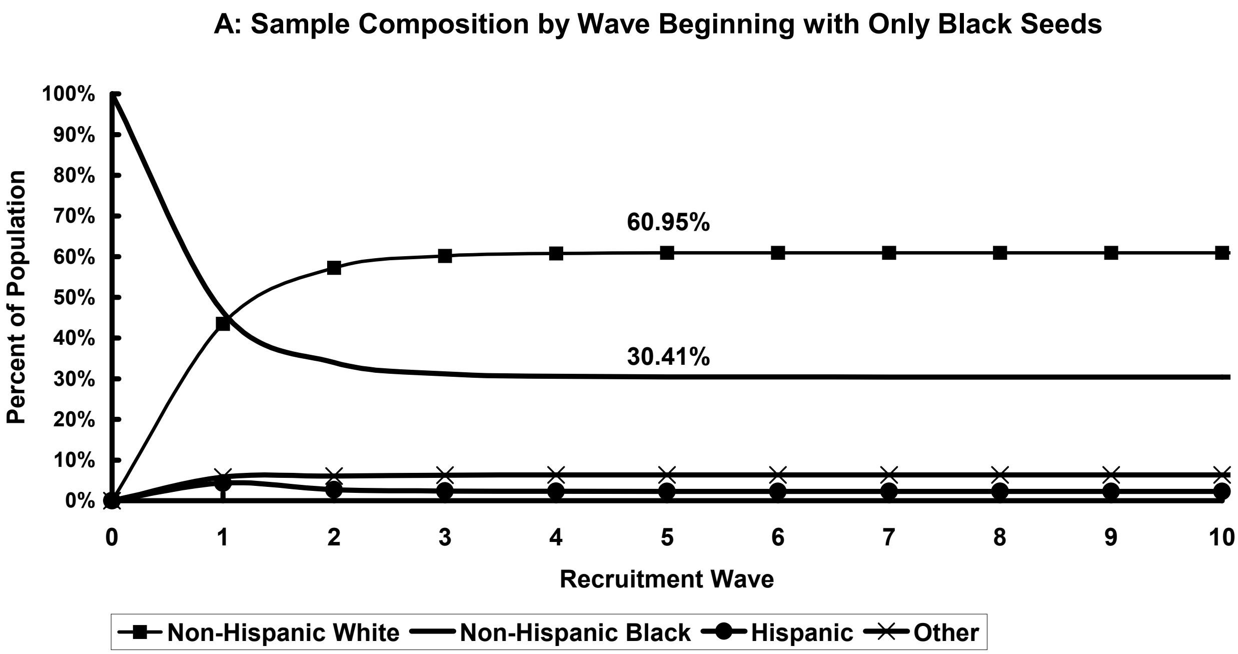


Figure 3: Two Simulations of Recruitment in a Respondent-Driven Sample: Race and ethnicity of recruits in a respondent-driven sample, beginning with all non-Hispanic white or black seeds.



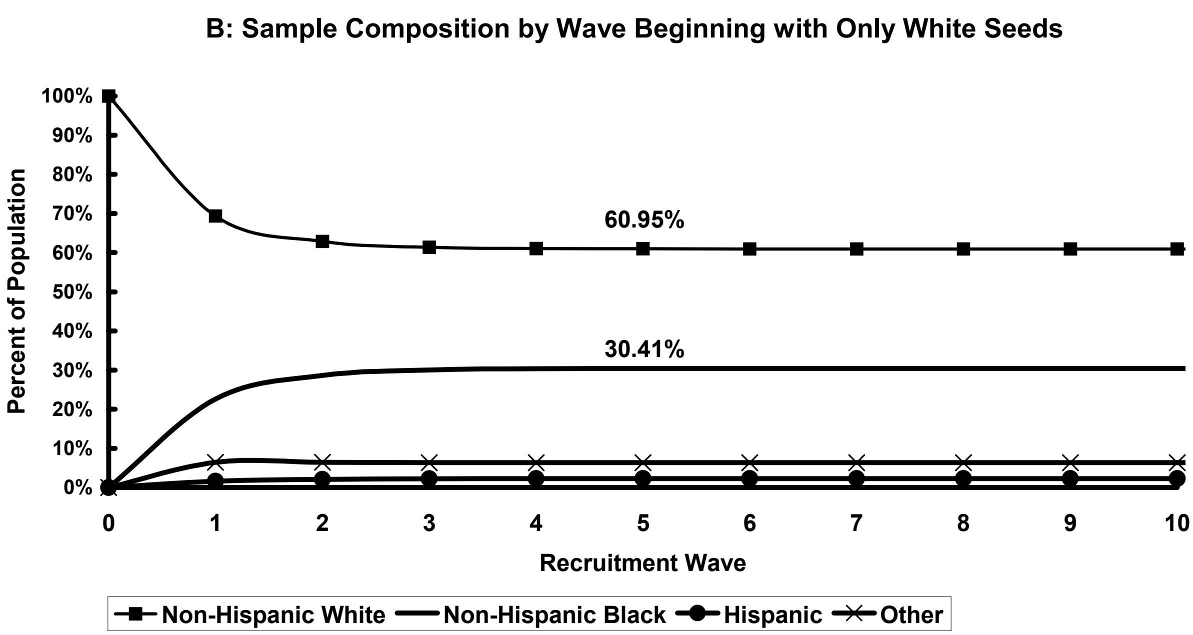
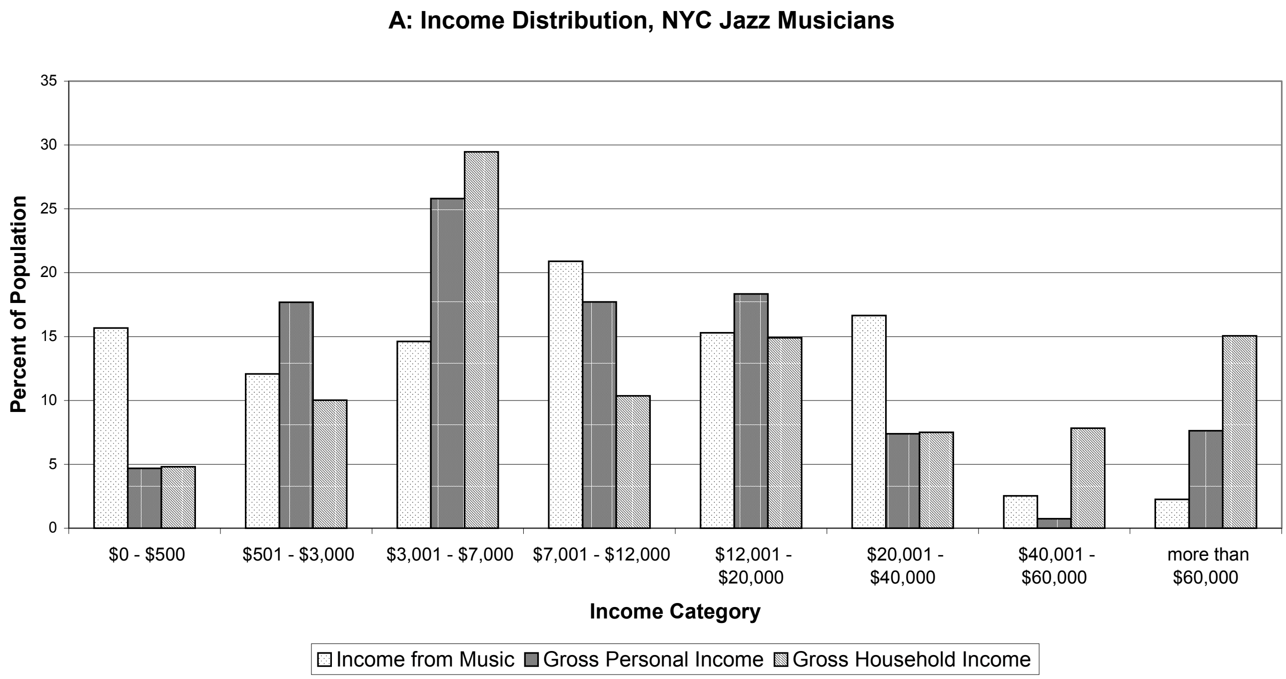


Figure 4: Estimated Income Distribution of Jazz Musicians, NYC and SF. Note that incomes are substantially higher in NYC, especially income from music.



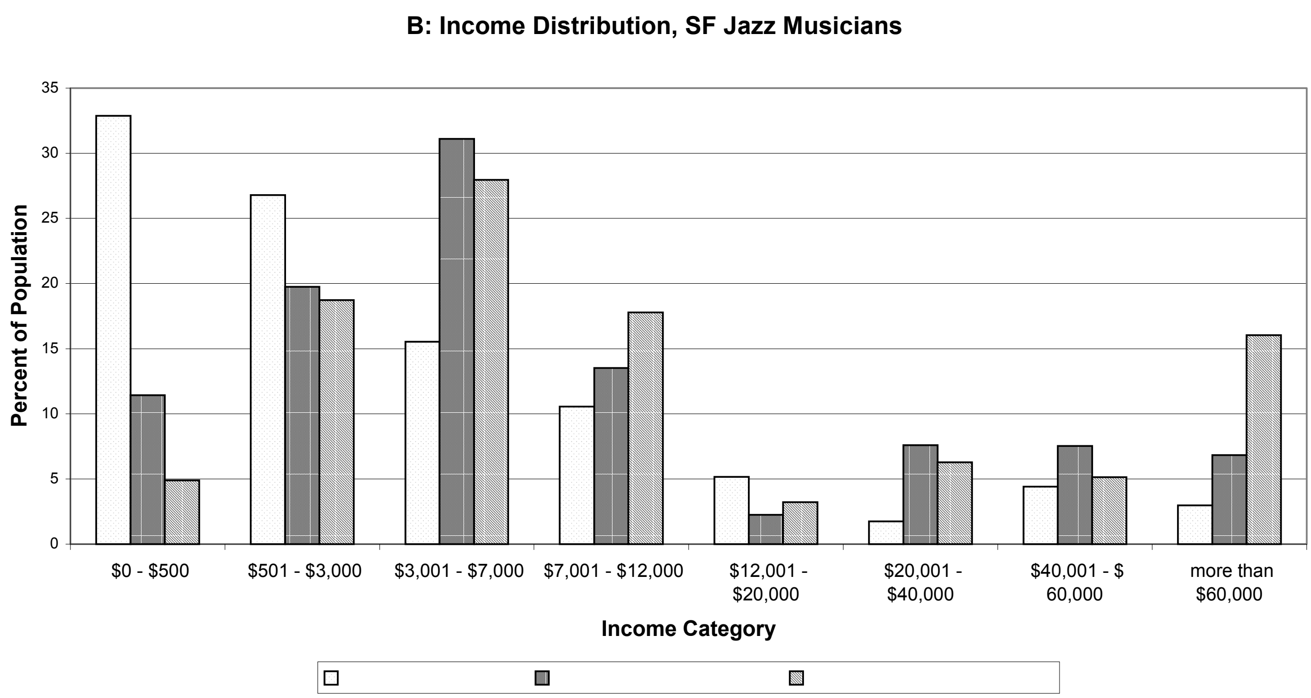
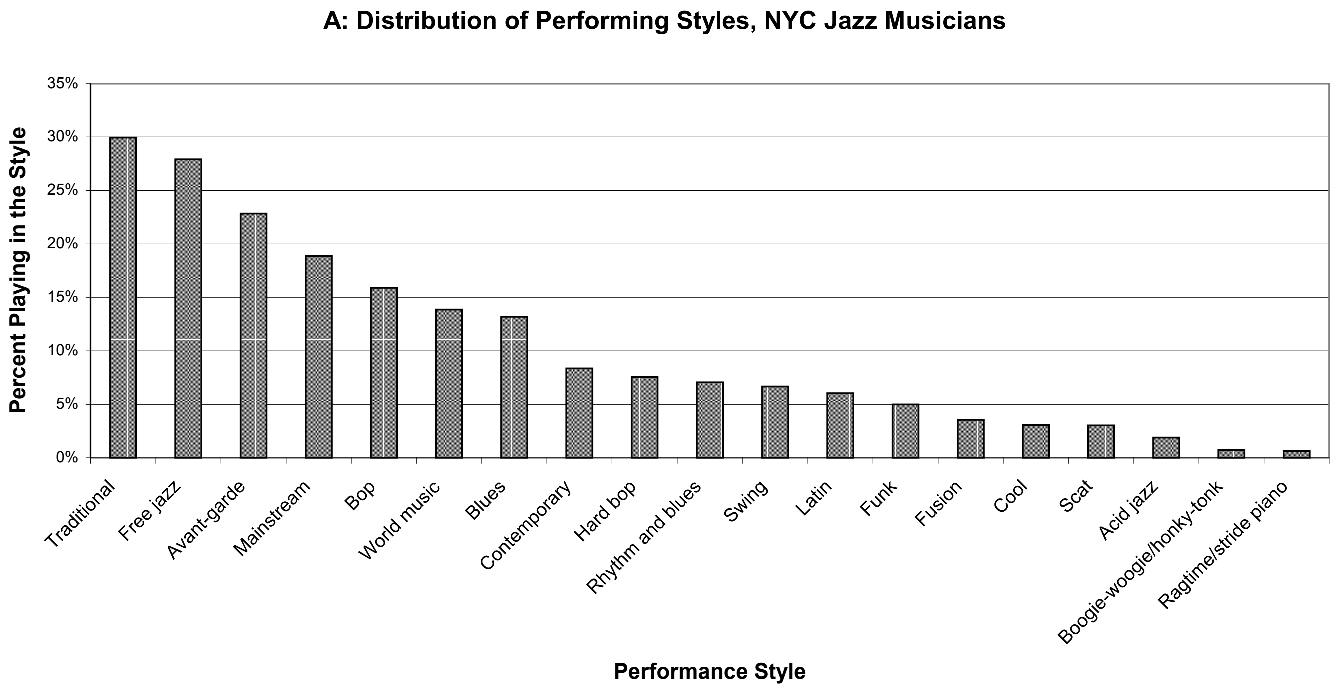


Figure 5: Distribution of Styles of Performance. Styles are listed in order of popularity. Musicians in NYC are more specialized, playing on average in only 2.3 styles, compared with 7.09 styles for SF musicians.



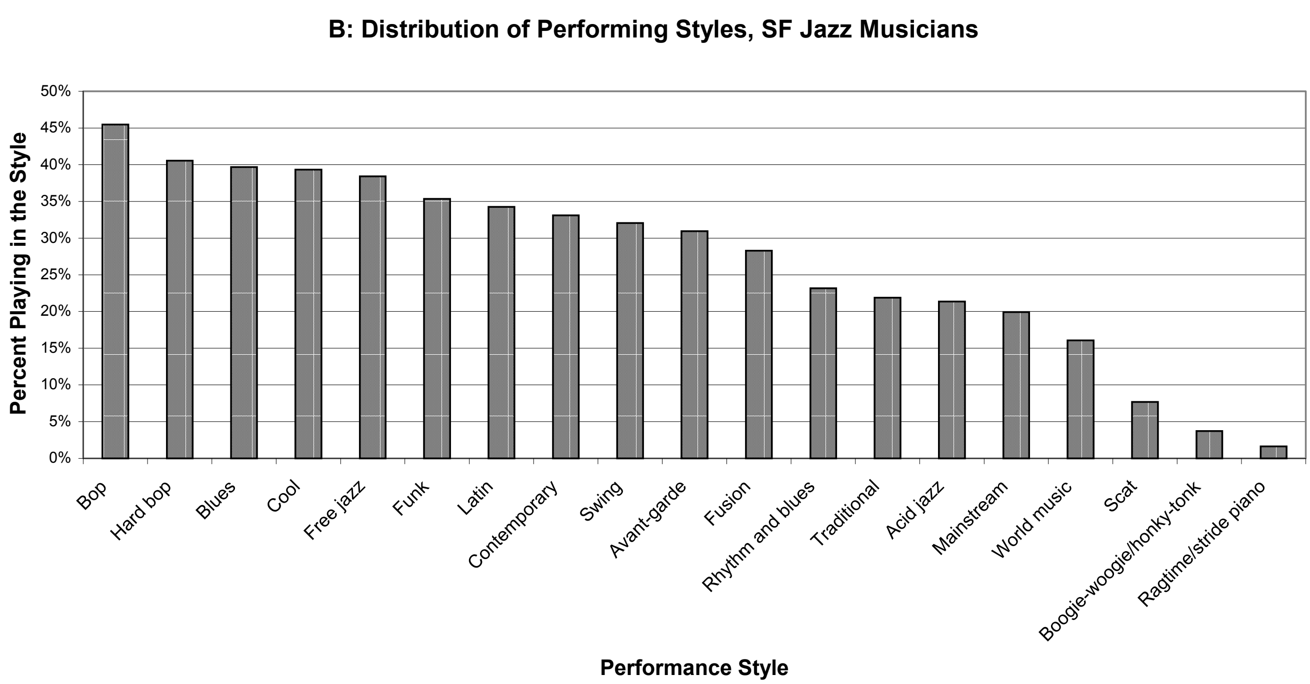
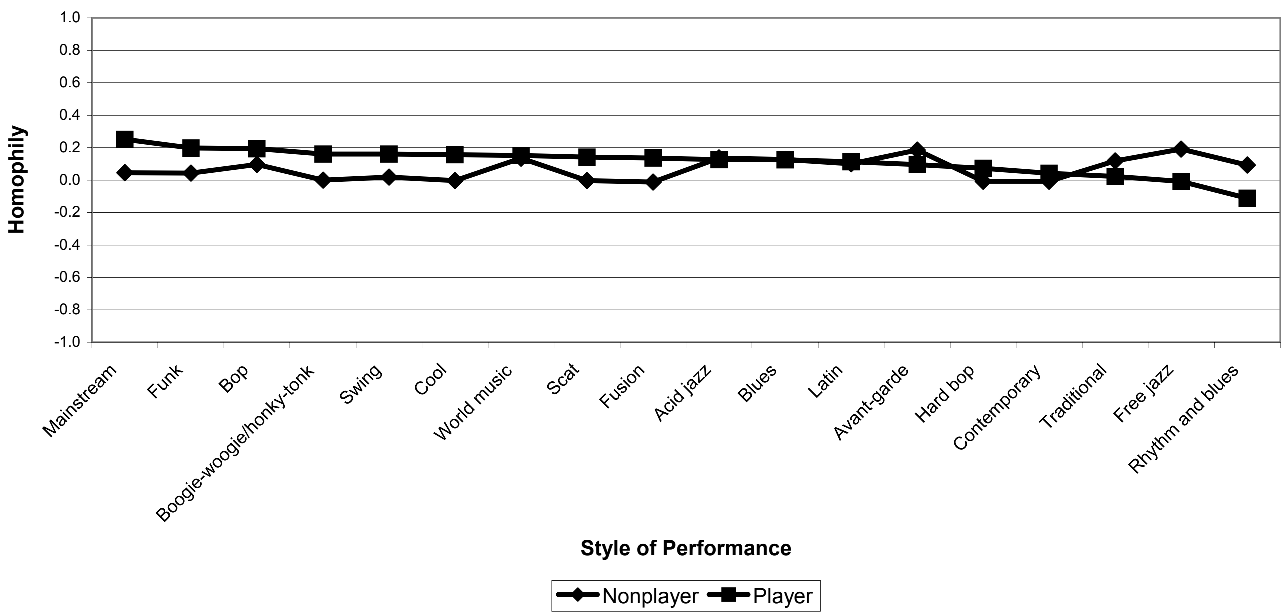


Figure 6: Homophily by Style of Performance. Styles are listed in descending order of homophily for those who perform in the style. Style of performance is a stronger determinant of affiliation in SF than in NYC, but style is a weaker determinant than network size, age, and income. This reflects a pattern in which most musicians are generalists who play in multiple styles, rather than specialists.

A: Homophily by Style of Performance, NYC Jazz Musicians



B: Homophily by Style of Performance, SF Jazz Musicians

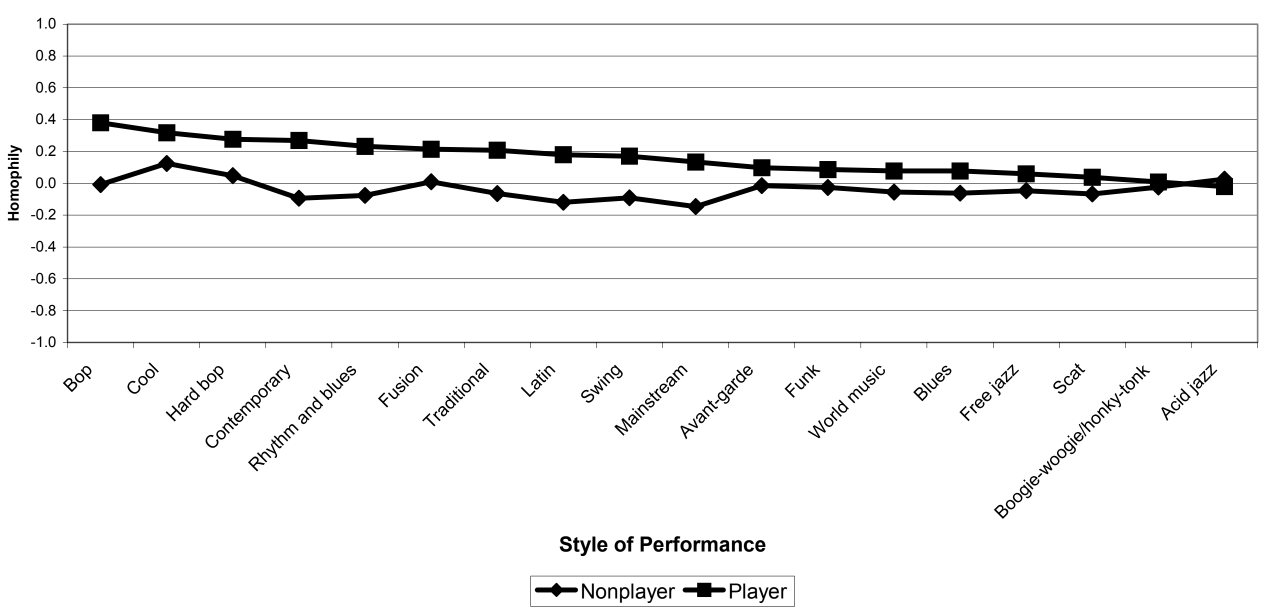
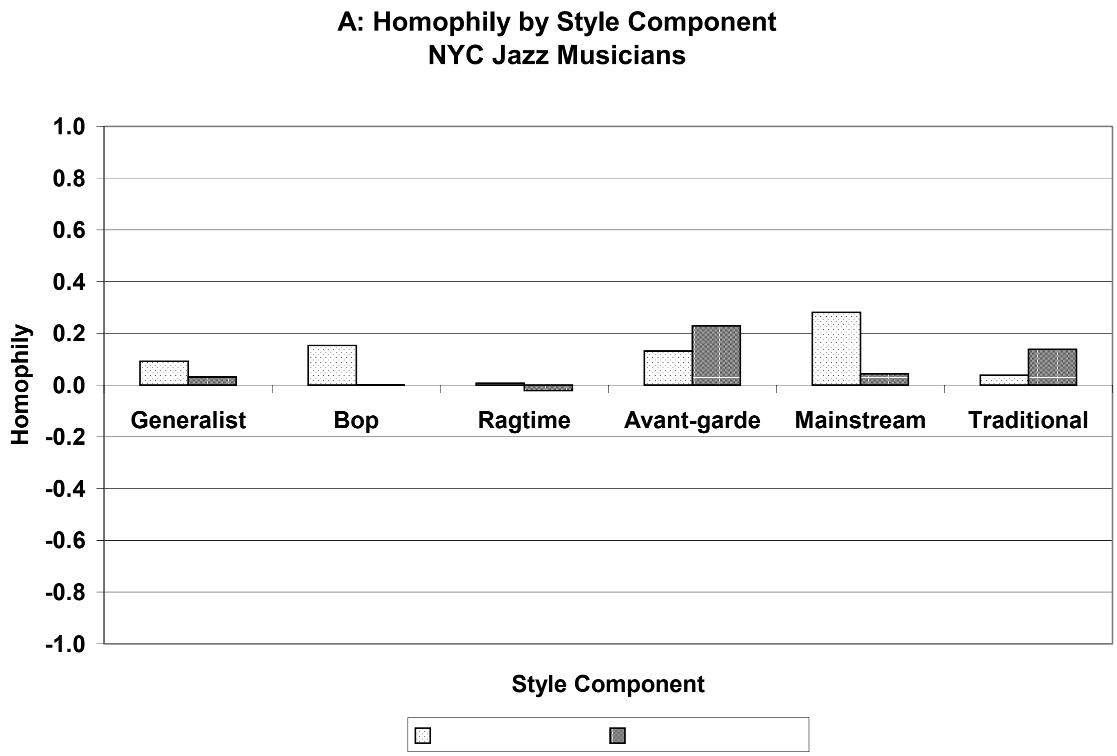


Figure 7: Homophily by Style Factor Scores. The first component is a generalist orientation. Other components are named for the most heavily loaded style.



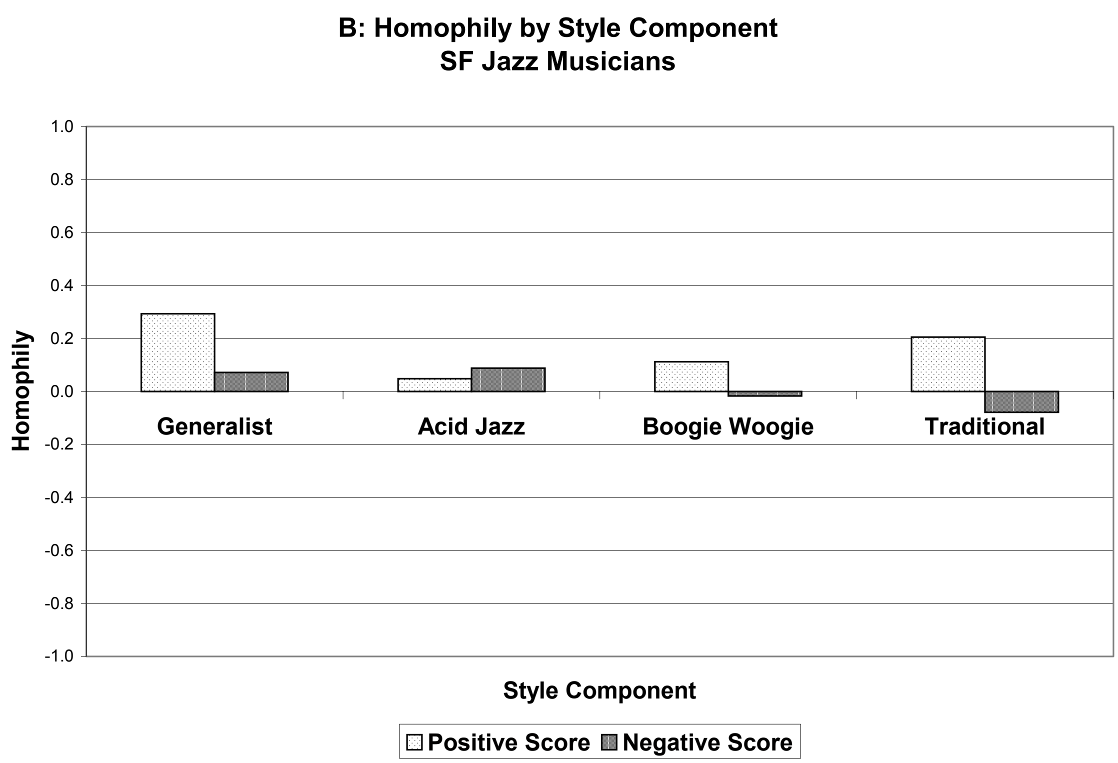


Figure 8: Homophily by Primary Instrument, NYC and SF Jazz Musicians

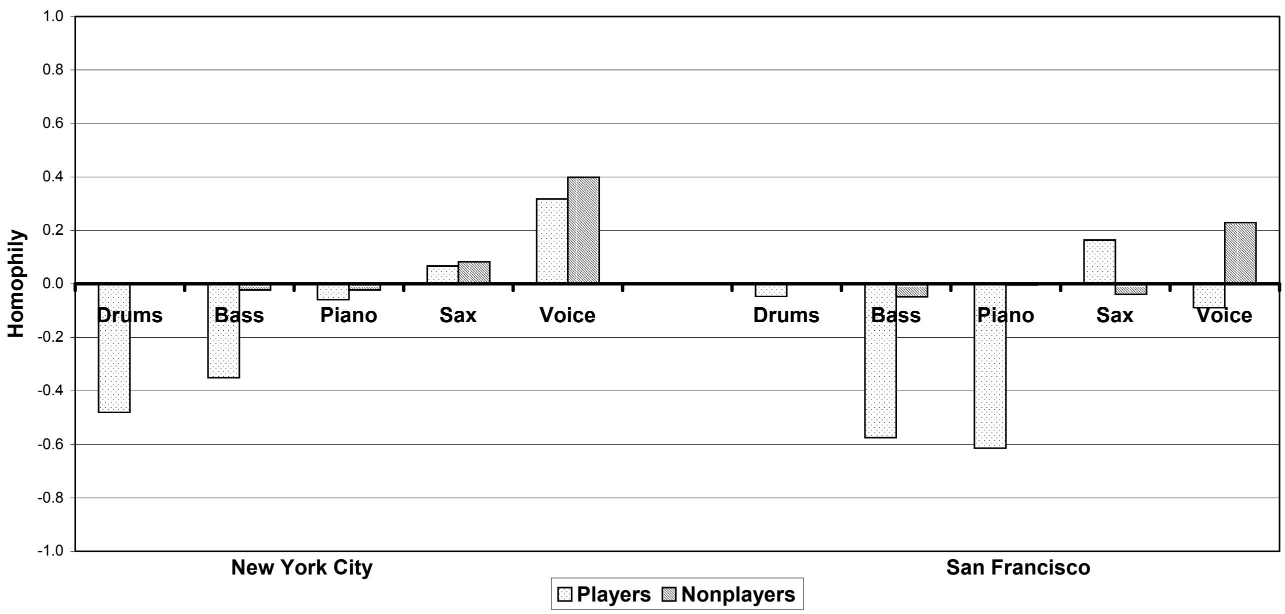
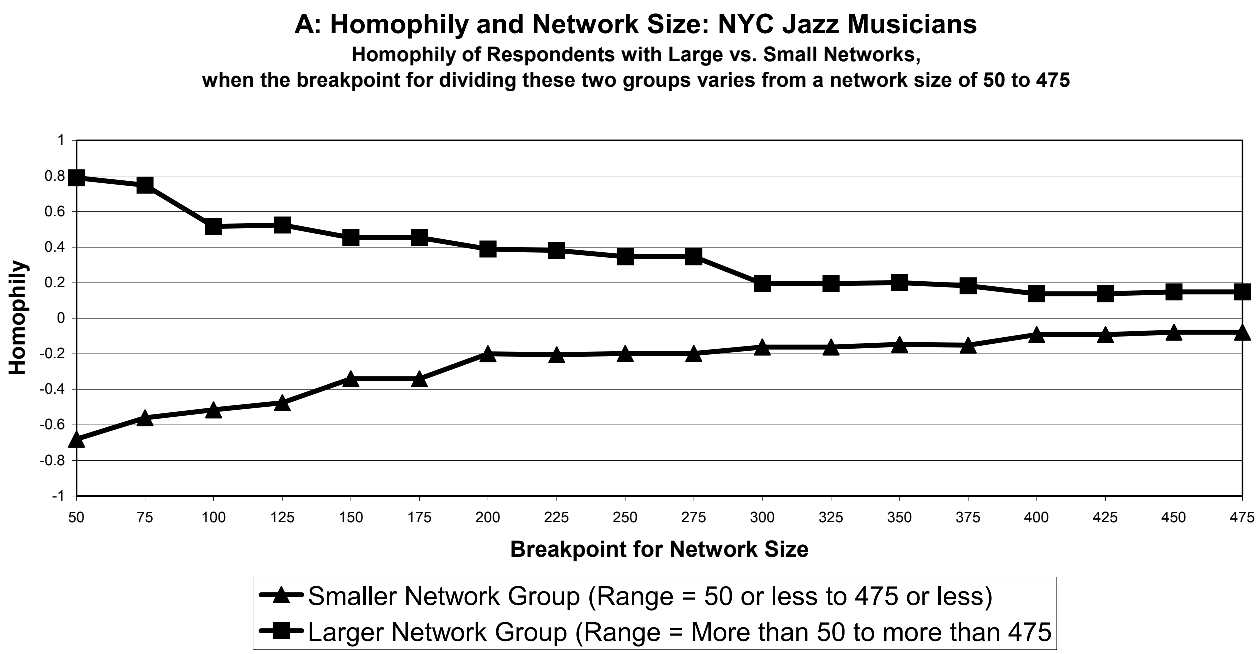


Figure 9: Homophily and Network Size. The graphs for both NYC and SF illustrate a core-periphery structure, in which well-connected persons are densely interconnected, and poorly connected persons interact primarily through those who are better connected. Despite the substantially larger networks in NYC, the structures are remarkably similar.



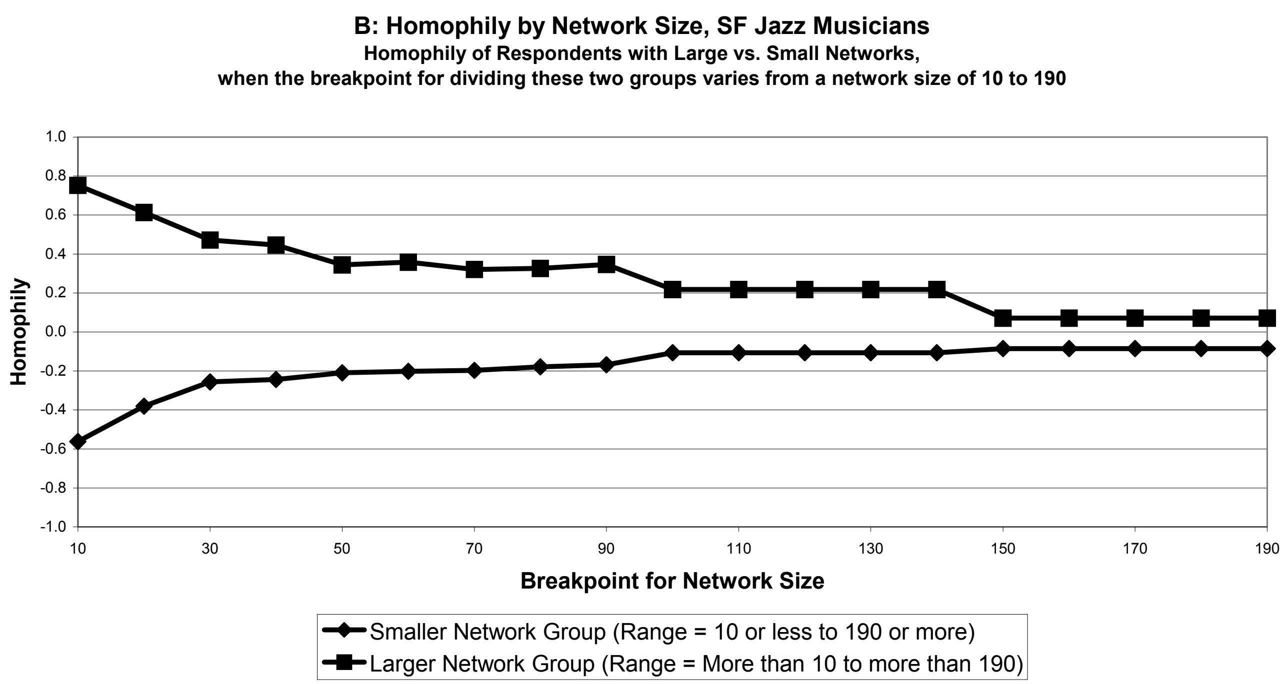


Figure 10: Homophily and Age. The graph for NYC illustrates a cohort structure, in which persons tend to affiliate with those of similar age; irrespective of breakpoint, then, homophily for the older and younger groups is positive. Homophily declines with the breakpoint for age because the significance of a fixed difference is age is less for older than for younger persons. In contrast, the graph for SF displays a form of core-periphery structure, in which older musicians are more densely networked, and younger musicians interact indirectly, through older musicians. Hence, inequality by age is substantially greater in SF than in NYC.

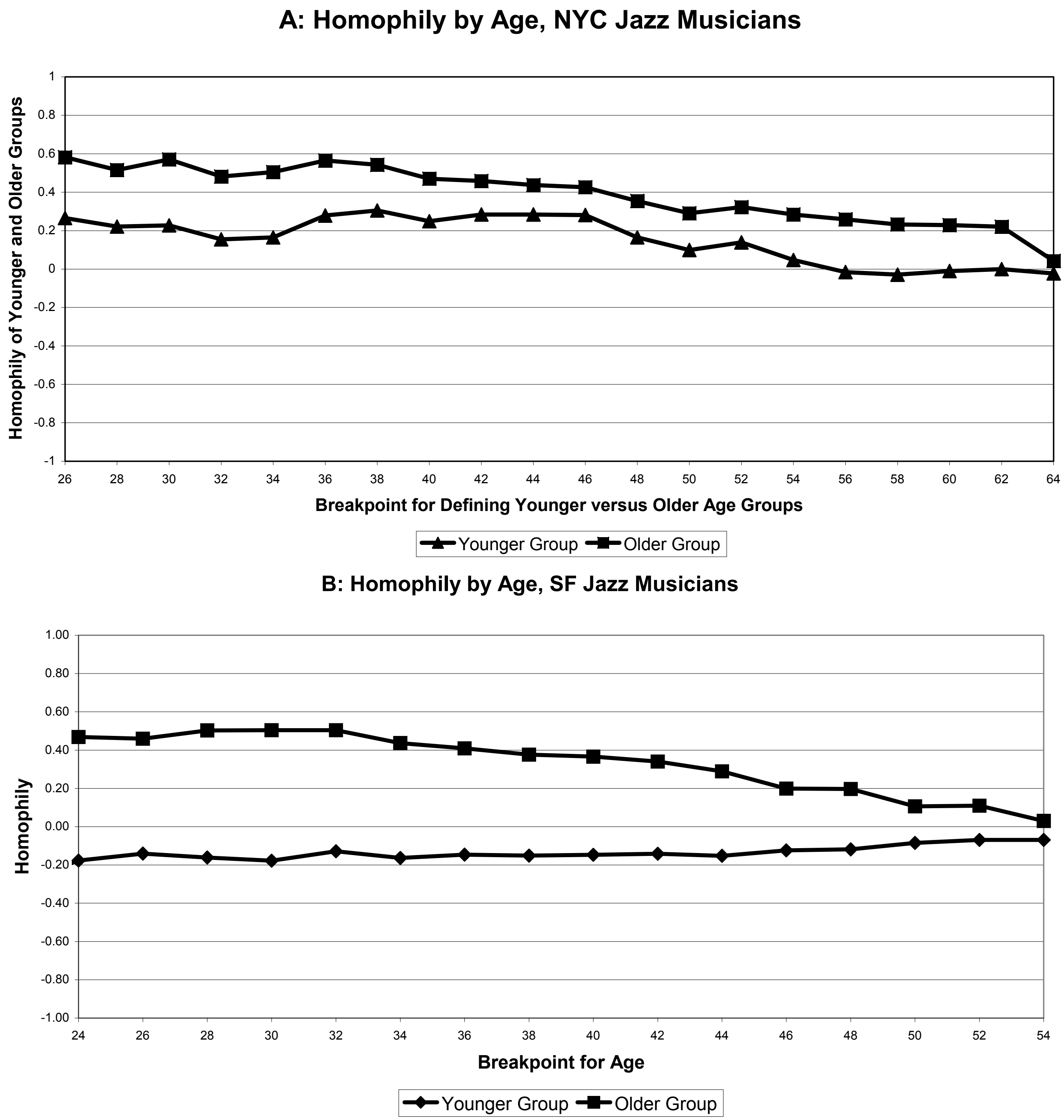
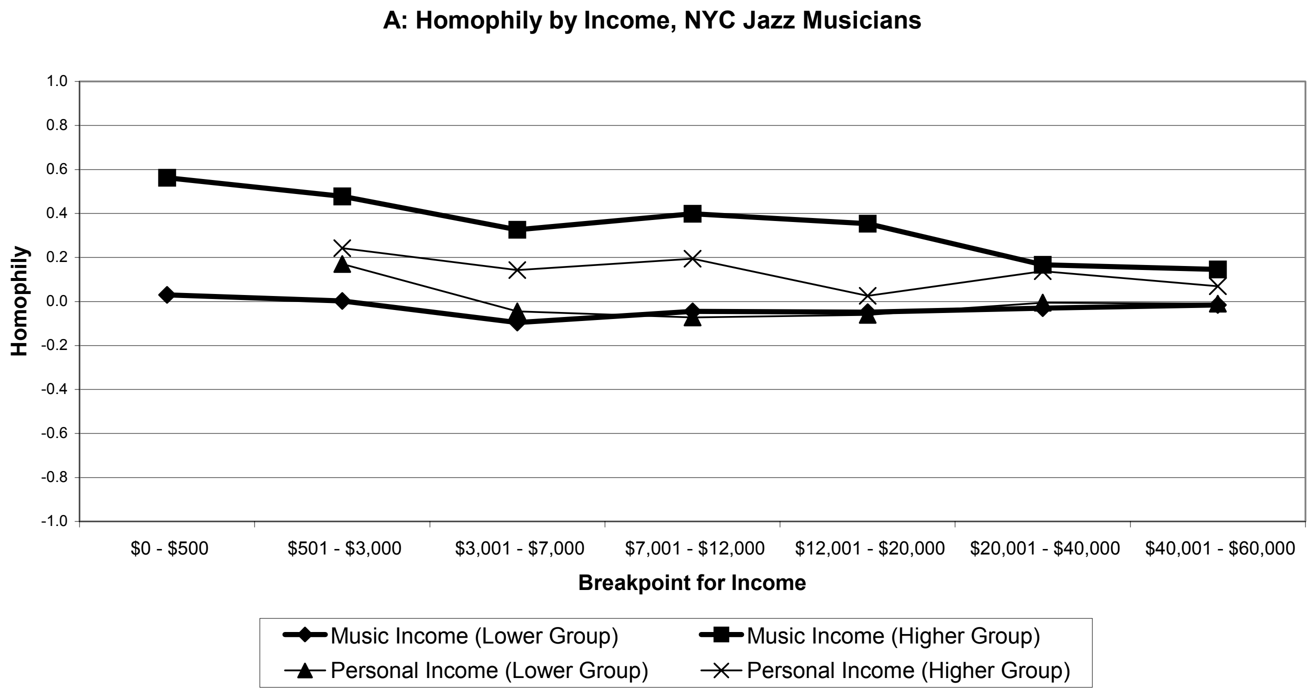


Figure 11: Homophily by Income from Music, Personal Income, and Household Income. Note that for each form of income, the higher income group has more homophily, and that differences in homophily decline for higher breakpoints for income. Note also that except for higher-income musicians in SF, income from music has the strongest effect on affiliation patterns. In NYC, too few musicians had personal or household incomes from $0 to $500 for homophily to be calculated.



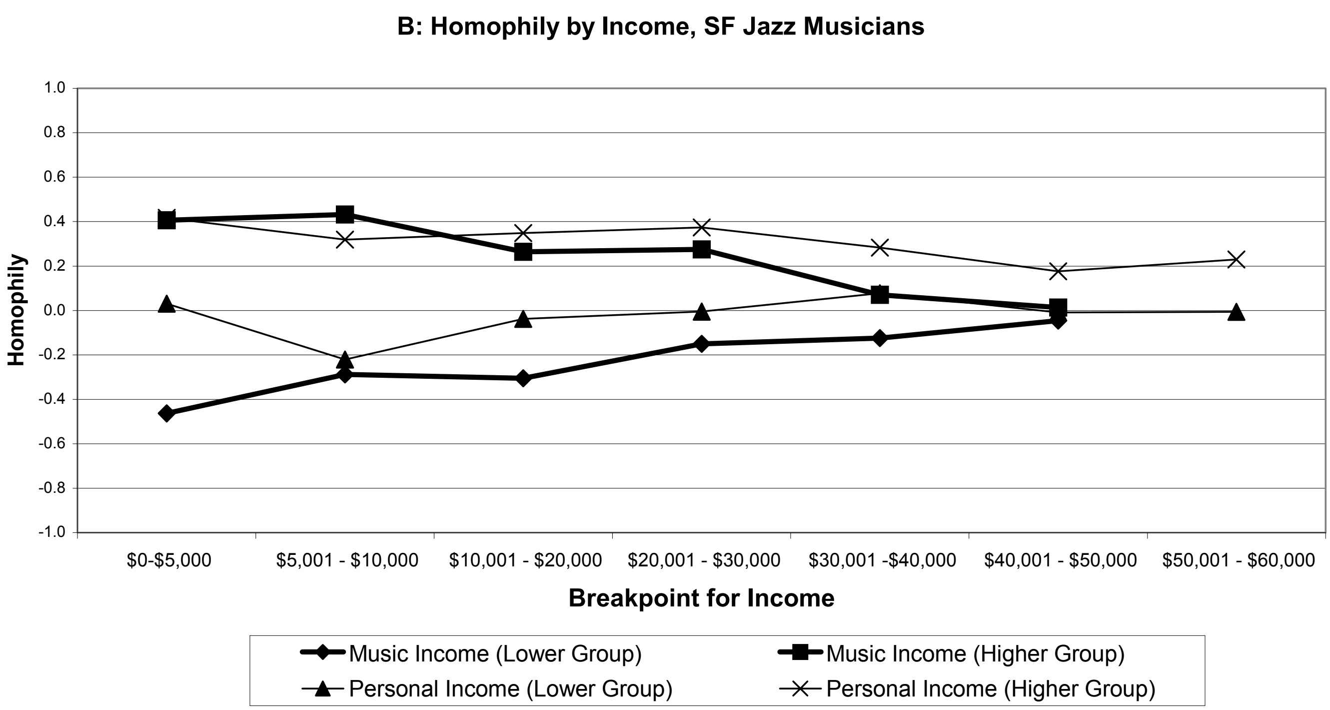


Figure 12: Homophily by Number of Groups in Which the Respondent Performs. NYC is characterized by a weak cohort structure, in which those who perform with more groups are more homophilous. In contrast, SF is characterized by a strong core-periphery structure, in which those who do not play with any groups are strongly heterophilous, and those who perform with more than one group are strongly homophilous.