

Discovering micro-genre Communities in Survey Data on Cultural Tastes

Omar Lizardo

February 5, 2022

1 Persons and Cultural Choices

Recent work in the sociology of taste has moved to explode the distinction, foundational for much early work, between “relational” network data (codifying the relationships between people) and survey data (codifying the relationship between people and “variables”). The basic idea is that all data, whether collected purposefully as “network” data, or collected as part of a social survey is *relational* data. The main difference is the types of entities that are related [Borgatti and Everett, 1997]. That is, any data source that can be stored in matrix form has “modes” (how many types of entities are related) and “ways” (the number of entities linked by a relation, usually two at a time). The standard network data set is “one mode” (we are usually looking at one type of entity at a time, like people, organizations, schools, countries, and the like) and two “ways” (people-by-people or organization-by-organization pairs). In the same way, the usual person-by-variables survey data has two modes and two ways. The relations are not people-by-people, but people-by-variables. People are connected to the survey items they answer by relations of affinity, disagreement, choice, or even negatively (by not answering). Following, Breiger [1974], it is possible to recover person-by-person (one mode, two ways) matrices from the usual person-by-variables data matrices. People can be connected to others if they share the same values, opinions, tastes, practices, or demographics recorded as variables in the columns of the matrix.

When transplanted to the sociology of taste, for which survey data has been the workhorse source of insights and empirical generalizations [Peterson and Kern, 1996, Bryson, 1996, Van Eijck, 2001, Savage and Gayo, 2011] this blurring of the boundaries between the two main sources of relational data in the social sciences can be revelatory. This goes beyond the fact, to be exploited below, that once we treat survey data as network data, then the entire methodological panoply developed by social network analysts (and increasingly “network scientists” working across many disciplines) for the last fifty or so years becomes part of the analytic toolbox. In addition to this, the entire *conceptual* arsenal of social network theory [Borgatti and Halgin, 2011] also becomes available. As Borgatti and Everett [1997, 244] once noted, once we convert “2-mode data sets into 1-mode matrices... we can then apply the techniques (if not the theories) of network analysis.”

Accordingly, a spate of recent work has begun the job of theoretical translation, enriching the first-generation of work in the sociology of taste with network-flavored concepts. Accordingly, Pachucki and Breiger [2010] extend Ronald Burt’s theory of structural holes, for understanding how people may be able to bridge gaps not just in one-mode person-to-person networks, but in two-mode person-to-culture networks. Following this lead, Lizardo [2014] provides a metric based on Burt’s conception of network efficiency that is meant to characterize the slippery concept

of “omnivorousness” with respect to genres and cultural forms that people engage in. This approach goes beyond just summing or counting the cultural engagements of people to take into account the audience overlap between the genres themselves. Thus, the true omnivore is a person who engages genres which themselves have low audience overlap. Other work in this same vein extends ideas related to positional equivalence in networks White et al. [1976] to uncover “blocks” of respondents that tend to dislike the same set of genres that others like [Okada, 2017]. A similar approach, based on ideas of positional equivalence in networks, can be extended beyond the study of beliefs and social attitudes more generally, to identify respondents who share cultural schemas [Goldberg, 2011]. Lizardo [2018] adapts techniques first developed for the study of economic complexity in two-mode networks of geographic sites and products/technologies for the characterization of genres (e.g., popular versus niche) and audience (e.g., omnivore versus “univore”) characteristics in survey data on cultural tastes.

2 The Problem of Genre in the Sociology of Taste

In this paper, I continue to ride this wave of adapting network approaches to survey data on cultural tastes to tackle a fundamental problem (both substantive and measurement-wise) in the sociology of taste, namely, the problem of *genre*. In a foundational paper, DiMaggio [1987] provided a prescient conceptualization of the core constructs of the sociology of taste, in a way that exploited the network imagery that has become a routine reality of late. In the paper, DiMaggio [1987, 244] asked us to

...imagine a matrix defined by persons on the vertical axis and artworks on the horizontal axis, with...signifying relationships (knowledge about, like for, dislike of) between person and artworks, genres consist of those sets of works which bear similar relations to the same set of persons. The logic behind this imagery will be familiar to students of network analysis as one of “structural equivalence.”

Thus, for DiMaggio, genres are dual entities (with respect to people) precisely in Breiger’s sense defined above. Genre categories are composed of the audiences that engage them (are subsets of the larger set of people). People, on the other hand, are related to one another (e.g., via relations of similarity, opposition, or non-overlap) by the genres they choose, and genres related to one another via overlaps (intersections) in the sets of people who choose them Lizardo [2018].

2.1 Overlapping Boundaries

This conceptualization, like a lot of network theorizing, is elegant, but it gets messy in the application. As Lena [2015, 149] has noted, “[n]o ordering principle is as fundamental to culture as genre” yet, none is also as vexed. One issue, raised by Lena, is that sociologists have relational intuitions about genres but tend to default to musicological definitions for convenience’s sake. Thus, when it comes to thinking about which genres to include in a survey on cultural tastes, the tendency is simply to pick off the shelf from the menu of genre categories that exist in the world. The problem with these folk genre classifications is that they tend to miss the fuzzy, overlapping ways that real world genres are organized. This lead us to treat what are contested boundaries as if they were natural, crisp boundaries (e.g., assigning respondent who like “Rap” to a different taste community than those who like “Opera.”). This is the problem of *overlapping boundaries among genre categories*.

2.2 Micro-genre differentiation

The other issue, has to do with the level of aggregation. Survey data, for convenience's sake, must settle on a "middle" level of categorization so as not to tax respondent's patience (or knowledge). Thus, the standard "lists" of cultural genres, such as "Pop Art," "Romantic Comedy," or "Blues." However, as recent work points out people, including organizations and other powerful institutions, make sociologically relevant distinctions within those macro-genres categories. These micro-genres are embedded within larger macro-genres in complex ways that are missed by the standard survey approach.

One solution, suggested by Vlegels and Lievens [2015, 2017] concerning musical genres is to "change the mode" and "drop the label" by asking respondents not just for their levels of engagement or liking of broad macro-genre labels but for performers within those labels. This work shows that once these "micro-genre" distinctions based on performers are made, a number of the core empirical generalizations in the sociology of taste (e.g., highly educated people like classical music) are controverted, or at least heavily qualified. This *micro-genre heterogeneity* critique thus asserts that the standard genre labels studied by social scientist hide as much as they reveal, because micro-genre communities can be internally diverse and even contain mutually opposed groups in the social space Flemmen et al. [2018]. For people, the unit of selection and judgment is the micro-genre not the "broad" genre classification, so it is unclear what to do (and how to interpret) the usual data collected by sociologists who study taste using surveys.

But what happens when all the data we have (as it is with the long-standing surveys such as the Survey for Public Participation in the Arts or the General Social Survey) is at the macro-genre label. Both overlapping boundary and micro-genre heterogeneity critics of traditional genre labels would say to simply ignore these data and collect new data that does not rely on misleading macro-genre label. Here I propose another solution: Get relational! More accurately, my point will be that by exploiting the inherent relationality already contained in the usual survey data (when treated as two-mode network data), old data can reveal insights that pertain to the very features sociological critics of the macro-genre label concept say is missing. In particular, it is possible to use embedded sets of relations between people and genres to discover the very heterogeneous micro-genre communities critics say are hidden by the macro-genre label (thus my point is that these are hidden but not necessarily absent). This approach is thus ideal for making new use of old data.

3 Data to be Used in the Study

The empirical part of the study uses data on the musical tastes of Americans collected with Sara Skiles in the summer of 2012 [Lizardo and Skiles, 2015, 2016]. The survey that was (partially) designed to replicate the General Social Survey (GSS) 1993 Culture Module. This is now a canonical data set providing the empirical basis for a variety of analyses (and reanalyses) in the sociology of taste [e.g., Bryson, 1996, Goldberg, 2011, Schultz and Breiger, 2010, Han, 2003, Tampubolon, 2008, Okada, 2017]. The data were collected by an organization then called Survey Sample International (SSI), a private firm specializing in sampling, data collection, and analysis. SSI managed recruitment and participation invitation tasks to generate a panel of adults from which our working sample was drawn. Survey respondents were selected from the panel for participation on the basis of age, gender, race, education, and geographic region to approximate a sample representative of the U.S. population ($n = 2,263$). Like GSS 1993, SSI 2012 included items assessing respondents' likes and dislikes (as well as a middle category of "mixed feelings") for 20 categories of musical style: classical/ symphony and chamber, opera/operetta, jazz, Broadway musicals/show tunes, mood/easy listening, big band/swing, classic rock/oldies,

country, bluegrass, folk, hymns/gospel, Latin/Spanish/salsa, rap/hip hop, blues/R&B, reggae, top 40/pop music, contemporary rock, indie/alternative rock, dance/club/electronic, and hard rock/heavy metal. In addition to providing a taste judgment for each item, respondents were also asked if they regularly listened to each macro-genre. The analyses that follows are based on a data frame in which people are coded as engaging the macro-genre if they report both liking and listening to it. The two mode network we will be working with is thus of dimensions has 2263 people and 20 genres. This defines an affiliation matrix A_{PG} of the same dimensions, with entries equal to 1 if person P likes *and* listens to macro-genre label G and zero otherwise.

4 Macro-genre Analysis Strategies: Three Variations

There are three standard analytic approaches in the quantitative sociology of taste. All begin from the same starting point, however, a rectangular (two-way, two-mode) matrix (A) with (usually) people ($P_1, P_2, P_3, \dots, P_N$) in the rows and cultural genres ($G_1, G_2, G_3, \dots, G_K$) in the columns. As noted at the outset, the *ways* of the matrix are the number of dimensions and the *modes* of the matrix are number of entity types for which data have been collected [Borgatti and Everett, 1997]. In the usual case, we have two-way, two-mode (people and genres) data [Lizardo, 2018]. The goal of the analysis is to learn some kind of *classification* of either the people, or the genres (or both) from the data. This setup is shown in figure 1.

In this (toy) example we have five people and four genres. Each cell in the A_{PG} matrix equals one when person P engages (e.g., likes, listens) genre G ; otherwise it is equal to zero. Reading across the rows we find each person's pattern of cultural choices [Peterson, 1983]. So we learn that person 1 engages genres 2 and 3, and person 3 engages all four genres; maybe they are an “omnivore”. Reading down each column we find the popularity of each genre, or the number of people who engage it. So we learn that genre 4 is engaged by four people (2, 3, 4, and 5) and genre 2 is only engaged in by two people (3 and 4). Maybe genre 4 is a “popular” genre, and genre 2 is a “niche” genre [Lizardo, 2018].

Of course, researchers usually want to go beyond describing the cultural choice patterns of individual people or the audience distribution of each genre. Instead, researchers use specific data reduction techniques to classify either the genres, the people or both. Let us consider each of these possibilities in turn.

4.1 Classifying Genres: Factor Analysis (FA)

One thing that we could do with the (real) data described earlier is to attempt to come up with a classification of musical genres. Recall that in these data we have 2263 people in the rows and 20 (musical) genres in the columns. One well-established and long-pedigreed technique allowing us to come up with clumps of the variables in a person by variables table (with a long history in psychometrics) is Factor Analysis (FA). This method uses the genres as variables, and relies on a factorization, (hence the name) of the correlation or covariance matrix between these variables to extract a smaller number of dimensions. We can then classify each genre based on their loading (e.g., correlation with) each dimension extracted. The output of such an exploratory factor analysis of the SSI 2012 data, featuring a four-factor solution, is shown in figure 2.

How do we interpret these pattern of results? One approach, for instance followed by Van Eijk [2001], is to inspect the factor loadings as indicating a classification of genres according to a higher order principle (e.g., logics or discourses); namely, a set of “meta-genres.” The results of the factor analysis is broadly consistent with such a hypothesis. For instance, factor 3 groups all the “rock and pop” related genres, and factor 1 groups together a related set of “Afro-pop” genres such as Rap, Reggae, and R&B. Together, they might constitute a discourse of “fun.”

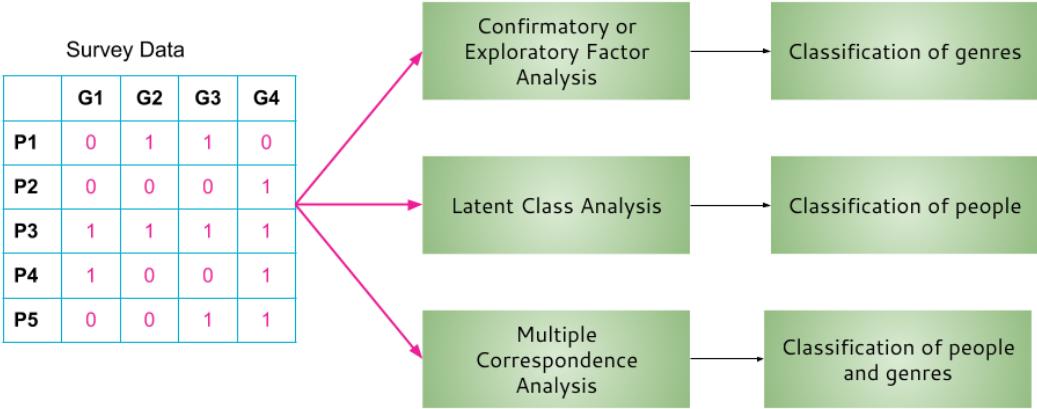


Figure 1: Three approaches to classifying survey data on cultural tastes.

Factor 2 groups together all the usual “highbrow” genres, such as Classical, Opera, and Jazz, reflecting a more austere “art” discourse. Finally, the genres loading highly on factor 4 such as Country, Bluegrass, and Folk may reflect a more down to earth “folk” discourse linked to roots authenticity Van Eijck [2001].

4.2 Classifying the People: Latent Class Analysis (LCA)

Another approach relies not on classifying the genres, but classifying the people. The idea here is to rely on the fact that each person produces a “response pattern” (a “pattern of cultural choice” in Richard Peterson’s 1983 classic locution). The pattern corresponding to a person (P_i) is thus the vector of zeros and ones corresponding to their row in the matrix shown in Figure 1. The idea is that we can then (probabilistically) clump people into classes based on how similar these patterns are. People with similar patterns end up in the same clump; people with dissimilar patterns end up in different clumps [Chan and Goldthorpe, 2007, Tampubolon, 2008]. One well-established and equally long-pedigreed technique that allows us to group people into clumps in a person by (categorical) variables table (with a long history in sociology, marketing and related fields) is Latent Class Analysis (LCA).

The main output from an LCA is an m -dimensional vector (where m is the number of latent classes requested) assigning each row i of the data matrix (in this case the people) a probability of belonging to the j^{th} latent class. Because memberships are mutually exclusive (each person can only belong to one latent class), these probabilities have to sum to 1.0. This means that, if the solution fits the data well, one of the probabilities in the vector will be much larger than the others (e.g., $m_{ij} > 0.90$). We can then assign each person to the class they have the largest probability of belonging to. Once we have assigned each person to a class, we can go back to the data, and see how the clumps of people assigned to different classes differ in their cultural choice patterns. The output of such an exploratory latent class analysis applied to the

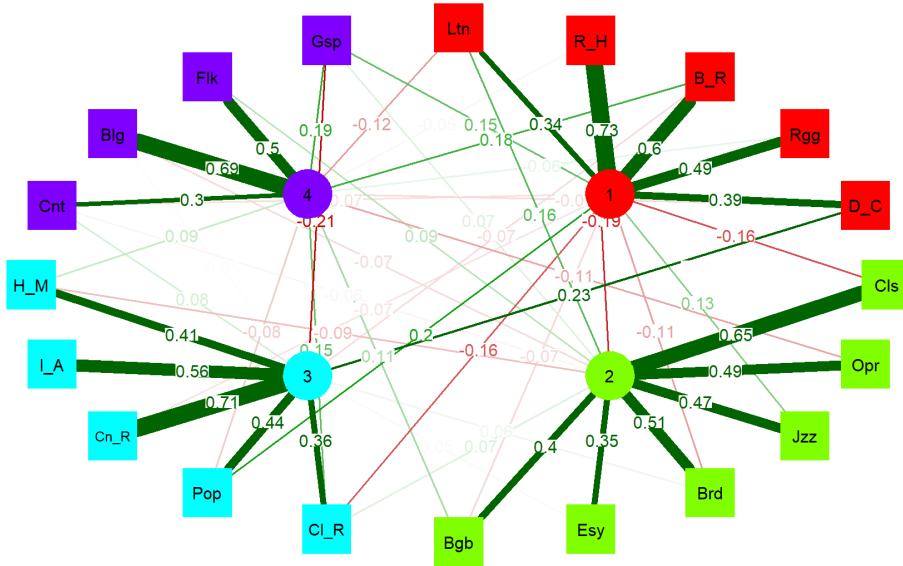


Figure 2: Point and line plot of exploratory factor analysis results (four-factor solution) of the musical taste data. Positive loadings between each factor (circles) and each of the genres (squares) shown as green edges. Negative loadings shown as red edges. Edge width is proportional to magnitude. Node labels should be read as follows. Factor 1 (red Nodes): Ltn (Latin/Spanish), R_H (Rap/Hip Hop), B_R (Blues and R & B), Rgg (Reggae), D_C (Dance/Club); Factor 2 (green nodes): Cls = Classical, Opr = Opera, Jzz = Jazz, Brd = Broadway Musicals, Esy = Easy Listening, Bgb = Big Band, Factor 3 (teal nodes): Cl.R = Classic Rock/Oldies, Pop = Pop/Top 40, Cn.R = Contemporary Rock, I_A = Indie/Alternative, H.M = Hard Rock/Heavy Metal; Factor 4 (purple nodes): Cnt = Country, Blg = Bluegrass, Flk = Folk, Gsp = Gospel.

same data as before, and featuring a six-class solution, is shown in figure 3.

How do we interpret these pattern of results? The usual approach in the sociology of taste, exemplified by Chan and Goldthorpe [2007] and Tampubolon [2008], is to use the conditional probabilities of engaging genre G given that a person belongs to class m (shown in the separate panels of figure 3) to baptize the classes with a given name. Thus, information from the genres (whether given by previous research, theory, or intuition) can help us make sense of what each of the people clumps means. For instance, we may find that one class of people (in this case class 4) has relatively high probabilities of engaging *all* genres. Maybe they are “omnivores” [Peterson and Kern, 1996]. Others, only engage have a high probability of engaging a couple of genres like Country and Classic Rock/Oldies (Class 2); maybe they are “univores” [Peterson, 1992]. Some univores (like Class 1) just engage very fancy, traditionally high-status genres like Classical; others (like Class 5) specialize in “Afro-Pop” genres like Rap/Hip Hop and R&B, while others (like Class 6) gravitate toward Contemporary Rock, Pop, Indie and Alternative music.

We could further investigate the demographic composition of each class, by for instance, fitting a regression model for categorical response variables with multiple modalities (e.g., like the Multinomial Logit model) in the “three-step” approach [Bakk et al., 2013], or doing everything in one-step (letting the classes be affected by exogenous covariates, as with Tampubolon [2008]). We could then examine hypotheses about whether omnivores are young and educated, highbrow univores are old, Country/Oldies univores live mostly in the American South, Afro-Pop univores



Figure 3: Stacked bar plot showing the proportion of people (shown in the x-axis) engaging ("yes" in dark blue) and not engaging ("no" in light blue) each genre (shown in the y-axis) by latent class assignment (plot panels).

are more likely to be young and non-white and so forth (fancier approaches would do the class assignment and the regression model in a more efficient single step, but it is the same idea).

4.3 Classifying the People and the Genres: Multiple Correspondence Analysis (MCA)

Finally, we may be interested in a technique that allows for the simultaneous classification of the people and the genres. The go-to technique for this job is Multiple Correspondence Analysis (MCA). Correspondence Analysis and related techniques have as old and storied histories as FA and LCA, but for a long time remained a well-kept secret (for English speakers) ensconced in multi-volume French mathematical monographs written by Jean-Paul Benzécri and collaborators.

Today the situation is very different, with MCA now introduced and extended in a variety of anglophone monographs and collections [Le Roux and Rouanet, 2004, Greenacre and Blasius, 2006], and algorithms implementing the basic mathematical ideas available in the standard statistical packages [Lê et al., 2008]. Because of this, and very much ironically, while MCA (and not regression FA, or LCA) could be considered one of the *first* techniques that reinvigorated the modern empirical study of taste in sociology (figuring prominently in Pierre Bourdieu's [1984] field-defining classic *Distinction*), it has not been until recent that the method has begun to be used widely [see Flemmen et al., 2018, Roose et al., 2012, Savage and Gayo, 2011, among many others].

How does MCA work the magic of providing us with a simultaneous classification of people and genres? The trick, revealed in Figure 4 using our running toy example, is to transform the original two-mode data matrix to something called an indicator matrix. In the indicator matrix, there is a column not for each variable (like in the original two-mode data) but for each *category*

Survey Data

	G1	G2	G3	G4
P1	0	1	1	0
P2	0	0	0	1
P3	1	1	1	1
P4	1	0	0	1
P5	0	0	1	1

Indicator Matrix

	G1 _Y	G1 _N	G2 _Y	G3 _N	G3 _Y	G3 _N	G4 _Y	G4 _N
P1	0	1	1	0	1	0	0	1
P2	0	1	0	1	0	1	1	0
P3	1	0	1	0	1	0	1	0
P4	1	0	0	1	0	1	1	0
P5	0	1	0	1	1	0	1	0

Figure 4: Transformation of original data matrix of people by genres, to an indicator matrix of people by genre-modality.

of each variable (these are usually called “modalities” in MCA-speak [Le Roux and Rouanet, 2004]). In our toy case (and our real data), each variable has two categories. So we go from a 5×4 matrix to a 5×8 one (see Figure 4). Each genre G column splits into two columns, G_Y and G_N and we put a one in the corresponding cell of the matrix under column G_Y if the individual engages (e.g., likes, listens) to the genre. If the individual does not engage the genre, we put a one under the G_N .

The singular value decomposition (a kind of “factor analysis for matrices”) of a suitable transformed version (each cell entry weighted by the row and column marginal sum) of the indicator matrix yields scores for each row and each column (both the “yes” and “no” answers to engaging the genre in our toy example) along a number of “dimensions” (also called “axes” or “principal components”). Like in FA the dimensions are ordered so that the first captures the closest representation of the original matrix (in the least squares sense), the second captures the best representation of the stuff that is left over, and so forth. The scores on each dimension are orthogonal and thus can be plotted in a Euclidean space (e.g., of two dimensions). This can be seen as a low dimensional representation of the original multidimensional matrix object. The clumps of people and genres thus emerge as clusters of (same-kind) entities that end up close together in this low-dimensional space. The capacity to visualize both people and especially the genres “social space” is one of the main appeals of the method [Greenacre and Blasius, 2006], as evident in Bourdieu’s [1984] classic work. These scores are the main thing that MCA learns from the data. The output of such an multiple correspondence analysis of the data, featuring a four-factor solution, is shown in figure 5.

How do we interpret these pattern of results? We can follow some of (many) recent papers that have used MCA to understand the link between people and genres [e.g., Roose et al., 2012]. As noted we can see the modalities that clump together along a given dimension as telling us something about what that dimension represents. So, looking at the horizontal (first) dimension

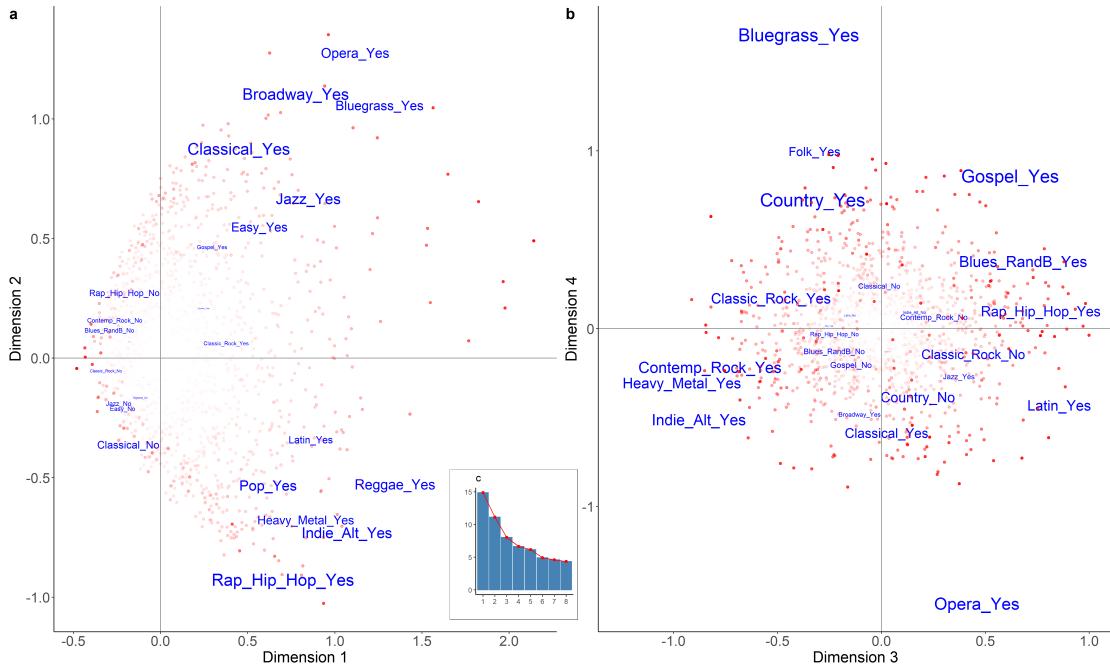


Figure 5

of the left-hand side biplot (panel a), it is easy to see that all the "yes" modalities are to left of that horizontal dimension (indicating values greater than zero) and all the "no" modalities are to the right. So the main thing that the MCA picks up is a distinction between people who engage the genres and people who do not (actives versus abstainers). Note that the both people (the red dots forming the "cloud of individuals") and the genre modalities are distributed in a classic "horseshoe" pattern in this space, indicating that this first dimension is *ordinal*, separating people who engage many genres from those who engage none. This is a finding that has been replicated repeatedly in many MCA analyses done on a variety of cultural taste and participation data across dozens of countries.

What does the second dimension tell us? Well, it seems like (looking at panel b) here the usual distinction between fancy, high-status genres (with high positive scores at the top) and popular, industry or just plain fun genres (with high negative scores at the bottom) reasserts itself. So "net" of engagement, we find that engagers are partitioned across a highbrow/popular divide [Roose et al., 2012, Flemmen et al., 2018]. Looking at genre clumps (really active engagement with genres clumps) in dimensions 3 and 4 reveals further insights. In fact, looking at how the genres and the people are distributed in this space, we can see that net of the quantitative distinction between engagers and non-engagers, and the broad binary genre distinction between "traditional and popular" what reappears is the van Eijkian pattern differentiating of genres into four discourses or meta-genres: The same genre classification revealed by FA (see Figure 2). Categories indicating engagement with (predominantly white-performer-dominated) rock genres clump together, engagement with Folk and Country genres clumps together, traditionally fancy genres clump together, and "Afro-pop" pop genres clump together. People are evenly distributed (in a nice circle) across all four quadrants of this space, indicating that these logics may be a powerful organizing principle of the "musical taste field" [Savage and Gayo, 2011] in the contemporary U.S.

More advanced applications of MCA would be beyond eyeballing the clumps and do some other things [Le Roux and Rouanet, 2004]. For instance, since MCA produced scores along a multidimensional Euclidean space, this means that *distances* among each pair of elements of the same kind (people or genre engagement modalities) are defined. We can thus compute such distances and produce data (not eye) driven clumps using hierarchical clustering (with Ward's criterion) [Bécue-Bertaut and Pagès, 2008], an approach that was first developed by Benzécri himself [Le Roux and Rouanet, 2004]. We could also “project” categories (modalities) corresponding to people’s demographic characteristics (e.g., by computing the average score in each dimension for people with e.g., a college degree, professional occupation, and so forth). We could even engage standard hypothesis testing, by for instance, figuring out if the mean score in Dimension 3 is statistically different for people who identify with different ethnoracial categories [Flemmen et al., 2018].

5 Getting out of the (Macro) Genre Bubble

There is no question that the three approaches so far discussed, using to classify genres, classify people or simultaneously classify people and genres are the workhorses of the quantitative empirical study of taste today. All, by the same token, are vulnerable to the “genre” critique outlined in the introduction. For instance, both FA and MCA suggest that engagement with “Opera” or “Classical” is different from engagement with “Rap” or “Country.” These (macro) genres belong to different universes, discourses or what have you [Van Eijck, 2001]. But a critic of the genre concept can point to other types of qualitative and historical evidence, collected on both people and genres that puts into question this conclusion. All provide fodder for a version of the micro-genre or genre differentiation critique of traditional quantitative empirical work on the sociology of taste that does not collect sufficiently fine-grained information on cultural preferences.

For instance, are we certain *all* types of classical music are “opposed” to or divergent from engagement with Country or Hip Hop? Aren’t there versions of Indie/Alt Rock that are as “high-brow” as Classical music? Aren’t there kinds of Country music that appeal to highly educated audiences and thus to people who might also listen to musicals? Overall, traditional techniques have to work with the (broad) genre categories included in the survey and are powerless to deal with these (reasonable) objections. Note that *what* technique you use (FA, LCA, MCA) does not really matter; from the micro-genre critique perspective, it is garbage in and garbage out. MCA will produce an “opposition” between macro-genres and even results that are consistent with techniques (such as FA) that are usually posed as enemies of MCA. The problem is that poor MCA (or LCA or any other method) only has the macro-genre based categories to work with. Tasked with finding patterns using these data, the algorithm delivers, but what it delivers is, from the micro-genre perspective, are misleading or only partially true results.

5.1 Relationality to the Rescue

I propose that exploiting *relationality* hidden in the usual survey data collecting information on people and genres can help us make headway on the micro-genre critique issue [Lizardo, 2018]. As noted at the outset, this requires that we stop looking at the usual survey data that forms the bulk of empirical material in the quantitative study of taste as “survey data” (e.g., data collected on the characteristics of individuals in the form of “variables”). This is a presumption that is baked into traditional techniques (like FA or LCA) that fit statistical models to such data. It is also embedded in “relational” techniques like MCA which, while not steeped in the preoccupations of standard statistical inference (although it could be), divert our attention to the indirect relations between macro-genres (or between people), usually conceptualized as distances

in a social space. This kind of indirect relationality is fine and dandy (I will exploit it in what follows) but as already noted, all the relationality built into MCA is powerless, if what is fed *into* the process is the same old, limited, overextended macro-genre categories.

5.2 The Link Clustering Approach

I address the micro-genre issue using a technique for overlapping community detection called link clustering [Ahn et al., 2010]. The basic idea is simple and is illustrated in Figure 6a. Breaking with approaches that attempt to cluster people and genres by focusing on the people or the genres, we take the person-to-genre *link* as the unit of analysis and focus our classification energies on those. Classifying the links gives (by definition) a classification of the entities at the end of each link (people and genres) for free. Not only that because most people choose multiple genres and all genres “choose” multiple people, this classification guarantees that genres get assigned to multiple *overlapping* clusters, at the same time breaking the macro-genres into more relationally focused clumps. This allows us to make headway on both horns the micro-genre critique (overlapping categories and internal differentiation) at once.

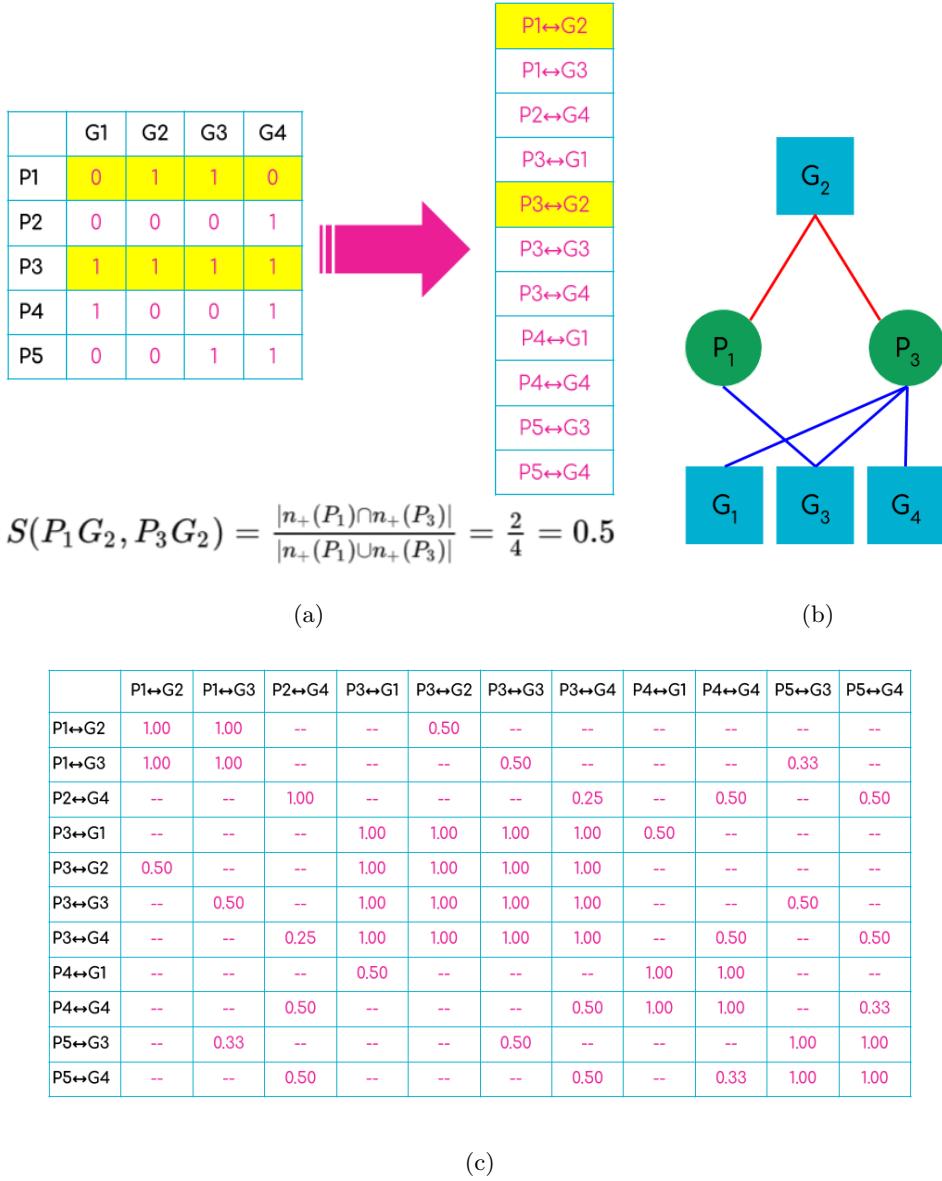
How does this work? As shown in Figure 6a, once we string out the original two-mode person-by-network into an edge list (where the cases are the person to genre pairs that are marked by a “1” in the original matrix), it is possible to ask: “How similar is one edge to another?” We can answer that question as follows. The similarity of one person-to-genre edge to another one, when both edges point to the same genre [Ahn et al., 2010], should be proportional to the overlap between the “cultural ego networks” of the two people at the tail end of the edges. As Lizardo [2014] defines it, for any respondent i , in a survey on cultural taste, the cultural ego network is simply the set of genres $\{G_1, G_2, \dots, G_k\}$ chosen by the person.

Let us go back to our running toy example. Suppose we wanted to calculate the similarity between two person-to-genre links in the two-mode network (P_1G_2 and P_3G_2), where the links join two different people (P_1 and P_2) to the same macro-genre label (G_2). In Figure 6a, these two links are highlighted in yellow the “strung out” edgelist and are shown in red in the corresponding bipartite subgraph shown in Figure 6b. The similarity (S) between the two links is then given by:

$$S(P_1G_2, P_3G_2) = \frac{n_+(P_1) \cap n_+(P_3)}{n_+(P_1) \cup n_+(P_3)} \quad (1)$$

Where $n_+(P_1)$ is the cultural ego network of Person 1 and $n_+(P_3)$ is the cultural ego network of Person 3. The formula thus says that the similarity between two person-to-genre links that share a genre (in this case, the similarity between the P_1G_2 edge and the P_3G_2 edge) is given by the intersection of Person 1’s and Person 3’s cultural ego network divided by their union (also known as Jaccard’s coefficient). When the cultural ego networks of two people are the same then $S = 1$, when they are completely disjoint (share no genres, meaning the intersection is the empty set), then $S = 0$. All other cases return a number between zero and one ($0 > S < 1$). In the toy example case, P_1 and P_3 consume two genres out of the four possible ones they could consume in common, resulting in an edge similarity score of $\frac{2}{4} = 0.5$.

If we do that for all pairs of person-to-culture genres that share one genre node in common, we end up with the two-way, one-mode (now the only mode is edges) similarity matrix shown in Figure 6c. Note two features of the similarity matrix. First, by definition two person-to-genre links that go from different genres to the same person are maximally similar (as the overlap is 1.0). This means that only links that go from different people to the same genre exhibit overlap variation, and this is driven (“reflectively” [Lizardo, 2018]) by the overlap between the neighborhoods of the other mode (people). Hierarchical clustering of this matrix would thus



yield *link communities* [Ahn et al., 2010, Kalinka and Tomancak, 2011], in which both people, but most crucially genres will be assigned to different clumps.

For instance, it could be that, due to their relatively low overlap (0.33), the link joining Person 1 and Genre 3, and the links joining Person 5 and Genre 3 get assigned to different communities. This takes the nodes labeled as “Genre 3” (in the macro sense) and puts them in different clumps, creating two variants of Genre 3 (hierarchically nested within the larger version). The idea then, is that the macro-genre nodes (e.g., “Classical,”) assigned to different clumps represent micro-variations of the macro-genre label that differ primarily in a relational or structural way; different “types” of “Classical” are different because their audiences make distinct choices *with respect* to the other genres in the survey.

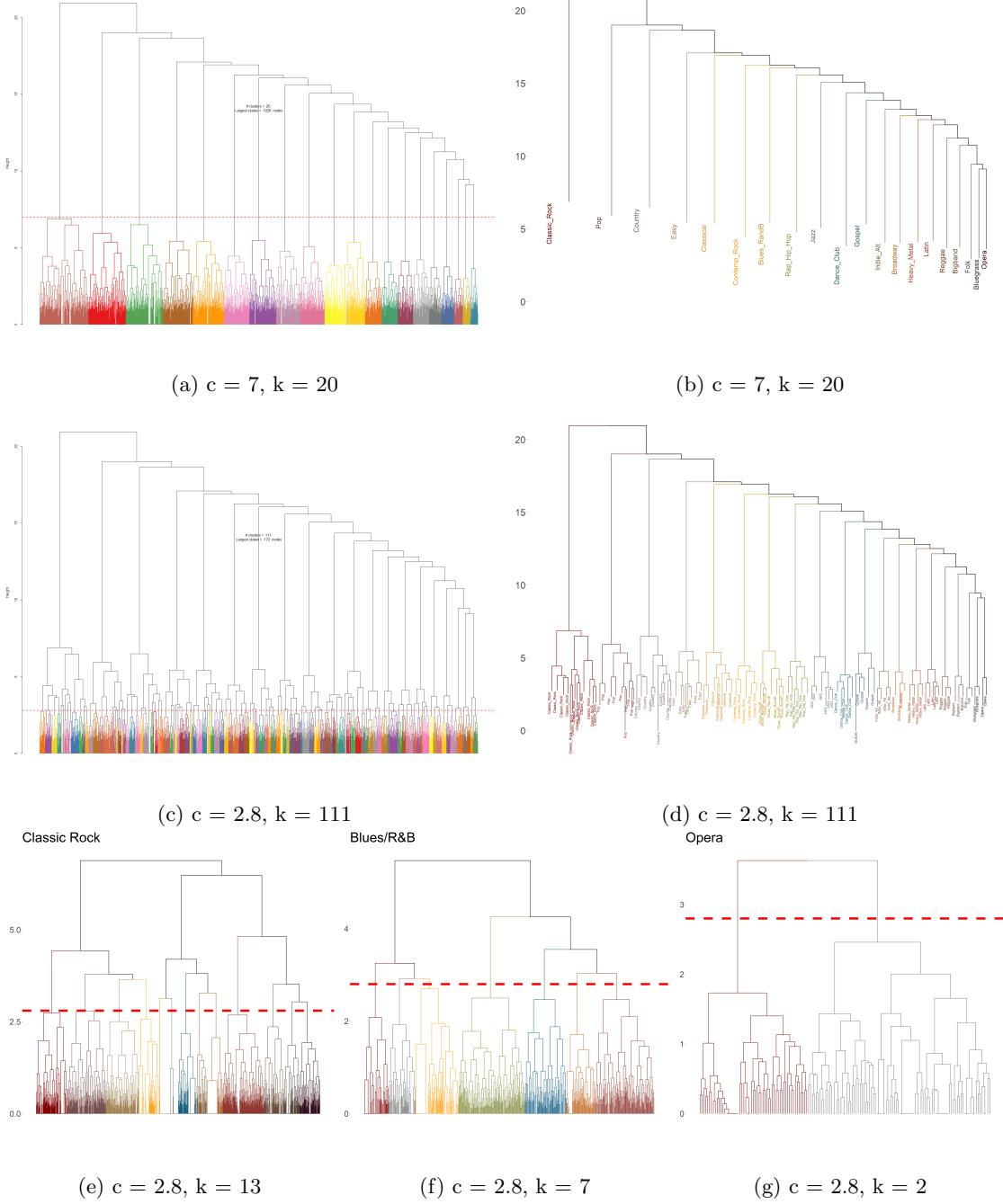


Figure 7

The resulting partition has two interesting (and desirable) properties. First, the number of genre communities that *people* end up belonging to is deterministic, and it is given by the number of macro-genre labels that they initially chose. Thus, link clustering preserves the cultural

ego network degrees (omnivorousness by volume) of the people mode. Second, the number of micro-genres into which the macro-genres is split is *not* deterministic. Instead, it is data-driven (discovered or learned) and cannot be pre-specified in advance. It is a function of relational information implicit in the overlap structure of the cultural ego networks of people in the data. Thus, link community detection of the person-to-genre network implies going from a situation in which we start with a network featuring a relatively small number of macro-genres, and end up with an enlarged two-mode network with the same number of nodes in one mode (the people) but many more nodes in the other mode (the micro-genres). How many micro-genres emerge is up to the analyst as it depends on where we “cut” the resulting clustering dendrogram.

5.3 Discovering Micro-genre Communities in Real Data

Let us see how this process works in the real data that we have subjected to standard Factor, Latent Class, and Multiple Correspondence Analysis. We will see that the link clustering can uncover valid micro-genres.

Recall that the data feature 2,263 people choosing up to 20 macro-genre labels. When strung out as an edgelist, this results in 9,216 person-to-genre connections in the data, the resulting $9216 \times 9,216$ matrix, containing Jaccard similarities among person-to-genre links sharing a node in either of the two modes, is then the input to a agglomerative hierarchical clustering algorithm. The hierarchical clustering process proceeds as follows. Initially, each person-to-genre link is initially assigned to its own community. Then, in the second time step, the pairs of links with largest Jaccard similarities are put in the same clump. This continues at each time step, where pairs of links with the largest similarity are chosen and their respective communities are merged. This process is repeated until all links belong to a single cluster. The history of the clustering process is then stored in a dendrogram, which encodes all the information of the hierarchical organization of the genre communities. The height of the dendrogram gives information about the strength of the genre communities. The highest levels reproduce the original macro-genre levels, while the lower levels uncover more focused micro-genres embedded within them. This is shown in Figure 7.

Figure 7a shows that when we cut the dendrogram at a high level (e.g., $c = 7$), we reproduce the original twenty macro-genres we began with as the “discovered” link communities. Note that the agglomerative link clustering procedure arranges the macro-genre levels roughly by their original popularity (number of person-to-genre links). From left to right, these are Classic Rock/Oldies, Pop/Top 40, Country, Classical, Easy Listening, Blues/R&B, Contemporary Rock, Rap/Hip Hop, Jazz, Gospel, Dance/Club, Indie Alternative, Latin, Broadway/Musicals, Heavy Metal/Hard Rock, Reggae, Big Band, Folk, Bluegrass, and Opera.

As shown in Figure 7c, micro-genre communities are produced by cutting the link clustering dendrogram at a lower level. Choosing a cut value of $c = 2.8$ results in 111 micro-genre communities extracted from the original twenty macro-genre labels. Data scientists wring their hand about how to choose a cut value when performing a cluster analysis. One advantage of the link clustering approach is that because micro-genres are strictly nested within the original macro-genres, we always know what we are doing. Choosing a smaller cut value produces a finer-grained micro-genres (perhaps at the expense of analytic tractability and interpretability) and choosing a higher value returns us to broader genre communities closer to the original macro-genre labels we began with (with the limiting case as shown in 7a being the original macro genre labels themselves). At any cut value c , the more popular macro-genres produce a larger number of micro-genre communities, while the less popular one produce a lesser number. Thus, as shown in Figure 7e, at $c = 2.8$, the macro-genre “Classic Rock/Oldies” (the leftmost set of branches in the dendrogram in Figure 7a) is split (at the point at which the branches intersect the red dashed

cut line) into thirteen distinct micro-variations, while the macro-genre “Opera” (the rightmost set of branches in the dendrogram in Figure 7a) is split into two micro-genre communities.

Each cut point in the dendrogram will have a micro-genre size distribution associated with it. Very low cut points return a micro-genre size distribution dominated by many micro-genres with tiny ($N < 10$) or trivial $N = 2$ audiences. There is no magical “right” answer here, but 111 is sufficiently fine-grained to showcase the analytic advantages of the micro-genre approach. The resulting size distribution is shown as Figure 8. The main panel shows the micro-genre communities chosen by 75 or more people, and the inset shows the full size distribution. micro-genre communities are named according to the original macro-genre label, followed by a number indicating the audience size for that micro-genre (e.g., “Classical_1,” is the largest Classical micro-genre, “Classical_2” the second largest and so forth). As the figure shows, the largest micro-genre communities in the data pertain to micro-variants of Pop, Jazz, Classic Rock, Rap, and Broadway musicals (the largest micro-genre at $N = 171$).

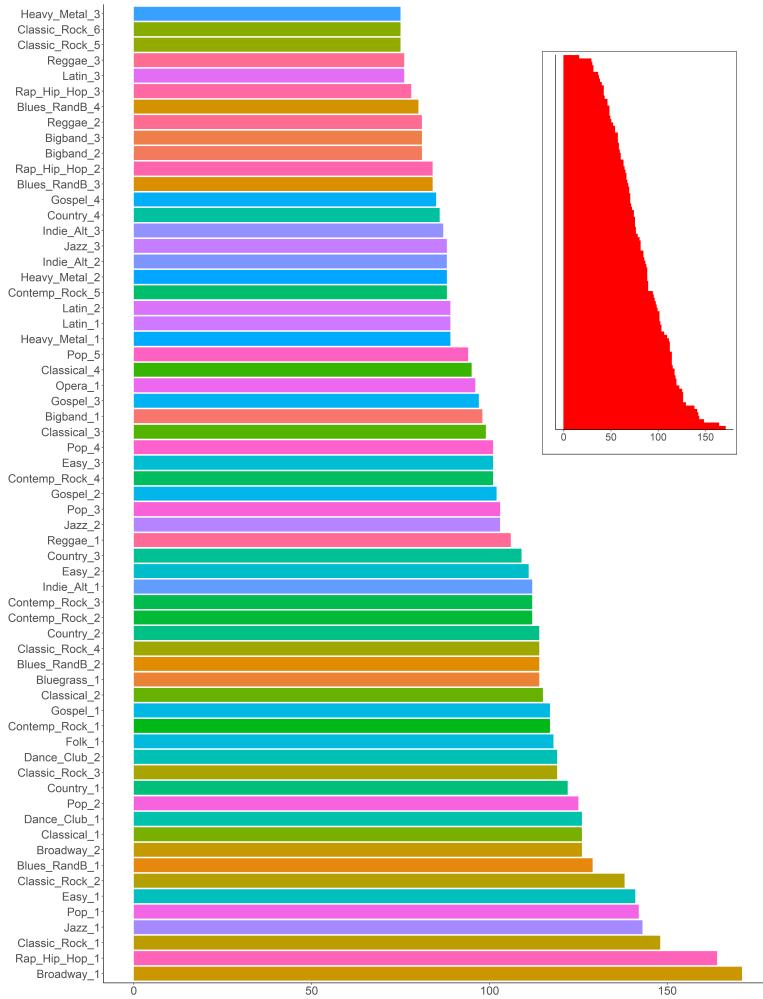


Figure 8

5.4 Doing Old Things with Micro-Genres

The link clustering procedure yields a “new” data set out of the old. Instead, of 2263 people choosing 20 vague macro-genre labels, we now have the same 2263 people, but their choices have been distilled and focused on 111 micro-genres. Since this is still a plain old survey data set, with people in the rows and micro-genres in the columns, we can perform some of the same old tricks (Factor Analysis, Latent Class Analysis, Multiple Correspondence Analysis) as before. But this time, the results will be more revealing, because we have honed in a bit closer upon the actual object that people made a taste judgment on and engage in their everyday musical listening practices. In addition, applying old techniques in this new data is likely to yield insights (e.g., combinations and classifications of micro-genres) that will not simply reproduce the standard findings of the past (e.g., clumping of vague macro-genre labels into even vaguer “discourses” like highbrow, pop, folk, and the like).

To show this, Figure 9 displays what happens when we subject the new 2263×111 matrix to Multiple Correspondence Analysis. Because now we have 111 points (positive “yes” modalities) to worry about, I used hierarchical clustering on the MCA dimensions to clump the micro-genre points in the MCA space. Labels rendered in the same color are in the same cluster, suggesting that people tend to choose them together.

Note that the resulting picture (compared to Figure 5) is now richer and more complex. Micro-genres trade focus for “legibility” at least when it comes to vague macro-genre clumps. However, this does not mean that there no interpretable results. On the contrary, it just means that there are *more* interpretable results. For instance, take the clump of genres that dominates the first (horizontal) dimension of the MCA space micro-genre classification. Here, we see all the macro-genre labels represented in close proximity (indicating people tend to choose all these micro-genres in tandem). The micro-genres in this region of the space (as given by their low number designation) tend to have the largest audiences of their kind (with signal exceptions such as Rap, Country, and Jazz). This, therefore indicates micro-genres that go together due to a rather indiscriminate mixing of styles, indicative of a *type* (but not the only type) of “omnivorousness.”

Other regions display more focused clumpings of micro-genre styles, but ones that are not easily interpretable using the usual macro-categories. Toward the top of the upper panel, we see a combination of genres usual classified as “contemporary” or “youth” but also containing Classic Rock. Immediately below this clump, we see a three-micro genre cluster featuring both Classic and contemporary rock styles. In the lower panel of the figure, showing the space formed by the third and fourth dimensions, we see other interesting micro-genre mixtures. Toward the top, a micro-genre cluster composed of traditional “high status” genres like Classical music, but also featuring two country micro-genres and classic rock. As we will see below, this Classical micro-genre (“Classical_3”), tends to feature individuals who do not mind making many choices.

6 Demographic Profiles of Micro-genre Communities

We could continue trying to derive novel “discourses” of taste by looking at micro-genre clumps in the MCA space. Suffice to say that these are probably not going to look like the ones that have usually been noted in the literature (although some might). However, as noted, the core feature of the link clustering algorithm for discovering genre communities is that takes somewhat fuzzy and vaguely defined macro-genre labels (e.g., “Classical,” “Classic Rocks/Oldies,” “Country,” or “Pop”) and decomposes them into more focused micro-genre communities. The remaining question is whether these micro-genres are valid, in terms of attracting systematically different types of people. To show this, I follow the venerable principle of audience segmentation to verify the validity of the procedure to discover interpretable micro-genres [Peterson, 1992].



Figure 9: Multiple Correspondence Analysis of Micro-Genre Data. Upper panel shows the Euclidean space formed by the first two dimensions, and the lower panel shows the space corresponding to the third and fourth dimensions. Micro-genres are hierarchically clustered (using Ward's criterion) into 13 clusters (see inset) based on their proximity in the four-dimensional MCA space. Micro-genre labels in the plot belonging to the same cluster are rendered in the same color.

6.1 Classical

Figure 10 shows the demographic profiles for the eight “Classical” micro-genres discovered by the link clustering procedure. Demographic profiles are depicted as the predicted probabilities

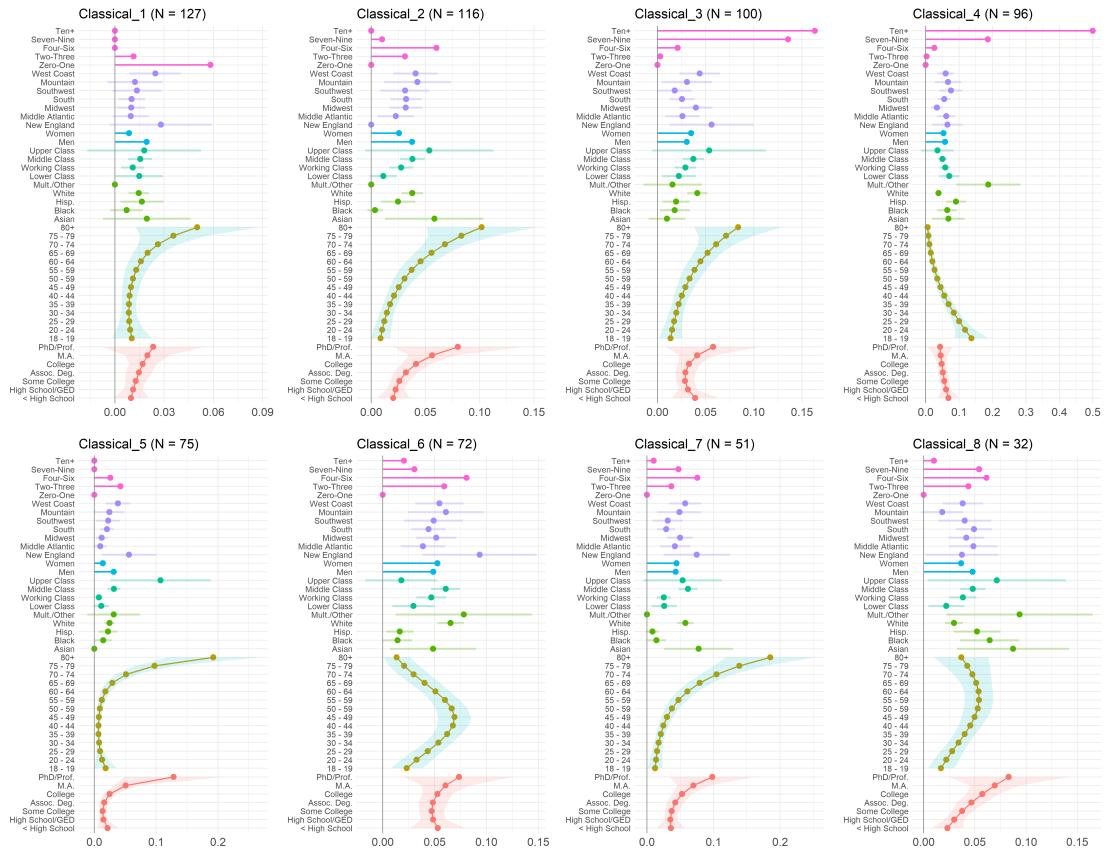


Figure 10

(and associated confidence intervals around these estimates) obtained from a logistic regression of the relevant sociodemographic characteristic (in this case, education (red), age (, ethnoracial identification (green), subjective class identification (teal), gender (blue), and region of residence (purple), along with a categorical breakdown of the total number of genres chosen (pink)) against the log odds of having chosen that particular micro-genre [King et al., 2000].

The first Classical micro-genre ($N = 127$), picks up a micro-genre that appeals to the Classical music “univore.” This person is likely to be old, likely to be a man living in a New England state; some of us have perhaps met this person and can imagine what version of Classical music this is (likely to be the most stereotyped around the established canon [Kremp, 2010]). The second Classical micro-genre ($N = 116$) seems like what is seen as the prototypical consumer of Classical music [Lizardo and Skiles, 2016]. This person is likely to be older, hold an advanced degree, be white and identify as middle class. Interestingly, other micro-genre variations of Classical picked by the analysis (3, 5, 7, and 8) capture variations of Classical music that also appeal to older, highly educated people; interestingly, these other Classical musics differ from the last variant with respect to the other markers of social position. The third Classical micro-genre (($N = 100$) is characterized by extreme omnivorousness by volume; the fifth micro-genre ($N = 75$) is distinguished by the fact that the people who choose it are men who are also likely to identify as “Upper Class” (a rare designation in the U.S. case). The seventh micro-genre ($N = 51$) is inordinately likely to be consumed by Asian respondents, while the eighth, and

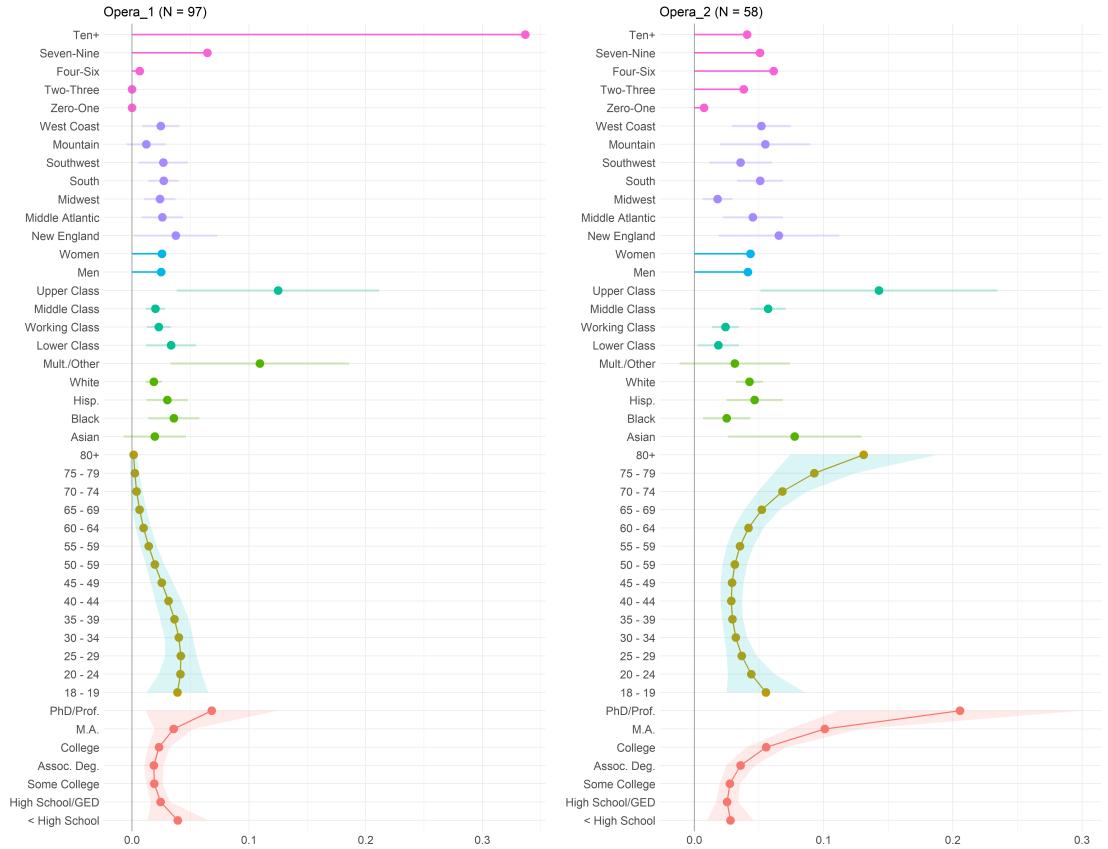


Figure 11

smallest, micro-genre ($N = 32$) is more likely attract nonwhite respondents. The link clustering approach is thus able to recover micro-genre communities that appeal to ethnoracial minorities even in the case of a stereotypically “white” macro-genre label such as Classical [Lizardo and Skiles, 2016]. The remaining Classical micro-genres have demographic profiles that deviate from the stereotypical Classical music consumer. People attracted to the fourth Classical micro-genre ($N = 96$) are young omnivores, while those choosing the sixth variation ($N = 72$) are middle-aged New Englanders who are either white or multiracial.

6.2 Opera

Not all macro-genre labels are decomposed into a multiplicity of micro-genres. Take the case of Opera, shown in Figure 11. The link clustering procedure discovers only two micro-genre variations in this case. A version chosen by omnivorous respondents who choose a multiplicity of other genres (and who are also younger, identify as upper class and are likely to be multiracial) and a micro-genre that looks a look like the stereotype; older, highly educated, upper/middle-class person. This suggests that, save for the version chosen by omnivores, the macro-genre label “Opera” already gets at a relationally focused micro-genre. A situation very different from “Classical.” This puts into question the common strategy of treating these two macro-genre labels as equivalent indicators of an overall “highbrow” dispositions. As we have seen, while

some versions of Classical do fit that bill in terms of their sociodemographic profile, others certainly do not.

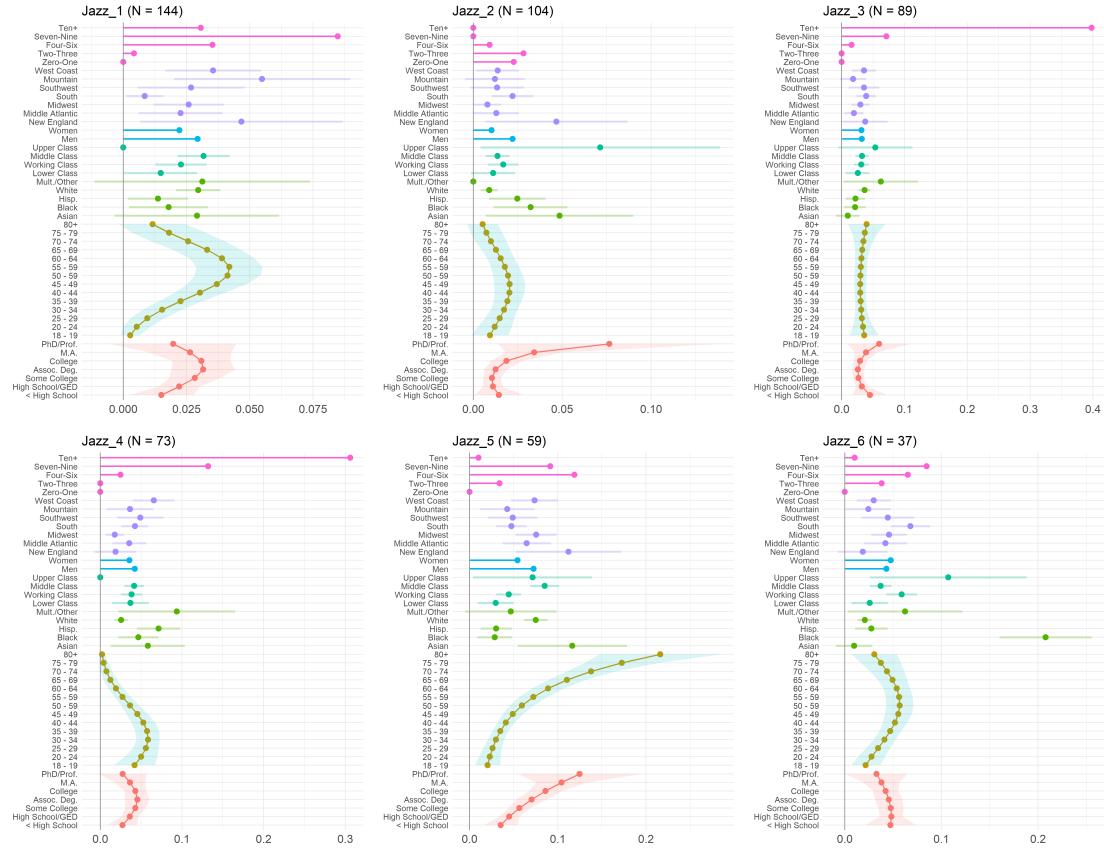


Figure 12

6.3 Jazz

The case of Jazz is an instructive one for the link clustering procedure, given the complex historical trajectory of the genre. It began as an “artistic” (expert performer-driven and focused) genre in the Black art world in the early 20th century, diffusing into both Black middle and working-class consumer communities and white aesthetic (both performer and consumer-driven) circles by mid-century, and then legitimized as a high art form in conservatories later on [see Lopes, 2002]. Accordingly, “Jazz” has always been suspected of being the ultimate multivocal macro-genre labels, condensing complex dynamics of ethnoracial, social class, and generational status. Whatever it may be, “Jazz” is simply not a single kind of thing and thus represents one of the linchpins of the macro-genre label critique.

The link clustering analysis shown in Figure 12, agrees; but with a caveat. Split into relationally structured micro-genres, Jazz is not a single thing, but neither is it an unruly multiplicity of unrelated things. Instead, it seems to be a small collection, summarizing in the synchrony of respondent choices in 2012, the complex history alluded to. In that respect, the micro-genre approach can tell the difference between the Jazz variant appealing to mainly white, older, highly

educated New Englanders (version 5, $N = 59$), perhaps the aestheticized and institutionalized conservatory version, and the one whose primary marker is appeal to middle-aged Black respondents who mainly reside in the American south, but otherwise lacking a strong education gradient (version 6, $N = 37$). Interestingly, there is also a version of Jazz (version 3, $N = 104$) that has a strong education gradient and a relatively weak relationship with age, that primarily appeals to non-white audiences. This version is also the most gendered, and univorous being more likely to be engaged in by men who do not choose a wide variety of other genres. Finally, in addition to two omnivorous versions (versions 3, 4) we discover one ultimate “middlebrow” version of Jazz, largest of micro-genre communities identified (version 1, $N = 144$); this genre appeals to middle-aged respondents with equally middling education levels who identify as middle class; Jazz as “cultural goodwill” [Bourdieu, 1984].

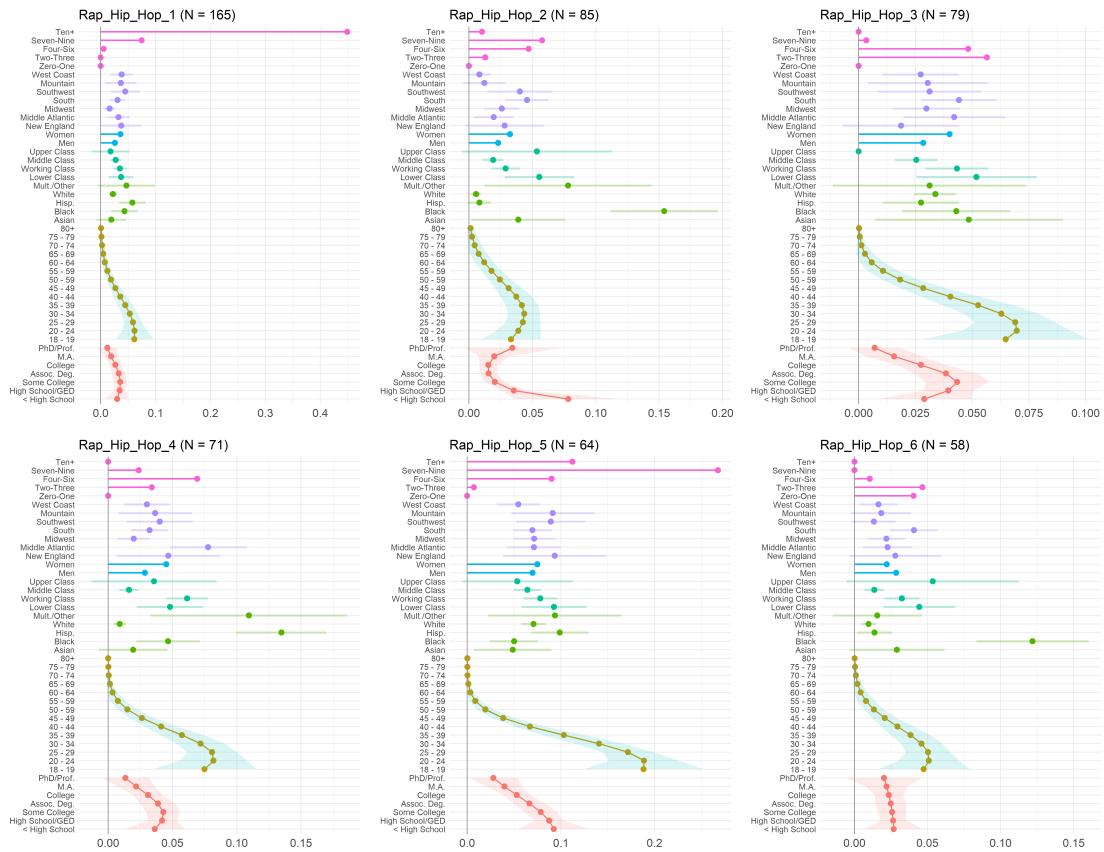


Figure 13

6.4 Rap/Hip Hop

While not having as long a career as Jazz or the Blues, Rap/Hip Hop has also endured vicissitudes as its sociological genre bases have been transformed from its inception as a “scene” genre in the late 1970s, followed by a slow proliferation into geographically segmented scenes in the 1980s and 1990s, its slow and hesitant uptake by industry during the same period, and its emergence as a full-fledged industry-backed “pop” genre in the 21st century. Throughout, however, the strong

association of Rap with young, Black, male performers and associated audiences (a situation that is in the process of being undone in the last decade or so with the emergence of powerful, iconic women rap performers with a strong foothold in industry) is one that is hard to shake [Lizardo and Skiles, 2016].

The micro-genre analysis, shown in Figure 13 discovers six variations, producing results that both support and contravene those expectations. Regarding the latter, none of the micro-variations of rap are strongly gendered, suggesting that this is not a powerful principle of audience segmentation here. Age is such a powerful principle, with rap audiences. Age/generation is such a powerful principle, with rap audiences across all six micro-genres being concentrated among young people; the only difference is the strength of the age effect, being strongest for variations 1, 2, and 3 ($N = 164$, $N = 84$, and $N = 78$, respectively), and weaker for 4, 5, and 6 ($N = 70$, $N = 63$, and $N = 57$, respectively). Variation 1, featuring the largest audience size, also features the least ethnoracial differentiation (save for the “most omnivorous” variation 4), and a penchant to combine with other genres; this suggests that this variation picks up the “industry” version of Hip Hop that presents itself today as radio-friendly pop music. This seems to stand opposed to variation 6; this is the smallest rap micro-genre, it has the strongest tilt toward Black, young audiences, and features the lowest propensity to be combined with other genres. Variation 5 is just like this last one, except that it features a strong tilt toward audiences without a college degree and a higher propensity to mix with other genres.

One issue, in the wake of the emergence of Rap and Hip Hop as general pop culture genres, is the extent to which some micro-variations appeal to non-Black audiences. The link clustering procedure can tell the difference, producing a micro-genre variation (2) that is much less likely to appeal to Black audiences and more likely to appeal to those who identify as “Hispanic” or “multiracial” (perhaps picking up ethnolinguistic variants of Hip Hop). In the same way, Neither variation 3 or 4 (along with 1 as already discussed) exhibit as strong a tilt toward Black audiences as 2, 5, and 6. Suggesting that audience segmentation along ethnoracial identity is strong within Rap/Hip Hop micro-genres despite the relatively weaker levels of segmentation along education or generational lines.

6.5 Gospel/Religious

The macro-genre fuzzy label “Gospel/Religious” is a lot like Jazz, in the sense that analysts know that it mixes micro-genre variations with distinct racialized audiences. As shown in Figure 14, the link clustering approach breaks this confounding, discovering three variations of Gospel primarily engaged in by Black people (1, 3 and 5) and one that primarily appeal to white people (4, $N = 86$), all of whom are likely to live the American South. Two of the Black gospel communities are mainly differentiated by age, with version 3 to working-class adults, and version 4 mainly appealing to middle class older people; both Black and white Gospel/Religious micro-genres are more likely to be engaged in by women. Version 1 is a Black Gospel univore community, and version 5 $N = 103$ is the omnivore micro-genre variation (which exists for all genres).

6.6 Latin/Spanish/Salsa

Even when macro-genre labels point to styles with some level of presumed audience homogeneity around specific markers (e.g., ethnoracial status), the link micro-genre community approach can discover variations that differ in terms of other social markers. The (already hybrid) macro-genre label “Latin/Spanish/Salsa” is a good case in point (see Figure 15). This macro-genre splits into five variations; all are overwhelmingly more likely to be engaged by people who identify as

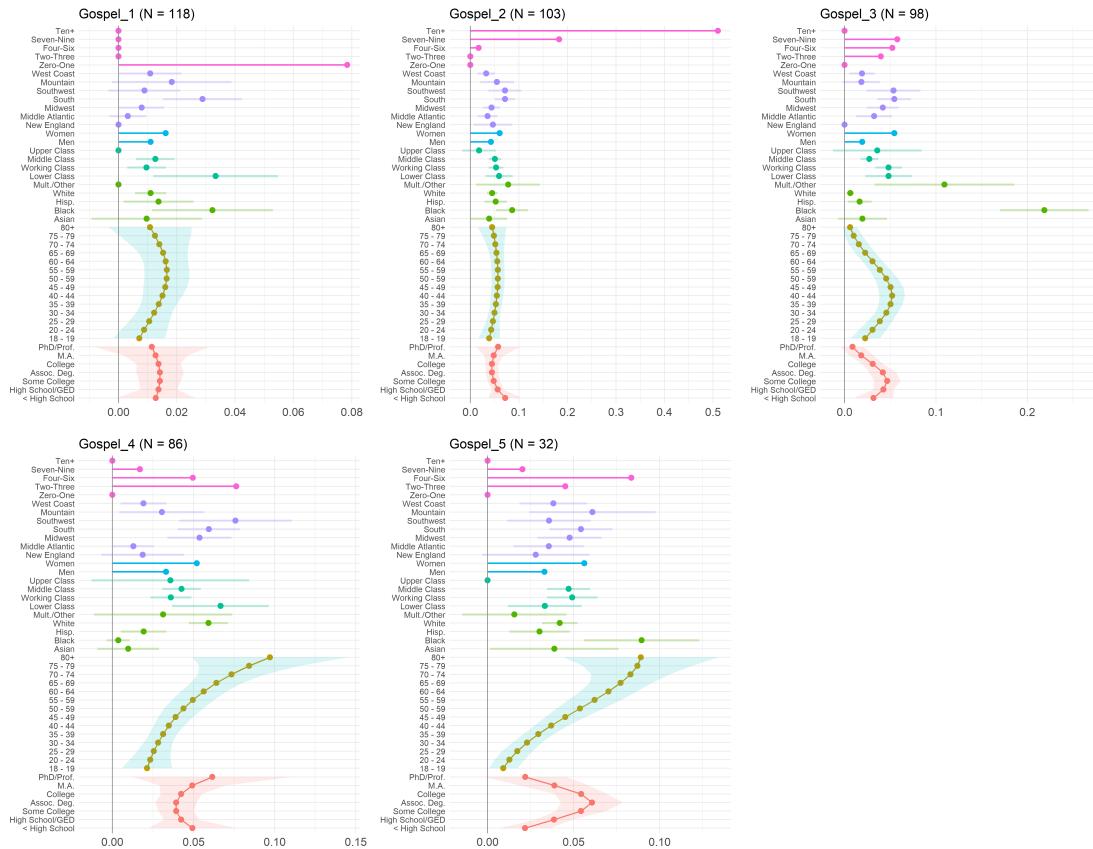


Figure 14

Hispanic (except for 3 which also attract people who identify as multiracial or “other”). Yet, they differ on key characteristics that cut across ethnic identification. Versions 1, 3, and 4 are more likely to attract young people of middling education, but those who prefer 1 and 3 more likely to combine this engagement with engagement of other genres. The second Latin micro-genre variation, on the other hand, appeals to middle-age people (mainly women) who have more education and identify as middle class.

6.7 Heavy Metal

Previous studies in the sociology of taste show that there is no genre more polarizing than Heavy Metal. Even the most “tolerant” of omnivores have no problem expressing dislike of it. Mass stereotypes of the genre [Lizardo and Skiles, 2016], show it to be associated with less educated, “lower” or working-class young white men, which accounts for its persistent role as a resource for symbolic boundary drawing, as encoded in the title of Bryson [1996] “Anything but Heavy Metal.” But in (liking or rejecting) metal are Americans affiliating or disaffiliating from a single monolithic thing? As shown in Figure 16, the micro-genre analysis discovers four variations of the “heavy metal” micro-genre. These results show support for the idea of “Heavy Metal” being one of the most sociodemographically consistent genres, but also show important *generational* variations in its micro-instantiations. On the consistency side, the sociodemographic reality of

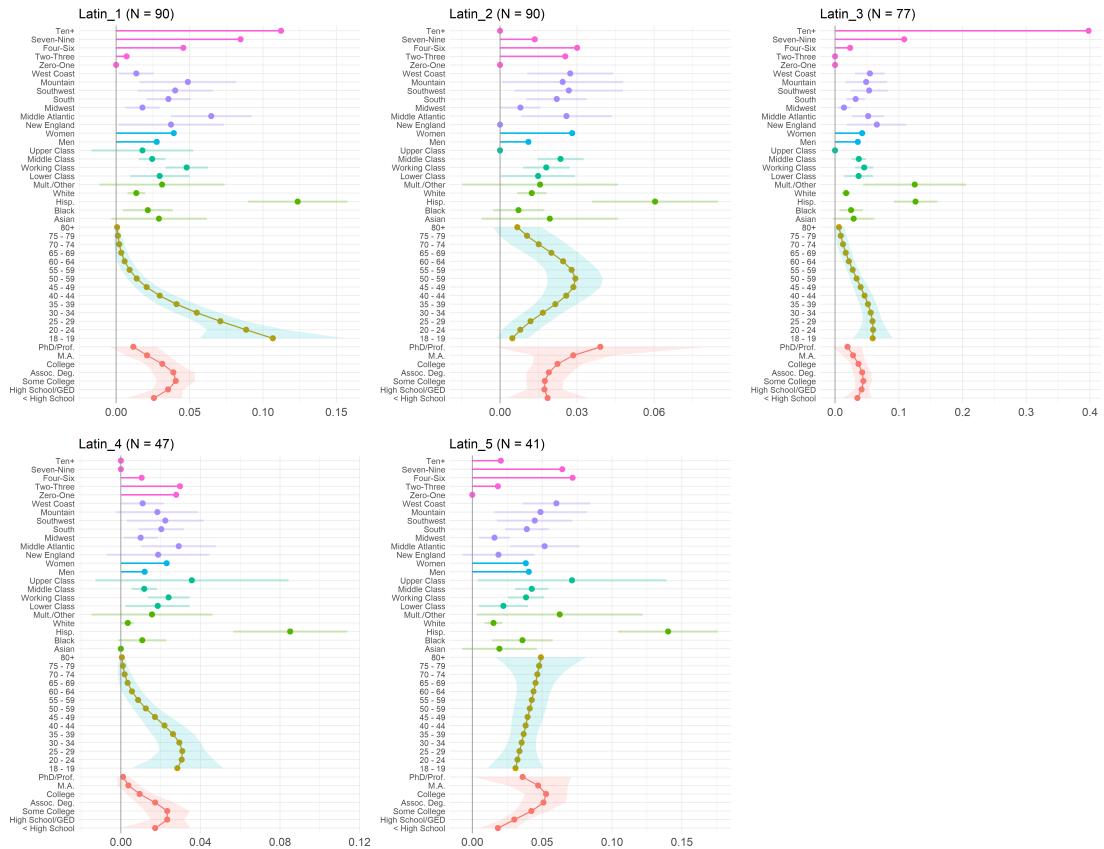


Figure 15

engagement seems to conform to stereotypical perceptions. Heavy Metal is gendered, racialized, and classed.

Men are more likely to engage all micro-genre variants more than women, consistent with previous work pointing to the strong gendering of the genre [Miller, 2016]. The only difference, among the four micro-genres uncovered is the size of the men-to-women disparity. For instance, “Heavy_Metal_3” ($N = 75$) is the most tilted toward men, while variations 1 ($N = 89$) and 4 ($N = 70$) are less so. White respondents are more likely to engage all four micro-genre variations, but so are those who classify themselves as “multiracial or other,” especially for the Metal micro-variation that mixes well with other genres (although the relatively small number of these respondents does not lead to very precise estimates). Surprisingly, for three of the four variations (excepting 4), respondents who identify as Hispanic are statistically more likely to engage the genre than Black respondents. In the same way, for three of the four variations, Metal fans are less likely to have a college degree, although variant 2 ($N = 88$) shows a tilt toward middling (e.g., “some college”) levels of education. Interestingly, this is also the micro-genre that has the highest concentration of older adult respondents (e.g., people in their 40s and 50s). The subjective class identification of Metal fans does controvert that projected to them by others; while it is true that some micro-variations are likely to feature people who identify as either lower or working class, they are likely to contain statistically indistinguishable proportion of people who identify as middle class.

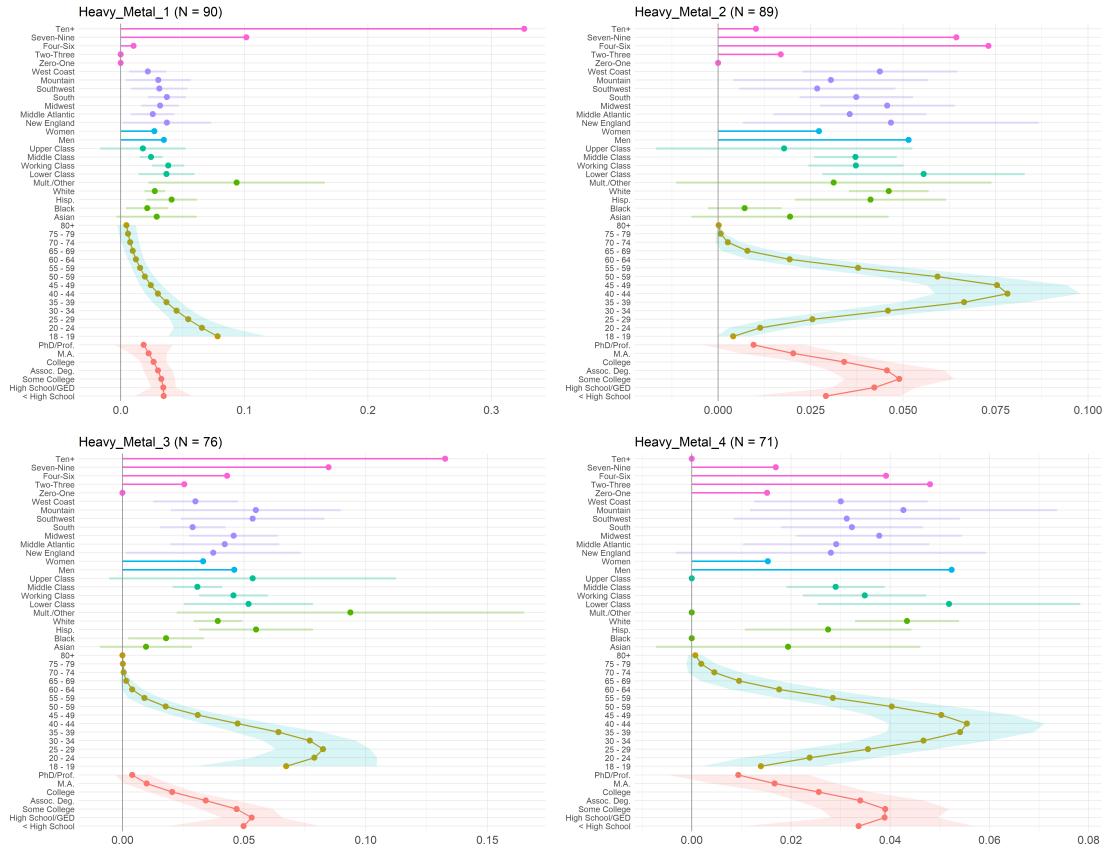


Figure 16

6.8 Classic Rock/Oldies

Perhaps the test case for the misleading nature of the macro-genre label approach are composite or vague labels like “Classic Rock/Oldies.” The suspicion is that such a vague designation hides a wide variety of more focused and perhaps stylistically and demographically disjoint micro-genres. The link-community approach produces results consistent with this claim. This macro-genre label (the most “popular” in terms of audience size in the original data) is also the one that splits into the largest number (thirteen) of micro-genres, shown in Figures 17a, and 17b. These vary substantially in terms of all the sociodemographic indicators considered.

The largest community (version 1, $N = 149$) depicts the Classic Rock univore; a middle-aged white person who resides in the Midwest, Southwest or Mountain West with less than a high-school education. The second-largest Classic Rock micro-genre ($N = 139$) is a lot like the first, except that the tilt to older respondents is much stronger as is concentration among respondents who either finished high school or received some college education. But perhaps the prototypical Classic Rock micro-genre is version 6 ($N = 76$), appealing to older respondents very much likely to identify as both white and working class, with relatively low education, who reside in the Mountain and Southwest region.

Some of the smaller micro-genre communities controvert the usual expectations as to who is the typical Classic Rock/Oldies fan [Lizardo and Skiles, 2016]. For instance, people who engage

variation 9 ($N = 60$) are of low education (which is expected), but they are also likely to be very young, working or middle class-identified Hispanics with less than a high school education and live in the Southwest. Version 10 ($N = 55$) is a lot like this last one, but concentrates among respondents with some college attainment; yet another variant of this micro-genre (version 13, $N = 31$) tilts toward middle-aged Hispanic-identified women who live on either coast.¹ There is even a version of Classic Rock that contravenes socioeconomic, gender, and regional expectations, appealing to middle-class, highly educated women who identify as white and live in New England (version 8, $N = 32$).

6.9 Pop/Top 40

As shown in Figure 18, landscape of the very vague macro-genre label “Pop/Top 40,” presents as variegated a landscape as that presented by Classic Rock, with ten micro-genre versions. There are consistent (and expected) patterns. In terms of age distribution, Pop micro-genres communities tilt toward the very young (3, 5, and 10, $N = 104$, $N = 95$, and $N = 17$, respectively) or toward young adults (1, 2, $N = 143$ and $N = 126$). The rest are either age-neutral (7, 8, 9, $N = 64$, $N = 43$, and $N = 30$, respectively) and only two (4, 6, $N = 102$, $N = 71$) tilt toward or are preferred by people in their 40s and 50s. None concentrate among older respondents. The very young pop micro-genres communities are differentiated by class: Version 3 is low education, working class, while version 5 is college-educated and middle class. Version 10 is differentiated by gender and race, attracting mainly young nonwhite women. Versions 3, 7, and 8 are pop micro-genres that combine well with others at the level of people’s repertoire (their audience tend to be omnivorous), while version 1 reveals a classic pop univore.

6.10 Country

The macro-genre label of “country music” has always been one suspected of hiding untold levels of micro-genre heterogeneity. This is not only the case style, but also sociodemographics, generations and even politics [Rossman, 2004]. As shown in Figure 19, the micro-genres discovered by the cluster analysis agree with this take. First, the analysis reveals ten distinct macro-genres within the larger vague label. Within these, there is strong generational differentiation across different versions of “country.” Five appeal mainly to older audiences (1, 2, 4, 6, 8, $N = 123$, $N = 115$, $N = 87$, $N = 70$, and $N = 58$, respectively), two tilting toward younger fans (3 and 7, $N = 110$ and $N = 61$), and two engaging middle-aged people (5 and 10, $N = 70$ and $N = 49$). Micro-genre 1 recovers the prototypical “country univore.” Fans of these micro-genre are largely white, shun the upper-class label, do not have a college degree, and engage only this micro-genre to the exclusion of others. Version 3, by way of contrast, is the obligatory micro-genre that mixes indiscriminately with other genres. Country micro-genres attracting older respondents exhibit class differentiation, with versions 1, 2, and 6 attracting respondents with less than a college education, and versions 4 and 8 reversing this pattern; the great bulk of fans of country version 4 identify as middle class making this micro-genre distinctive in that respect.

7 Discussion

Let us take stock. We began our discussion with the network or “relational” revolution in the quantitative study of taste in sociology [Pachucki and Breiger, 2010]. We noted that many good

¹Perhaps these last set of respondents understand “Classic Rock” to include such long existing American genres as “Tex-Mex.”

things have come from thinking of survey data on taste using a relational lens, inclusive of a deeper specification of core concepts in the literature, and even the discovery of new phenomena and empirical patterns. However, we also noted that survey data are as good as the labels chosen to collect the data on. A resurgent line of critics question whether the labels that appear in our most venerable survey-based studies are true “genres” in the sociological (or even stylistic) sense [Lena, 2015, Vlegels and Lievens, 2015]. The labels are broad, likely to be interpreted by people in heterogeneous ways, and thus hide as much as they reveal. I noted recent attempts to just drop the idea of vague macro-genre labels and either study actual sociological genres on the ground (thus partially abandoning studies of audience segmentation or at least radically reconfigure them) or “dropping the label” [Sonnett, 2016] by querying people about more focused objects of taste (e.g., performers within genres). These are all important and good developments, but we also noted that they may be throwing away a good thing. Upping the relationality by exploiting hidden patterns lying in the same old vague macro-genre based data that has been collected before can reveal focused micro-genres.

I proposed to do this using recent developments in the discovery of *overlapping communities* in networks [Ahn et al., 2010], which partially answers two of the challenges of macro-genre critics: The fact that actual genres are overlapping and not crisply bounded, and the fact that there is hidden heterogeneity within the broad labels we usually focus on. Our analysis contrasted the application of the usual techniques to the same data, conceived in the usual “macro” way, and *after* upping the relationality and discovering the micro-genres hidden within. It is clear that focusing on micro-genres reveals variation and patterns of cultural choice (as well as audience segmentation) that we would not have noticed using the standard approach (in this, the critics are right), but we were able to do this without dropping the label or querying people about hundreds of micro-styles (most of which they’d be unfamiliar with). Instead, we exploited the venerable principle, etched into classical approaches to defining genres in network terms [DiMaggio, 1987], noting that if genres are defined by the people that choose them, then they are also defined by the other choices that people make when they choose them; the country that combines with one version of Classical may not be the same country that combines with a version of the Blues (and neither is the same as the country that refuses to combine with any other style).

The approach here is general and can be applied obviously to the study of other genre complexes beyond musical taste (allowing for economic data collection), but also for the study of other processes beyond taste. This includes belief, opinion, and attitude data. Essentially allowing us to move from “vague” responses, to more focused responses, by exploiting the hidden patterns in the inter-response network formed when people respond to other items. Overall, we also learned a general lesson with respect to our usual people or item classification techniques: To the extent that these are applied to items which themselves are “vaguely” defined, they will also yield even vaguer (and perhaps less than useful) macro-classifications of those items (like “popular” and “high status”). The same may be said for such macro-classifications of opinions such as “liberal” and “conservative,” which while being the primary way we interpret how people split themselves into groups by the beliefs they choose to hold, may be decreasingly apt today, if they ever were.

References

- Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *nature*, 466(7307):761–764, 2010.
- Zsuzsa Bakk, Fetene B Tekle, and Jeroen K Vermunt. Estimating the association between latent

class membership and external variables using bias-adjusted three-step approaches. *Sociological methodology*, 43(1):272–311, 2013.

Mónica Bécue-Bertaut and Jérôme Pagès. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics & Data Analysis*, 52(6):3255–3268, 2008.

Stephen P Borgatti and Martin G Everett. Network analysis of 2-mode data. *Social networks*, 19(3):243–270, 1997.

Stephen P Borgatti and Daniel S Halgin. On network theory. *Organization science*, 22(5):1168–1181, 2011.

Pierre Bourdieu. *Distinction: A social critique of the judgement of taste*. Harvard university press, 1984.

Ronald L Breiger. The duality of persons and groups. *Social forces*, 53(2):181–190, 1974.

Bethany Bryson. " anything but heavy metal": symbolic exclusion and musical dislikes. *American sociological review*, 61(5):884–899, 1996.

Tak Wing Chan and John H Goldthorpe. Social stratification and cultural consumption: The visual arts in england. *Poetics*, 35(2-3):168–190, 2007.

Paul DiMaggio. Classification in art. *American sociological review*, pages 440–455, 1987.

Magne Flemmen, Vegard Jarness, and Lennart Rosenlund. Social space and cultural class divisions: the forms of capital and contemporary lifestyle differentiation. *The British journal of sociology*, 69(1):124–153, 2018.

Amir Goldberg. Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. *American journal of sociology*, 116(5):1397–1436, 2011.

Michael Greenacre and Jorg Blasius. *Multiple correspondence analysis and related methods*. CRC press, 2006.

Shin-Kap Han. Unraveling the brow: What and how of choice in musical preference. *Sociological Perspectives*, 46(4):435–459, 2003.

Alex T Kalinka and Pavel Tomancak. linkcomm: an r package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*, 27(14):2011–2012, 2011.

Gary King, Michael Tomz, and Jason Wittenberg. Making the most of statistical analyses: Improving interpretation and presentation. *American journal of political science*, pages 347–361, 2000.

Pierre-Antoine Kremp. Innovation and selection: Symphony orchestras and the construction of the musical canon in the united states (1879–1959). *Social Forces*, 88(3):1051–1082, 2010.

Sébastien Lê, Julie Josse, François Husson, et al. Factominer: an r package for multivariate analysis. *Journal of statistical software*, 25(1):1–18, 2008.

Brigitte Le Roux and Henry Rouanet. *Geometric data analysis: from correspondence analysis to structured data analysis*. Springer Science & Business Media, 2004.

- Jennifer C Lena. Relational approaches to the sociology of music. *Routledge International Handbook of the Sociology of Art and Culture*. New York: Routledge, pages 149–160, 2015.
- Omar Lizardo. Omnivorousness as the bridging of cultural holes: A measurement strategy. *Theory and Society*, 43(3-4):395–419, 2014.
- Omar Lizardo. The mutual specification of genres and audiences: Reflective two-mode centralities in person-to-culture data. *Poetics*, 68:52–71, 2018.
- Omar Lizardo and Sara Skiles. Musical taste and patterns of symbolic exclusion in the united states 1993–2012: Generational dynamics of differentiation and continuity. *Poetics*, 53:9–21, 2015.
- Omar Lizardo and Sara Skiles. Cultural objects as prisms: Perceived audience composition of musical genres as a resource for symbolic exclusion. *Socius*, 2:2378023116641695, 2016.
- Paul Lopes. *The rise of a jazz art world*. Cambridge University Press, 2002.
- Diana L Miller. Gender, field, and habitus: How gendered dispositions reproduce fields of cultural production. *Sociological Forum*, 31(2):330–353, 2016.
- Sosuke Okada. Structure of cultural rejection. *Sociological Perspectives*, 60(2):355–377, 2017.
- Mark A Pachucki and Ronald L Breiger. Cultural holes: Beyond relationality in social networks and culture. *Annual review of sociology*, 36:205–224, 2010.
- Richard A Peterson. Patterns of cultural choice: A prolegomenon. *American Behavioral Scientist*, 26(4):422–438, 1983.
- Richard A Peterson. Understanding audience segmentation: From elite and mass to omnivore and univore. *Poetics*, 21(4):243–258, 1992.
- Richard A Peterson and Roger M Kern. Changing highbrow taste: From snob to omnivore. *American sociological review*, pages 900–907, 1996.
- Henk Roose, Koen Van Eijck, and John Lievens. Culture of distinction or culture of openness? using a social space approach to analyze the social structuring of lifestyles. *Poetics*, 40(6):491–513, 2012.
- Gabriel Rossman. Elites, masses, and media blacklists: The dixie chicks controversy. *Social Forces*, 83(1):61–79, 2004.
- Mike Savage and Modesto Gayo. Unravelling the omnivore: A field analysis of contemporary musical taste in the united kingdom. *Poetics*, 39(5):337–357, 2011.
- Jennifer Schultz and Ronald L Breiger. The strength of weak culture. *Poetics*, 38(6):610–624, 2010.
- John Sonnett. Ambivalence, indifference, distinction: A comparative netfield analysis of implicit musical boundaries. *Poetics*, 54:38–53, 2016.
- Gindo Tampubolon. Revisiting omnivores in america circa 1990s: The exclusiveness of omnivores? *Poetics*, 36(2-3):243–264, 2008.
- Koen Van Eijck. Social differentiation in musical taste patterns. *Social forces*, 79(3):1163–1185, 2001.

Jef Vlegels and John Lievens. Music genres as historical artifacts: The case of classical music. *Connections*, 35(1):53–62, 2015.

Jef Vlegels and John Lievens. Music classification, genres, and taste patterns: A ground-up network analysis on the clustering of artist preferences. *Poetics*, 60:76–89, 2017.

Harrison C White, Scott A Boorman, and Ronald L Breiger. Social structure from multiple networks. *American Journal of Sociology*, 81(4):730–780, 1976.

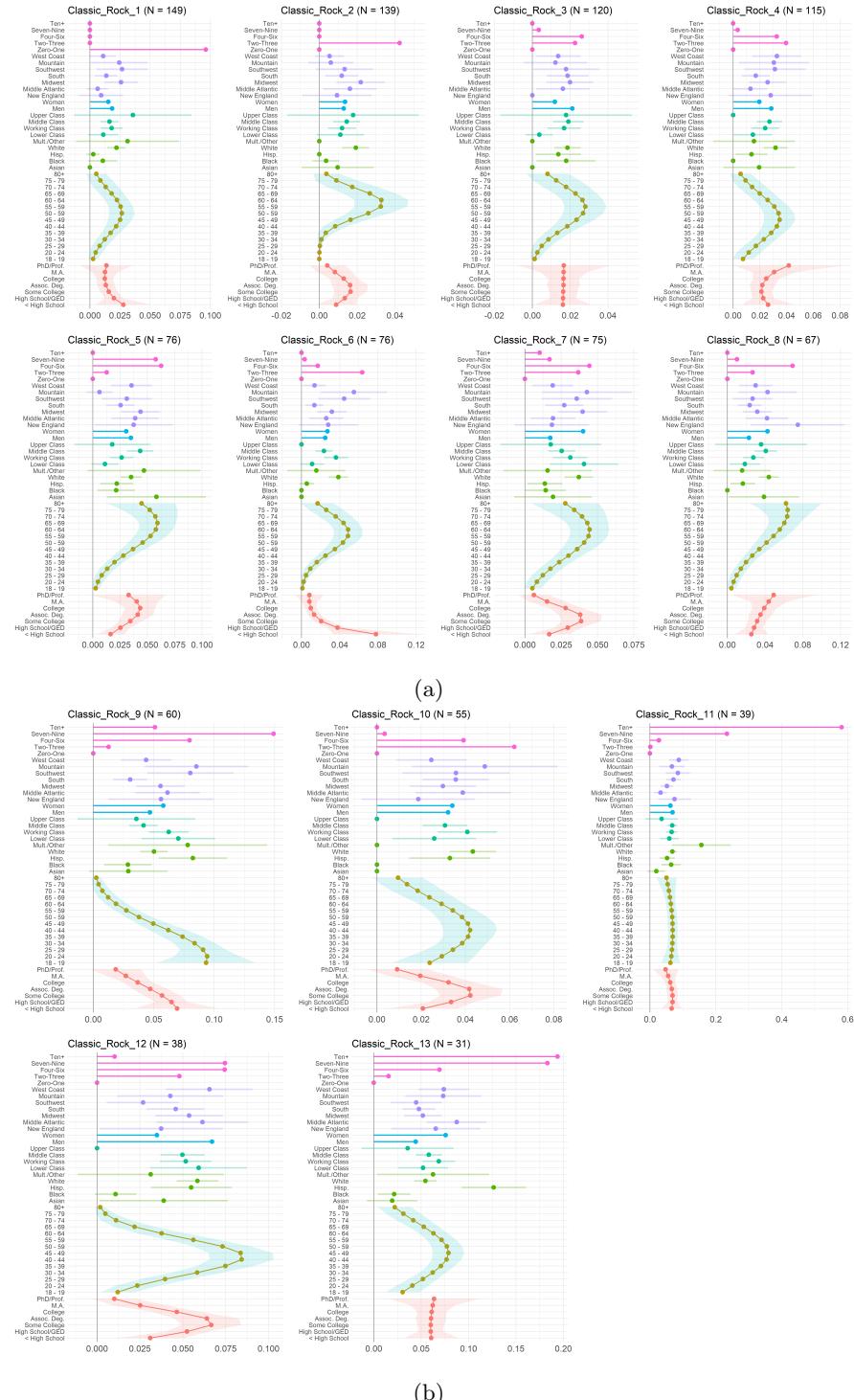


Figure 17

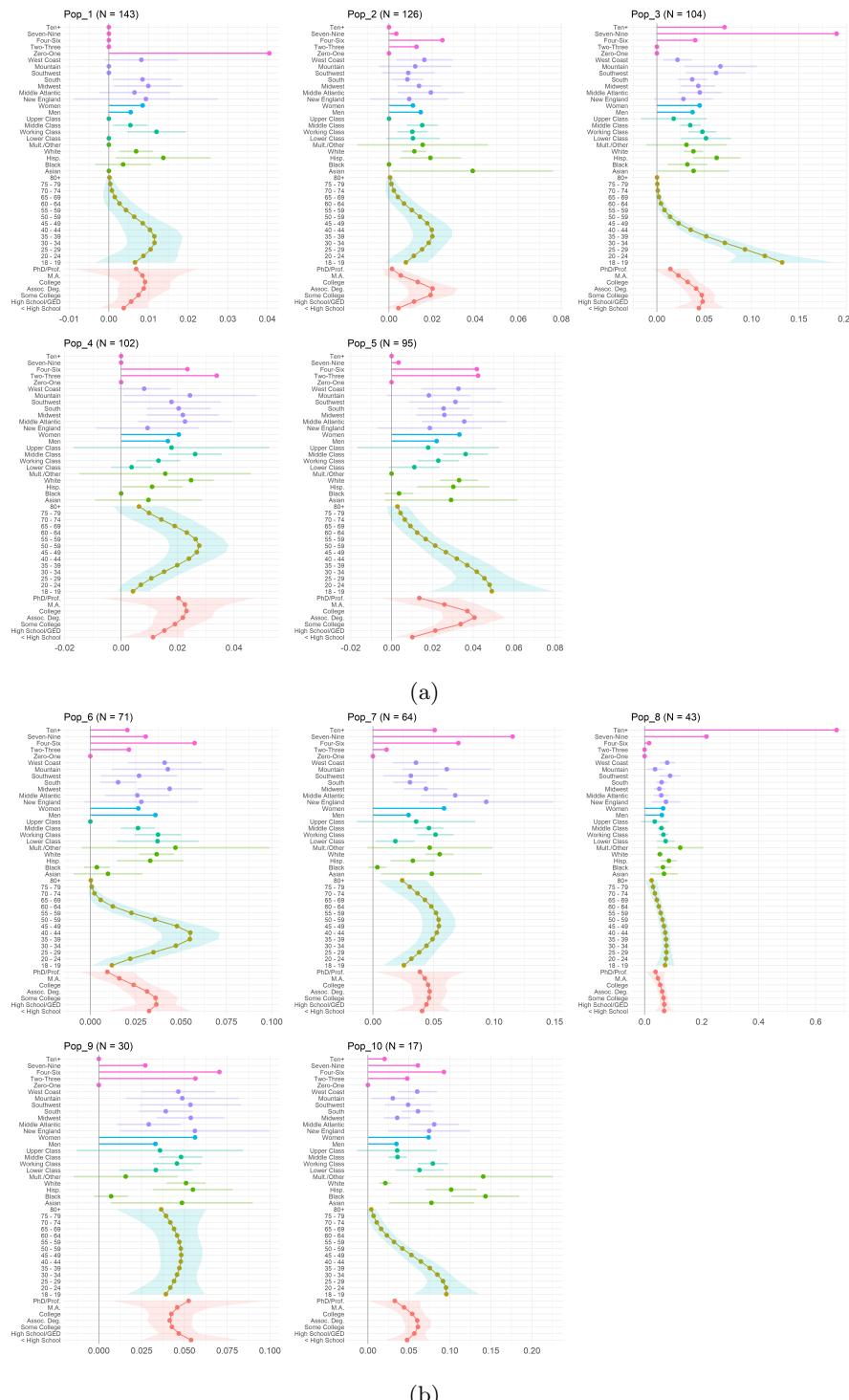


Figure 18

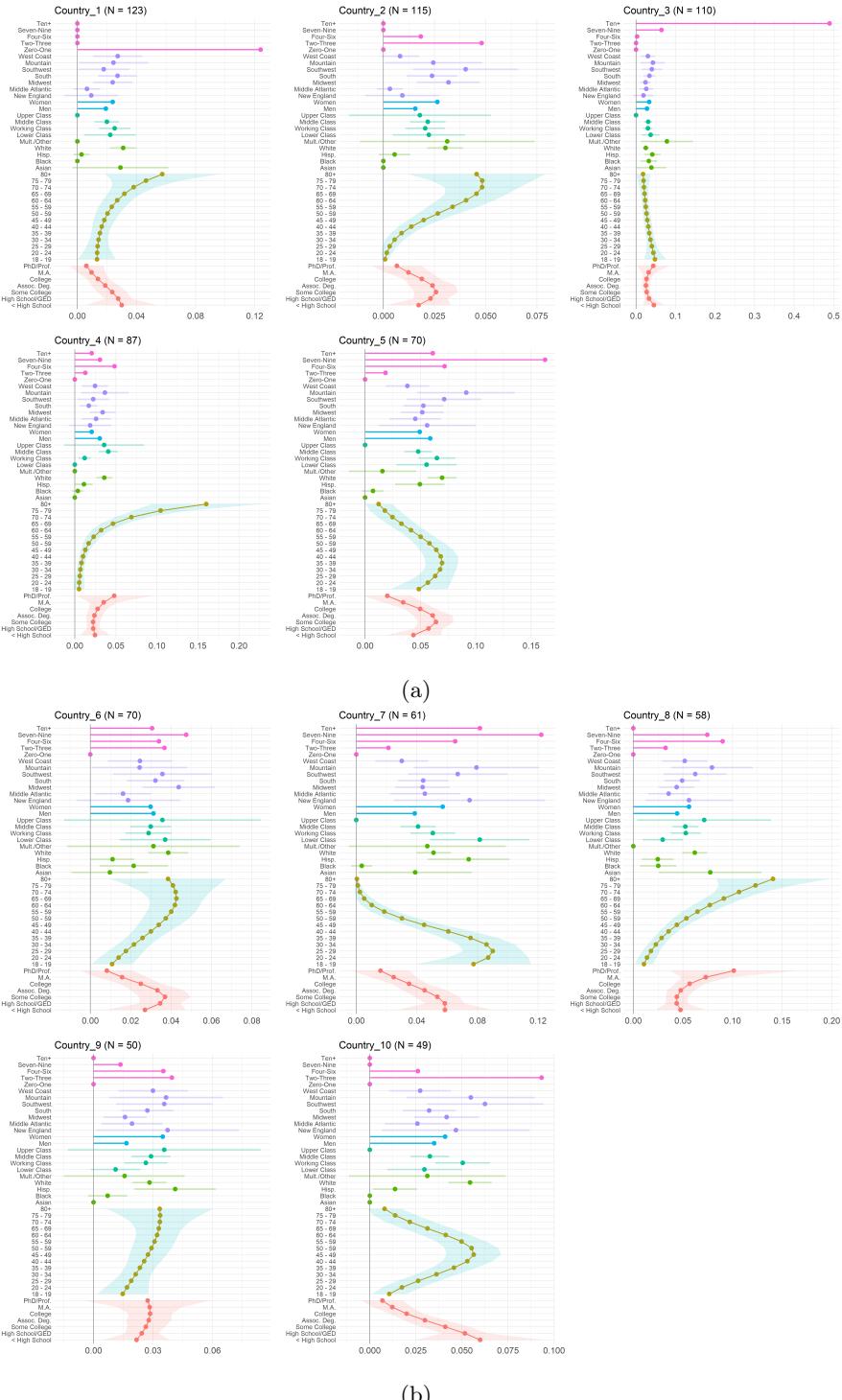


Figure 19