Socially-Structured Mobility Networks and School Segregation Dynamics:
The Role of Emergent Consideration Sets*

Julia Burdick-Will
Jeffrey A. Grigg
Kiara Millay Nerenberg
Johns Hopkins University

Faith Connolly
McREL International

May 2020

Key words: segregation, social networks, decision-making, school mobility

*Acknowledgements:
This research was made possible by a grant from the 21st Century Cities Initiative at Johns Hopkins University and indirect support from the Baltimore Education Research Consortium. Curt Cronister provided invaluable technical support. We are grateful for feedback and assistance from the Baltimore City and Baltimore County Schools. All errors and opinions are our own.

Citation:
Burdick-Will, Julia, Jeffrey A. Grigg, Kiara M. Nerenberg, and Faith Connolly. 2020. "Socially Structured Mobility Networks and School Segregation Dynamics: The Role of Emergent Consideration Sets." *American Sociological Review* 85(4).

Abstract

This study proposes and applies a novel method for empirically evaluating the role of social structure in the school sorting process. We use administrative records from Baltimore City and suburban Baltimore County public elementary schools (2011-2015) to generate a network of schools based on student transfers. We then apply repeated calculations of the Louvian method of community detection to estimate emergent sets of schools that are likely to be considered by similar parents – which we term *emergent consideration sets* – and use gravity models to explore the role of social structure, demographics, and geography in observed enrollment patterns. We find that our network-derived emergent consideration sets are better defined by structural boundaries than student composition or proficiency alone. Within consideration sets, students tend to avoid schools with relatively higher levels of free- and reduced-price meal eligibility and flock towards schools with higher proficiency levels. School racial composition, however, plays a much smaller role in predicting movement between schools, in part because structural constraints generate racially homogeneous consideration sets. Together, these findings highlight how regional social and geographic organization shapes school segregation processes and the policies used to combat them.

Introduction

American schools are highly segregated by race and there are large gaps in school quality between white and non-white students (Logan and Burdick-Will 2017; Logan, Minca, and Adar 2012; Reardon, Kalogrides, and Shores 2019). These disparities in educational opportunity play a large role in racial gaps in achievement, attainment, and employment (Hanushek and Rivkin 2012; Johnson 2011). Historically, this inequality has been tied to questions of access, as districts first explicitly maintained segregated schools and later drew district and enrollment zone boundaries in ways that maintained racial patterns in enrollment (Reardon and Owens 2014; Saporito and Sohoni 2006). Today, integration is considered something that must be achieved voluntarily. Court ordered desegregation and race-conscious assignment policies are being overturned or eliminated (Reardon et al. 2012) and the increasing availability of charter schools and open enrollment policies now allow families to decouple their residential and school decisions (Lareau and Goyette 2014). Unfortunately, this increasing reliance on individual choice has resulted in more, not less, segregated schools (Bifulco and Ladd 2007; Kotok et al. 2017; Renzulli and Evans 2005; Stein 2015; Tatian and Lei 2015).

Recent insights from complex decision theory help explain the relationship between increasingly available choice and segregation by suggesting that existing inequalities in social structure lead parents to racially isolated schools, regardless of their stated preferences (Bruch and Swait 2019; Krysan and Crowder 2017). Parents do not make enrollment decisions in isolation or at a single point in time (Altenhofen, Berends, and White 2016; Cahill 1994; Lareau and Goyette 2014; Schwartz 2004). Like other consumers and complex decision makers, parents are likely to use a two-stage method of information gathering (Buckley and Schneider 2003; March 1994). To select a school, parents first narrow their options and define the set of schools they are willing to consider and then evaluate their options within that smaller set (Frisch and Clemen 1994; Kahneman and Tversky 1979). During the narrowing phase parents rely heavily on past experiences, informal networks, word of mouth, and school reputation to restrict their options (Ball and Vincent 1998; Holme 2002; Smrekar and Goldring 1999). In other words, parents frequently rely on their social contacts' perceptions of schools and districts, rather than objective, observable information, to inform their choice sets and often rule out large numbers of schools or districts without giving them careful consideration.

In making school enrollment decisions, parents rely on social contacts and past experiences that are embedded within a stratified social structure (Ball and Vincent 1998; Fong 2019), but this structure remains loosely defined and difficult to measure. Our study proposes a novel method for empirically identifying and evaluating the role of social structure in the school sorting process. Specifically, we use population-level school enrollment files from more than one hundred thousand Baltimore City and suburban Baltimore County public elementary school students in the 2010-11 through 2014-15 school years to generate network ties between schools based on student transfers. We then use repeated calculations of the Louvian method of community detection (Blondel et al. 2008) to derive emergent clusters of schools that frequently share students and are therefore likely to be considered by similar parents – which we term *emergent consideration sets*. We then evaluate the relative importance of demographic composition and structural boundaries in the formation of those aggregate consideration sets. Next, we estimate gravity models (Sen and Smith 2012) that predict actual student flows between schools and compare the relative performance of models with different assumptions about the range of schools parents are likely to consider.

We find that our network-derived emergent consideration sets fit the data much better than alternative consideration sets based on demographics and distance alone. Moreover, these emergent consideration sets are better defined by structural boundaries, such as being on the opposite side of an interstate or district line, than student composition or proficiency. We also find that gravity models are better at predicting flows of students when they take these emergent consideration sets into account than with naïve models that include every possible school. Finally, observed school characteristics become much better predictors of mobility flows within emergent consideration sets than across all possible school. Within those consideration sets, students tend to avoid schools with relatively higher levels of free- and reduced-price meal eligibility and flock towards schools with higher proficiency levels. School racial composition, however, plays a much smaller role in predicting movement between schools, in part because structural constraints generate racially homogeneous consideration sets.

These findings have important implications for our understanding of school segregation processes and the policies used to combat them. First, they highlight the incorrect assumptions that underlie much of the existing policy analysis in this area. Families do not consider every school in their district, let alone their region. Unless we design policies that take the socially structured nature of consideration sets into account, increased individual choice is likely to lead to more, not less school segregation. Second, the findings highlight the intractability and self-reproducing nature of segregation and the role of sometimes subtle structural boundaries in that sorting process. Since parents of different races often do not interact with each other in the same social or geographic circles, they do not end up considering the same places. They therefore have little chance of enrolling in the same schools, even if they are looking for the same objective characteristics. Reducing school segregation is going to take more than widening access to schools through open enrollment policies; we must also remove the geographic and social boundaries that shape patterns of interaction and geographic exposure.

Complex Decision Making

Deciding where to live and where to send one's children to school are complex, multi-dimensional, and potentially overwhelming decisions (Buckley and Schneider 2003; Cahill 1994; Schwartz 2004). Getting good information about all possible districts and schools in an area is often quite difficult and stressful (Delale-O'Connor 2018; Pattillo, Delale-O'Connor, and Butts 2014; Stein and Nagro 2015). Even the most active school consumers, such as those who voluntarily change schools, are unable to gather enough information to fully weigh all available information (Buckley and Schneider 2003). Instead, a long history of sociological research on bounded rationality and decision-making shows that both informational asymmetries (March 1994; March and Simon 1958; Simon 1990) and cognitive bandwidth restrictions (Cowan 2010; Miller 1956) impose limitations on our ability to give careful consideration to all possible options.

These limitations lead individuals to engage in a two-step process of decision-making (Bruch, Feinberg, and Lee 2016; Bruch and Swait 2019; Manski 1977; March 1994). Upon recognizing the need to make a decision, individuals first construct a smaller and more manageable choice set from the available options, known as the "editing" stage; after doing so, they can proceed to assess the options within this set and make a final selection (Kauko 2004; Manski 1977). Consider how people purchase cars: car buyers investigate each model (e.g., Toyota Camry and Ford Fusion) only after determining the class of vehicle (e.g., mid-size sedan). Similarly, to select a school, parents first narrow their options and define their choice sets in the "editing" phase and then evaluate their options within the smaller set (Frisch and Clemen

1994; Kahneman and Tversky 2013). Many parents never gather any information on a large number of schools before ruling them out (Holme 2002). Thus, the exact set of schools that end up under consideration determines in large part where a student ultimately enrolls.

Socially Structured Sorting

Research on the construction of choice sets in other contexts suggests that individuals rely on social networks in order to learn about their available options, evaluate these sources of information, and make more satisfying decisions (Centola and Macy 2007; Fong 2019). Choice set construction is framed by prior experience with similar decisions, known as "anchoring" (Furnham and Boo 2011) and may also be enabled by the use of heuristics, in which people make quick judgements about a set of options based on a few characteristics (Hertwig and Herzog 2009). These heuristics vary substantially across individuals as well as environmental contexts (Swait, Brigden, and Johnson 2014), but often rely on geography and space (Phillippo and Griffin 2016), culture (Vaisey and Valentino 2018), or social networks (Hertwig and Herzog 2009).

Traditional models and methods of estimating choice sets consider this editing phase to be a conscious part of the decision process (Buckley and Schneider 2003; Kahneman and Tversky 1979). Individuals may use different criteria to limit their choice sets than in their final selection, but are still making an active decision to include or exclude different options. Krysan and Crowder (2017) expand on these ideas in the context of housing searches and residential segregation with what they call the "social structural sorting perspective." This models recognizes that what individuals consider is shaped by previous experiences in specific geographic areas, as well as social ties to friends and family who themselves have geographically and socially limited experience. Rather than making a conscious decision to exclude specific places from their housing search, they simply look in areas that they know or their social contacts recommend. In other words, the construction of a choice set is not necessarily a conscious decision. Instead, it is a product of an already socially and geographically segregated world. The result is a highly segmented market in which residents remain racially and economically segregated despite explicit stated preferences for more integration.

This socially-structured model of housing search aligns with the qualitative literature about how families approach their school searches. In fact, housing searches are often an important part of the school selection process. In interviews and surveys, parents repeatedly say that they would like to pair their housing and schooling decisions into one “package deal” with affordable housing that provides guaranteed access to a high quality school (Rhodes and Warkentien 2017). The results of these joint decisions can be seen in the dramatic variation in housing prices across districts and even enrollment zones (Fack and Grenet 2010; Gibbons, Machin, and Silva 2013) as well as the apparent premium that families are willing to pay to be zoned to schools with higher test scores (Bayer, Ferreira, and McMillan 2007). For those who cannot afford to live in a neighborhood with their preferred school, school choice and open-enrollment policies allow parents to separate these two decisions and send their children to schools that are not tied to their residential address (Lareau and Goyette 2014). In sum, the literature on school preferences suggests that parents are rarely passive consumers of schooling and will use whatever resources they have to ensure that their children attend a school that they find appealing.

In contrast, what parents find appealing and how they gather information about schools are not straight forward. When asked directly, parents frequently highlight two aspects of school that they care most about. First, parents repeatedly say that they want a “high quality” school for their children (Burgess et al. 2015; Hastings and Weinstein 2008; Holme 2002). Proficiency rates on state-level standardized tests are the most readily available measure of quality, but they are often a better measure of the economic background of students than any real measure of quality instruction (Koretz 2008). Second, parents frequently discuss the importance of convenience and talk about a school's location in relation to home or work as well as travel time during rush hour (Bell 2007; Burgess et al. 2015; Denice and Gross 2016). Although parents rarely say that demographics are a determining factor, surveys and experimental studies show that white parents tend to avoid schools with high proportions of black students (Billingham and Hunt 2016; Lareau and Goyette 2014; Roda and Wells 2013; Saporito and Lareau 1999; Schneider and Buckley 2002) and many families also stay away from schools with high numbers of poor children (Kotok et al. 2017).

However, despite these measures (achievement, convenience, and composition) being relatively readily available, it is not clear that parents actually do much objective research on them during their search. Instead, parents are much more likely to rely on word-of-mouth to decide which schools to consider (Ball and Vincent 1998; Holme 2002; Smrekar and Goldring 1999). This information often comes from local neighborhood-based networks, both because parents are more likely to interact with those who live nearby on a daily basis and because those who live nearby likely face some of the same options and trade-offs (Bader, Lareau, and Evans 2019).

Much of this shared perception is negative rather than positive—where to avoid as much as where to apply. Holme (2002), for example, finds that most of the white, affluent families in her study who left urban neighborhoods because of the "bad" public schools had not gathered any objective information about the schools in their old neighborhood and had relied entirely on the opinions of other high-status parents in their personal networks. Similarly, many parents of all income levels in Cleveland relied heavily on the reputation of all traditional Cleveland public schools as unacceptable and did anything they could to either move to the suburbs or get into a charter school (Rhodes and Warkentien 2017). In other words, parents of all races and classes use their personal experiences and social connections to rule out whole districts or subsets of schools without carefully weighing each school's individual plusses and minuses. Importantly, those sources of information are likely to be socially structured and segregated in ways that exacerbate inequality (Bell 2009).

Emergent Patterns in Complex Systems

Similar to what Krysan and Crowder's model shows with housing, none of these individual school enrollment decisions likely take place in a vacuum. For example, current social ties shape a family's enrollment decision, but their decisions, in turn, shape the composition of their social and geographic exposure in the future. Systems with this kind of complexity often generate emergent social patterns in which aggregations of individual decisions can produce district-level outcomes that no one person could have necessarily anticipated (Macy and Willer 2002).

The earliest demonstration of these collateral consequences comes from Schelling's (1971) classic residential segregation simulation. Using a very simple model, he shows that residents of an abstract grid do not need to prefer segregated environments to end up in very segregated contexts. In the decades since this theoretical exercise was published, numerous studies have replicated the findings with more realistic models and data on actual residential patterns (Benard and Willer 2007; Benenson, Hatna, and Or 2009; Bruch 2014; Bruch and Mare 2006, 2009; Clark 1991; Flache and Hegselmann 2001; Fossett 2006). The role of population structure and this kind of dynamic feedback has also been demonstrated in marriage markets (Blau, Blum, and Schwartz 1982), friendship formation (Moody 2001; Wimmer and Lewis 2010), organizational survival (Lomi and Larsen 1998), and labor market discrimination (Duong and Reilly 1995).

The key to these findings is that global complexity does not necessarily reflect the cognitive complexity of individuals (Macy and Willer 2002). In other words, the group-level sorting process can appear non-linear and complicated even when people themselves are making simple and straightforward decisions. What generates complexity in these systems is the complicated nature of the underlying social structure, not the ways in which individuals react to it. Moreover, people respond to their environments with relatively simple strategies of moving, imitating, avoiding, or learning, rather than being completely rational at every decision point (Holland 1995). Therefore, in order to examine these emergent sorting processes, one cannot simply study the individual-level factors that predict a single student's preferences or decisions. Instead, one must also study the patterns that these decisions generate in the aggregate.

Applying these findings to school enrollment suggests that the decisions of individual families may be idiosyncratic, but in the aggregate their collective decisions reveal an emergent structure that segments the regional market for schools (Bowe, Ball, and Gewirtz 1994; Sirer et al. 2015). These sub-markets reflect the combined choice sets of many individual families in similar circumstances and are reinforced by their collective enrollment decisions over time. In other words, if choice sets are framed by socially structured sorting patterns, then repeated decision-making by individuals with similar social positions will contribute to further market segmentation (Bruch, Hammond, and Todd 2015; Kauko 2004). For example, through repeated instances of residential mobility, individual-level preferences create even more housing market segmentation than one would expect from financial constraints alone (Adair et al. 2000; Cahill 1994; Kauko 2004). Similar segmentation has been identified in online dating, which is even less reliant on financial constraints (Bruch and Newman 2018). Thus, drawing on decision-making repetition can allow us to infer the structure of consideration sets (Bruch and Atwell 2015; Tuljapurkar, Bruch, and Mare 2008) and deduce the logic which drives decision-making within them (Leishman et al. 2013).

Therefore, in this paper we focus not on the individual choice set of specific families, but instead on the structure of what we call *emergent consideration sets*. We use this term, rather than the more common “choice sets” for two reasons. First, by using “consideration” rather than “choice” we hope to de-emphasize the active construction of these sets. Given the socially-structured sorting perspective proposed by Krysan and Crowder (2019) and what we know about how families make enrollment decisions, these sets are rarely the result of an active editing process based on specific school characteristics. Instead, they represent the often unconscious filtering of information that takes place in our highly structured social and geographic environments. Second, these sets are an emergent property of the network and do not represent the individual choice set of any specific family. Rather, they reflect the segmentation of the elementary school marketplace and capture clusters of schools that are likely to be considered by similar families in similar social and geographic positions.

Since the structure of these aggregate consideration sets is an emergent property of a complex system, the best way to assess it is not through experiments, survey responses, interviews, or even analysis of individual administrative choice forms. Asking individuals how they decide – or even relying on their listed preferences – does not always reveal the informal social networks that structure their choices. The exact preferred combination of attributes may not exist in real life, sources of information used are not always fully conscious (Small 2017), and when it comes to schools, many students end up enrolling in something other than what they initially listed on their choice form (Stein, Burdick-Will, and Grigg 2019). In contrast, our proposed method uses network clustering algorithms to track actual transfers between schools in ways that highlight the emergent social structure that shapes decision-making. These network methods reveal the connections between what would otherwise appear to be static and isolated entities, such as neighborhoods, schools, and districts (Browning et al. 2017; Graif, Gladfelter, and Matthews 2014; Sampson 2008; Wang et al. 2018).

Hypotheses

Existing theory of socially structured complex decision-making and qualitative evidence regarding school selection processes leads to four concrete hypotheses that can be tested with our data and methods. To our knowledge, these theories have not yet been tested using large scale quantitative data using these or any other methods.

Structure of Emergent Consideration Sets

> Hypothesis 1: Emergent consideration sets derived through community detection methods will better predict the flow of students between schools than those derived from distance, demographics, and test scores alone.
>
> Hypothesis 2: Structural barriers (distance, district boundaries, interstates, etc.) will be more important than demographics in predicting the composition of emergent consideration sets.

We expect to find that our emergent consideration sets will be better predictors than simple demographics because they capture the hard to measure social structures that shape interaction. We also know from the qualitative literature on school selection that parents often rule out whole districts or large groups of schools without doing any detailed research. This would suggest that structural boundaries that can be easily observed (i.e. school sector, in a particular district, on a particular side of the interstate) will be strong predictors of our emergent consideration sets. If demographics and distance are instead good predictors of student movement and the division of emergent consideration sets, it is a sign that these characteristics play a "deal breaking" role in decision making (Bruch et al. 2016) and that past experience and informal flows of information matter less than expected.

Logic of Mobility Flows

> Hypothesis 3: Incorporating emergent consideration sets will improve the prediction of mobility flows compared to models that assume every student considers every school.
>
> Hypothesis 4: Observed measures of student composition will be stronger predictors of student flows after taking into account the emergent consideration sets.

We expect that when it comes to a final detailed deliberation between a small set of schools, families are likely to weigh their options the way they say they do in interviews and surveys, which is value academic quality, distance, and possibly student demographics. However, since they often rule out large groups of schools without careful consideration, these factors will not likely predict mobility in models that assume every student considers every school. In other words, relative proficiency rates may appear to matter very little when comparing across an entire district, but when you limit the options to just those that are likely to be considered, parents are more likely to move to the higher achieving schools within that set. Therefore, as the models approach more realistic assumptions about emergent consideration sets, we would expect the observed characteristics of schools to matter more in terms of both the size of the coefficients and overall ability to predict the flow of students.

Baltimore Context

The Baltimore metropolitan region consists of Baltimore City and six surrounding county-based school districts. The two largest of these jurisdictions are Baltimore City and Baltimore County. Despite sharing a name, they are completely independent counties. Together they account for more than half of the metropolitan area's 2.8 million people. Suburban Baltimore County almost entirely surrounds Baltimore City on all sides. This means that it is possible to examine the city and its inner suburbs using these two districts alone. Moreover, Baltimore County represents a much wider range of land use density and median income than the city alone. Within the more than 600 square mile county, there is everything from high density, high poverty apartment complexes to concentrated areas of affluence in sprawling estates. These concentrations of affluence are less frequently the subject of choice and mobility studies and will provide a more appropriate contrast to experiences than could be found in the city alone. According to the 2010-2014 American Community Survey, Baltimore County is similar in terms of population density and demographics to the other suburban counties in the Baltimore region, suggesting that patterns found in this county are likely to be similar in the rest of the region.[i]

Figure 1 about here

Figure 1 shows census tract level median household income in Baltimore City and County from the 2010-2014 American Community Survey. The tract level shading indicates that central Baltimore County has very high-income residents with median incomes above 100,000 dollars. These areas have much higher income levels than most of the city, or even the sides of the county. The lowest-income parts of the region are located towards the middle of Baltimore City, but there are also some affluent neighborhoods in the north and around the harbor.

Figure 2 about here

Figure 2 shows the tract-level percent black. Here the residential segregation in the region becomes apparent. Tracts tend to be either predominantly black or have relatively few black residents. Baltimore City's residential segregation pattern follows what is known locally as the "Butterfly and the L": the Eastern and Western sides of the city are predominantly black with white neighborhoods in the middle extending to the east of the harbor. In Baltimore County there are far fewer predominantly black neighborhoods. The western wing of the County has a string of predominantly black neighborhoods that lead outwards from the city. These and a single tract along the eastern side of the northern boundary are the only sections of the county in which there are substantial numbers of black residents. The rest of northern and eastern Baltimore County are overwhelmingly white with a few pockets of Hispanic and Asian residents along the highway that heads north-south down the middle of the county.

Together, these two school districts serve more than 92,000 students in grades K-5 each year. While all students in the region are assigned a default elementary school based on residential address, both districts allow for some degree of non-residential assignment, either through magnet programs in the County or charter schools and out-of-zone enrollment in neighborhood schools in the City (BCPS 2020; BCPSS 2018). [ii] These programs mean that school enrollment decisions are not dictated exclusively by residential patterns and provide some room for parental decision-making based on school characteristics alone.

Data

The data for this study come from de-identified administrative records from the Baltimore City and County Public Schools from 2010-11 through 2014-15 that are stored at the Baltimore Education Research Consortium. The data include every child ever enrolled in either district during this time period, along with ids for every school attended and the date of record for each school enrollment and withdrawal. Students are matched across districts using a state identifier generated by the Maryland State Department of Education (MSDE).

We focus our analysis on regular elementary schools, as defined by National Center for Education Statistics (NCES). This excludes students who are assigned to special education or alternative schools by either district and whose mobility patterns, therefore, do not represent informal flows of information in the same way. The population of elementary schools include all schools with at least one grade between kindergarten and fifth grade. These schools tend to have smaller enrollments, serve smaller geographic catchment areas, and have higher mobility rates than middle or high schools (MSDE 2015).

School-level data comes from the NCES Common Core of Data (NCES 2015) and the MSDE School Report Cards (MSDE 2015). NCES reports the aggregate racial composition and free-meals status of all schools in the country. They also provide geocoded school addresses. The MSDE reports include attendance rates, special education and English language learner status, as well as standardized test score proficiency rates. We will use the 2014-15 school year to describe the schools. It should be noted that 2014-15 was the first year that Maryland adopted the Common Core-aligned test (the Partnership for Assessment of Readiness for College and Careers [PARCC]) and pass rates on standardized tests are lower than in previous years. Nevertheless, the relative differences in test performance between schools are comparable to other years.

The first three columns of Table 1 describe these measures for all schools and for the city and county districts separately. Means and standard deviations are weighted by the number of students in each school. Each district contains approximately the same number of public-school students in grades K-5. Twenty-six percent of students in the two districts combined are white. Fifty-eight percent are black, 9 percent are Hispanic, and 4 percent are Asian. Baltimore County has a much larger proportion of white students than the city (43 versus 10 percent) and fewer black students (36 versus 79 percent). A substantial proportion of all students are eligible for free- and reduced-price meals, including 87 percent of students in the city, but only around half of students in the county. Attendance and achievement scores are also higher in the suburbs than the city, but special education and English language learner status are similar across the two districts.

Table 1 about here

Our models also include measures of social and geographic separation between schools. First, we include an indicator for whether the sending and receiving school are in different districts. Administratively, it is harder to make cross-district school changes since they require a simultaneous change in residential address. Families are also likely to rule out an entire district in the early stages of decision-making.

Similarly, charter schools tend to appeal to distinct sets of families (Posey-Maddox, Kimelberg, and Cucchiara 2014). Parents that have opted out of traditional schools may be more likely to keep their students in that sector even when changing schools. Therefore, we include an indicator for whether the sending and receiving schools are different types: charter versus traditional. This variable has a value of zero if both schools are traditional or both schools are charter schools.

Just as districts and school sector serve different sets of families, major roads and interstates have been used historically to divide communities. Alexander et al. argue that especially in Baltimore City, major roads and parks were used to delineate "defended communities" where white residents who decided to remain during periods of white flight in the 1980s and 1990s were able to exclude blacks (2014:41–42; Green, Strolovitch, and Wong 1998;

Suttles 1972). The region's interstates can be seen on Figures 1 and 2. They often appear as dividing lines between census tracts with very different demographic compositions. In order to capture these social dividing lines, we create an indicator for whether there is a major road or interstate between two schools. If one of these roads intersects the straight line drawn between schools, we consider it to be on the path between two schools. This does not take into account whether traveling along a highway would be necessary to get from one school to another, but indicates that schools fall on different sides of the road on a map. To capture additional non-linearities in distance, we also include an indicator for whether each receiving school is the closest school to each sending school.

Methods

Building a Network of School Ties

The hard to measure and individually-specific nature of choice sets make them difficult to model on a large scale, which has limited their utility in empirical studies of decision-making and segregation. (For more detail on the limits of existing methods of estimating and incorporating choice sets see Appendix A). In this study, we propose a novel method that can be applied to any large population dataset or highly-saturated survey. Rather than make assumptions about what families use to limit their consideration sets or rely on active search behavior and stated preferences, we use the observed connections and flow of students between schools to estimate the emergent structure of consideration sets. These structures are based entirely on the actual patterns of flows between schools and are not limited by researchers' assumptions of how families behave or what families say they look for.

In order to estimate the emergent structures in school enrollment, we first create a network of ties between schools established by student mobility. Specifically, we calculate directional flows of elementary students from each school in both districts to every other school during the summer across all five academic years. We focus on summer moves only because these are the moves that are most likely to be related to school characteristics rather than personal family hardship (Welsh 2017).[iii] We include students enrolled in kindergarten through fifth grade and avoid all promotional school moves (i.e. moves required because the current school closed or does not offer higher grade-levels). This means that all moves are the result of individual decisions rather than administrative features of the districts.

We use summer school transfers to build the network for four reasons. First, by virtue of their change in enrollment, these families are "voting with their feet" and making an active decision to seek out a different school environment. These movers are exactly the kind of "active chooser" that Buckley and Schneider (2003) argue are most likely to gather information about their schooling options. This makes their behavior especially important for understanding the logic of overall enrollment patterns.

Second, school transfers are often related to but do not always coincide with residential moves. Nationally, only about half of school moves are the result of a residential move (Gasper, DeLuca, and Estacion 2010). Therefore, examining school movers allows us to look at enrollment patterns that are influenced by, but not dictated directly by, residential segregation. When looking at initial enrollment it is harder to separate earlier residential decision-making from school enrollment decisions (DeLuca, Darrah-Okike, and Nerenberg 2018).

Third, while students may change schools for a wide range of reasons, some reactionary and other strategic, research shows that where they go is often quite similar to where they left (Kerbow 1996; Welsh 2017). This is because their destination decisions are likely structured and constrained in many of the same ways as their initial enrollment decision. Even if the stated reason for a school change is the need to move in with a relative for financial reasons, where that relative lives tells us something about that student's social and geographic place in the region and the options that similar families might also consider. Given the similarity in the origin and destination schools, there is no reason to think that the structure of choice sets and final decision-making would be different for initial enrollment and transfers.

Finally, we know that all decisions are "anchored" by past experience (Furnham and Boo 2011). Fortunately, in the case of movers, that past experience is easily observed by the researcher. This means that we are able to model relative change in observed characteristics between sending and receiving schools rather than compare across all schools. In other words, we can observe whether schools with low proficiency rates but high in-mobility rates are receiving students from even lower achieving schools. This would be evidence that families value achievement even if they do not end up in the highest achieving school in the district.

Table 2 about here

Around one quarter of all students ever make a summer move. Rates of mobility are somewhat higher in the City than in the County (28 percent vs 18 percent). Table 2 compares the characteristics of students who ever make a school transfer with those who stay stably enrolled during the five-year observation period. Mobile students are substantially more likely to be black and come from schools with somewhat higher proportions of free- and reduced-price meal-eligible students, but with comparable proficiency ratings.[iv]

Defining Emergent Consideration Sets

In order to empirically derive emergent structure from the network of schools, we apply the Louvian method of community detection (Blondel et al. 2008) to the network of school transfers with directed ties, weighted by the number of movers using Gephi 0.9.1. This method starts by optimizing on specific randomized nodes to form small groups and then iterates this process to combine them into larger groups. The final number of groupings is derived by optimizing the fraction of ties within groups compared to ties between groups. The extent to which a network is divided into more groups than would appear if the connections were established at random is known as *modularity*. The maximum value of the modularity index is 1 and the minimum is -1.

The Louvian method of community detection was chosen for a few reasons. First, this method of community detection does not require any prior assumptions about the nature of the network graph or the size, density and shape of the clusters. This means that we are able to let the data generate patterns without having to impose any pre-conceived notions about the role of school characteristics or location. Second, the Louvian algorithm is a “bottom-up” method, meaning that it aggregates individual ties to build hierarchical structures that allow us to examine emergent consideration sets at different levels of clustering. The random starting point for the algorithm means that running it multiple times allows us to measure the probability that any two schools are in the same cluster. Finally, this method is popular and widely available in existing network software and can therefore be easily adopted by others interested in replicating these methods with other data sources.

The starting point for the optimization is randomly selected each time the modularity command is run. Starting with a different random seed may result in a different optimized solution for the same network graph, especially if the divisions between clusters are not clear cut. In order to account for this randomness and to provide a probabilistic rather than deterministic measure of the consideration set, we run the modularity algorithm 100 times with 100 randomly selected starting points.[v] For every school tie, we calculate the number of times out of 100 that those two schools are in the same subset. We consider this to be an approximation of the likelihood that these two schools would be considered by similar families. We do not claim that this means that these are the only schools that specific families would consider, only that in the aggregate, these are the ways consideration sets tend to be structured. Since individual choice sets tend to be small, we create a second, more restrictive measure that is a dichotomous variable for whether or not a school tie is always in the same emergent cluster in all 100 of our runs.

Structure of Emergent Consideration Sets

As a comparison for our network-derived consideration sets, we also create alternative consideration sets determined by observed school characteristics alone. Since research shows that parents are heavily influenced by distance, proficiency, and school composition, we consider two schools to be in the same naïve consideration set if their racial composition and proficiency levels differ by no more than 20 percentage points and they are no more than three miles apart. To test whether the emergent consideration sets capture something beyond demographics and distance (hypothesis 1), we use ordinary least squares regression to predict the number of moves along each school tie using only a dummy variable for belonging to the same the simple demographic consideration sets, our probabilistic measure of emergent consideration sets, or our dichotomous measure of always being in the same consideration set. Comparing the predictive power (R-squared) of these different definitions of consideration sets will allow us to assess whether emergent patterns and unobserved characteristics drive consideration sets more than observed school characteristics.

To further explore the role of socially structured consideration set formation, we build a model that predicts the number of times that two schools will be included in the same network-derived consideration set in any of the 100 runs.[vi] A tie between schools does not define an emergent consideration set on its own. However, examining what predicts that two schools will be included in the same set provides important information about what school-level factors drive the clustering and segmentation of school enrollment patterns. Specifically, we compare the predictive power of demographics and proficiency alone to that of structural divisions between schools, including distance, district boundaries, different school types (charter vs traditional), and location across major roads and interstates. Since families tend to rule out large swaths of a district or metropolitan area based on categorical markers during the editing phase, we expect that these structural barriers will be much more predictive than school demographics and proficiency (hypothesis 2). The model is as follows:

$$I_{ij} = \beta_0 + \beta_1 Abs(T_{ij}) + \beta_2 Abs(Blk_{ij}) + \beta_3 Abs(Hsp_{ij}) + \beta_4 Abs(Frm_{ij}) + \beta_5 Log(Dist)_{ij} + \beta_6 Closest_{ij} + \beta_7 Chart_{ij} + \beta_8 Cross_{ij} + \beta_9 Road_{ij} + e_{ij} \quad (1)$$

Where I_{ij} is a count of the number of times the schools i and j were in included in the same set in each of the 100 runs; T_{ij} is the difference in math standardized proficiency rates; Blk_{ij}, Hsp_{ij}, and Frm_{ij} are the differences in the percent of black, Hispanic, and free and reduced lunch eligible students; $Dist_{ij}$ is the distance between schools in miles; $Closest_{ij}$ is an indicator for whether either school is the closest receiving school for the other in the pair; $Chart_{ij}$ is an indicator for whether the schools differ in charter status; $Cross_{ij}$ indicates whether the two schools are in different districts; $Road_{ij}$ indicates whether there is an interstate or major state road between the two schools; and e_{ij} is the tie-level error term. The unit of analysis is the pair of schools, without any directionality: the pair of schools A and B only appear as one observation, not separately as A to B and B to A. Therefore, we measure all differences in school characteristics between the two schools as absolute values. The coefficients can be interpreted as the degree to which differences between schools predicts higher or lower numbers of inclusions in the same set. We expect the coefficients for the demographic difference to be negative. This would indicate, for example, that the larger the difference in the percent of black students between the two schools, the less likely they are to be in the same cluster. While the coefficients in and of themselves are interesting, the focus of the hypothesis in this section is on the increase in predictive power that comes with adding structural constraints to the model.

Logic of Mobility Flows

We then turn to predicting the actual movement of students, given the structure of the consideration sets. The focus is on how our understanding of the logic of mobility changes when we incorporate emergent consideration sets into the analysis.[vii] To do this, we predict the size of the flow of students between two schools using a gravity model. These models were developed in the context of international economics to measure the volume of trade between countries based on their size and distance from one another (Sen and Smith 2012). In this context, we are estimating the volume of student movement between schools and our outcome is the number of students who make a specific move in a specific direction. We then use the difference in observed characteristics between the sending and receiving schools to understand the overall pattern of mobility.

Unlike the emergent structure analysis, we include directionality here. In other words, for every school tie there are two observations: one from school A to B and another from B to A. Therefore, measures of demographic difference also include directionality. Positive values mean that the sending school has a higher value than the receiving school, while negative values indicate the reverse. We include measures of student demographics, proficiency, geographic distance, and the structural boundaries used in the analysis of consideration sets. Similar to traditional gravity models, we adjust for measures of size and overall instability that are likely to predict larger flows regardless of school characteristics. Our formal model is as follows:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 Blk_{ij} + \beta_3 Hsp_{ij} + \beta_4 Frm_{ij} + \beta_5 Log(Dist)_{ij} + \beta_6 Closest_{ij} + \beta_7 Chart_{ij} + \beta_8 Cross_{ij} + \beta_9 Road_{ij} + \beta_{10} Size_i + \beta_{11} Size_j + \beta_{12} Out_i + \beta_{13} In_j + e_{ij} \quad (2)$$

Where Y_{ij} is the number of moves from school i to school j; T_{ij} is the difference in math test score proficiency rates, are Blk_{ij} and Hsp_{ij} are differences in racial composition; Frm_{ij} is the difference in the proportion of free and reduced meals recipients; $Log(Dist)_{ij}$ is the logged miles between schools as the crow flies; $Closest_{ij}$ is a binary indicator for whether receiving school j is the closest school to sending school i; $Chart_{ij}$ indicates different school sectors (charter versus traditional); $Cross_{ij}$ indicates whether the schools are in different districts; and $Road_{ij}$ indicates whether they are on opposite sides of a major road or interstate. We adjust for the size of both schools ($Size_i$ and $Size_j$) and the overall level of connectedness of each school to the rest of the network via the total number of exits for the sending school (Out_i) and the total number of entries for the receiving school (In_j). Robust standard errors are used to account for any heteroscedasticity in the tie-level error term (e_{ij}).[viii]

A positive value for a coefficient means that, on average, more students move to schools with higher values of that measure than their starting school. A negative coefficient means the reverse. In other words, if achievement scores are a strong attractor, we would expect students to be more likely to move to a school that has higher achievement scores than their current school (the coefficient for T_{ij} to be positive). If student composition, particularly race and poverty, are strong deterrents, we would expect students to be less likely to depart their school for one with greater numbers of poor and minority students (therefore the coefficients for Blk_{ij}, Hsp_{ij}, and Frm_{ij} would be negative). If convenience in the form of travel time is important, we would expect schools that are close together to be more likely to share students and the coefficient for $Log(Dist_{ij})$ to be negative. We do not claim that measures included in these models represent an exhaustive list of factors that could be important to families. There are certainly many other unobserved reasons that a specific family might choose to change from one school to another. However, these are the measures that are most frequently cited in the literature on school preferences and their relative importance across models reveals valuable information about the sources of school segregation.[ix]

To compare the influence of emergent consideration sets on decision parameters we run this model several times with different sample restrictions. Our first set of estimates allows any school to have an equal probability of sending students to any other school. In this case, all potential moves between schools are included, even if there are no observed moves along that path. Our next two models incorporate two different measures of the emergent consideration sets in order to limit the scope of realistic options for parents leaving a particular school. In the first version, we weight each observation (a directional tie between two schools) by the probability that those two schools are included in the same subset across all 100 runs. Ties between two schools that are never included in the same subset have a weight of zero and are, therefore, excluded from the analysis. Ties between two schools that are always in the same consideration set have a weight of one and the largest influence on the estimated parameters. Our third and most restrictive model limits the emergent consideration sets to only those pairs of schools that are always in the same cluster across all 100 runs. This narrows the consideration sets to an even greater degree and provides smaller, more realistic limitations on the network of schools.

The difference in the R-squared for these three models will indicate whether incorporating emergent consideration sets improves the prediction of mobility flows (hypothesis 3). The difference in the estimated strength of the coefficients between these three models will demonstrate how taking into account complex decision-making processes shapes our understanding of school choice and segregation processes (hypothesis 4).

Results

Building a Network of School Ties

Over 42,000 summertime school moves were made by kindergarten through fifth grade students between regular elementary schools during this five-year period. Approximately 25,000 of these moves originated from a Baltimore City school and 17,000 of these moves originated from a Baltimore County school. The majority of school moves remained within the same district. Only around 20 percent of moves from a city school ended up in the county and around 15 percent of the moves from a county school ended up in the city.

The ties between schools created by mobile students follow a highly skewed distribution and create a dense network, as shown in Figure 3. There are no isolates; every school is connected to at least one other school during this period. When two schools are connected by mobility, the average flow is approximately 13 student moves, but around half of all ties between schools represent only one student move. On average, each school is connected to 76 different schools by student exits and 71 different schools by student entries, representing about 30 percent of the region's 236 elementary schools. As Figure 3 shows, the level of activity in the network makes it difficult to visually discern any structure.

Figure 3 about here

Defining Emergent Consideration Sets

The sheer density of ties in the network may be obscuring underlying patterns. Therefore, we use Louvian community detection algorithms to parse the network in ways that the naked eye might not be able to detect. The modal solution contains seven subsets (55 percent of runs), with a minimum of six (8 percent) and a maximum of nine (4 percent). The degree to which a network easily divides into distinct subgroups is often measured by the modularity index. This index represents the fraction of edges that fall within the given groups minus the expected fraction if edges were distributed at random and can range from -1 to 1. The average modularity of the estimated subgroups is 0.34 (s.d. 0.004). This means that schools are more clustered than if ties were distributed at random, but that student mobility also establishes quite a few connections between schools in different clusters.

Another way to examine the distinctiveness of the subsets is to look across all 100 runs. Specifically, we examine the number of times each of the 27,730 potential school ties show up in the same cluster. The distribution of cluster ties across runs highlights three things: first, there are many zeros. More than 20,000 school ties (72.4 percent) never fall into the same subset. We interpret this to mean that many schools are very unlikely to be jointly considered by similar families. This aligns well with the literature on decision-making that suggests that families only actively consider a very small set of schools. Second, there is a moderate number of ones. Approximately 1,250 ties (4.5 percent) show up in the same subset in every single run. These schools are very likely to be jointly considered by similar families when choosing a school or making a school transfer. Finally, around 6,300 school ties (23 percent) fall somewhere in between and fall into the same cluster only a fraction of the time. We interpret this to mean that there is some chance that a family would consider these two schools at the same time, but that not everyone will do so. Given the dense network shown in Figure 3, this uncertainty is expected and is why it is important not to rely just on one run of the modularity algorithm when the underlying network is dense.

Figure 4 shows what the emergent consideration sets look like on a map. Lines in red represent ties between two schools that are always in the same cluster. These groups are very small geographically and tend not to include large numbers of schools and never cross the district boundary. Orange lines represent ties between schools that were in the same subset in 75-99 percent of the runs. Yellow lines represent ties between schools that were in the same subset more than half the time, but less than 75 percent of the time. As the colors get lighter the geographic area of the clusters gets larger. Smaller clusters also combine to form large clusters that cover a lot geographic territory and occasionally cross the district line. However, even with this level of uncertainty, there are clear dividing lines between emergent consideration sets. For example, there is an east-west dividing line that begins at the harbor in Baltimore City, extends northwest along a highway and continues through Baltimore County.

Figure 4 about here

Structure of Emergent Consideration Sets

In order to better understand what drives these emergent network clusters, we turn to our two consideration set hypotheses (hypothesis 1 and 2). First, we will compare our emergent cluster definitions to simpler ties driven only by school demographics and distance. In this naïve counter-example, schools are considered to be in the same demographic consideration set if they differ by no more than 20 percentage points in their racial and economic composition or their proficiency rates and are less than three miles apart.[x]

Table 3 describes the overlap between these simple demographic ties and our emergent clusters. The table shows a cross tabulation between whether two schools have similar demographics and distance and whether they are always or never considered in the same cluster. Although there are a relatively similar number of ties identified by each definition, there is substantial disagreement about which schools should share a tie. Only around a quarter of ties that are demographically similar and relatively close together are identified as always in the same cluster by our community detection algorithm and only around 20 percent of the "always" ties are similar enough to have ties based on demographics and distance. The ties that are never in the same cluster show a similar pattern. One third of ties that are not demographically similar have some chance of being in the same cluster and around 12 percent of demographically similar ties never appear in the same cluster.

Table 3 about here

If families only consider schools that are nearby and demographically similar, we would expect that most moves should take place along these school ties. This does not appear to be the case. These demographic ties account for approximately 4 percent of all school ties and approximately 18 percent of moves. In comparison, the schools that are always in the same emergent consideration set account for approximately 4.5 percent of all school ties, but around 31 percent of all moves. In other words, more students flow along the paths derived by our emergent consideration sets than along demographically similar schools. Similarly, the demographically-defined consideration sets are very poor predictors of flows across the network. The R-squared predicting the total number of moves along each school tie with an indicator of a demographic tie is only 0.05, while using an indicator for whether two schools are always in the same consideration set is substantially larger at 0.17. If we use the continuous measure of the probability of consideration set inclusion, the R-squared increases to 0.23. Together this suggests that the consideration set definitions derived from the network clustering algorithms better describe the observed mobility patterns than demographics alone (hypothesis 1).

Our second consideration set hypothesis (hypothesis 2) has to do with the importance of socially structured constraints and barriers. To test this hypothesis, we focus on what predicts that two schools would be included in the same network cluster. We know from the previous analysis that demographics and proficiency alone are not very good predictors, but what about other kinds of structural and geographic divisions? Since families are known to rule out large swaths of a district or metropolitan area based on very little information, we expect that families will be less likely to consider schools that are in another district, of a different school type, or on the other side of an interstate or major road (hypothesis 2).

To test this hypothesis, we estimate an ordinary least squares regression that predicts the number of times (out of 100) that two schools appear in the same emergent consideration set. Since there is no directionality in this measure, each school tie is only included once. The results are shown in Table 4. The first model includes only demographics and academics. As expected, schools that are more similar are more likely to be included in the same emergent consideration set (all of the coefficients are negative), but the predictive power of the model is quite low (only 0.10). In contrast when we add in the structural constraints (model 2) the predictive power jumps to 0.37. A formal test of these nested models yields an F statistic for the 5 additional parameters with 27720 degrees of freedom of 1923.39, which is highly statistically significant at the less than 0.001 level. This means that we can easily reject the null hypothesis that these additional measures add nothing to the model.

The structural boundary coefficients are also much larger than the demographic coefficients. For example, a 10 percentage point difference in the proportion of black students decreases the predicted number of times two schools are included in the same consideration set by just 1, but being in different districts predicts 16 fewer times and being on different sides of a major road predicts 8 fewer times. Together, this highlights the importance of the social structural sorting process described by Krysan and Crowder (2017) with housing and suggests that large groups of schools that are either too far away or of the wrong type are never considered by families, regardless of their observed characteristics.

Table 4 about here

Notably, the coefficient for relative proportion of free- and reduced-price meal eligible students switches signs when the structural controls are introduced. This suggests that much of the differentiation along socioeconomic lines is driven by high levels of income segregation and that within smaller areas of the city, students may actually be more likely to transfer between schools with different levels of free and reduced-price meals eligibility. Now, it must be noted that we do not include direction in these models and cannot yet tell whether the patterns are driven by families that are always headed from schools with poorer populations to those with higher income populations.

Table 5 about here

It is important to note that despite the structure of these emergent consideration sets being driven largely by social and geographic boundaries, the results are still relatively racially homogeneous. Table 5 shows the Theil index of multi-group segregation for racial composition, the proportion of students eligible for free- and reduced-price meals, and proficiency for between all schools and between the emergent consideration sets (Reardon and Firebaugh 2002). Of the three measures, students are most highly segregated by race. Our emergent consideration sets explain around half of the overall segregation across schools in the area, on average, across all 100 runs. When we use the most restrictive definition of a consideration set and create distinct sets of schools that are always in the same set (the red lines in Figure 4) we explain around three quarters of the total between-school segregation in both districts. This means that there is relatively little variability by race, and to a lesser extent socioeconomic and proficiency status, between schools in the same consideration sets. It will be important to keep this in mind when we predict the flow of students within these sets.

Logic of Mobility Flows

Now that we have a working definition of emergent consideration sets, we move on to assess how incorporating those restrictions into our models changes our understanding of the logic driving mobility flows. Table 6 compares two sets of three gravity models that make different assumptions about the range of schools parents are likely to consider. The outcome is the number of students who make a specific directional transfer and there are two observations per school tie, one for moves from school A to school B and one for school B to school A. The first model in each set gives all potential flows equal weight and makes the naïve assumption that parents consider every other school in the area. Schools that do not share any students simply have an outcome of zero. The second model limits the observations to ties that occur along the probabilistic consideration sets derived above. Each directional tie is weighted by the likelihood that the two schools will be included in the same emergent consideration set. The approximately three quarters of ties with no chance of ever being considered together are excluded from the analysis. (Note the difference in the number of observations in each model). The final model uses the smallest definition of emergent consideration sets: only ties of schools that are always in the same subset. This is the most restrictive sample and only includes around 2,496 ties, but is more likely to represent realistic consideration sets for similar types of parents.

Each model adjusts for school size and total entries and exits so that we do not confound high levels of instability at some schools with the characteristics of the school those students move to. Since the ties in this analysis are directional, the coefficients represent the difference in characteristics between the sending and receiving schools. For example, the math coefficient is positive when more moves happen along ties where the receiving school has a higher math proficiency rate than the sending school. If parents are looking to improve their child's academic context, we would expect that coefficient to be positive. Similarly, if we expect families to move to schools with lower proportions of minority or poor students, we would expect those coefficients to be negative.

Table 6 about here

The first set of models (Model 1) includes only the demographics and proficiency levels of the schools. We exclude the various measures of structural and geographic distance since they are very strong predictors of the emergent consideration sets (see Table 4). The first thing to note is the difference in each model's predictive power (hypothesis 3). The R-squared for the naïve unweighted regression is only 0.05, while the R-squared for the models that incorporate emergent consideration sets is 0.16. This suggests that the observed characteristics of schools are better predictors of student flows when we take into account more realistic parental consideration sets. A similar pattern, although less dramatic, can be seen in the second set of models (Model 2) that include the measures of social and geographic distance. Here the change between models is smaller because the additional measures capture much of what defines the emergent consideration sets in the first place. In fact, adding these measures only improves the predictive power of the "always" model by a small amount (0.16 vs 0.19).

The second point of interest in Table 6 is the pattern of the coefficients themselves. In each set the pattern is the same: the coefficients for the observed characteristics of schools become larger as we limit the analysis to more realistic consideration sets (hypothesis 4). This change is most dramatic for free- and reduce-priced meals eligibility percentages, where the magnitude of the coefficients go from around -0.02 in the naïve model to -0.36 in the most restrictive definition of emergent consideration sets. This means that students are much less likely to move to a school with a higher percentage of free- and reduce-priced meals eligible students within the emergent consideration sets than when all other schools in the district are included. The change for proficiency levels is similarly large (0.04 to 0.33). Again, this means that when we take into account likely consideration sets, students seem much more sensitive to the relative proficiency levels of the sending and receiving schools. The overall pattern of the racial composition measures is similar, but the coefficients are much weaker. This means that at this stage, families are less sensitive to differences in racial composition than income or proficiency.[xi]

Discussion and Conclusion

This study proposes and applies a novel method for understanding school segregation dynamics and the role of social and geographic structure in student enrollment patterns. To do so, we begin with the premise that not every family will consider every school and that aggregate flows of students between schools can reveal emergent differences between sets of schools that are likely to be considered together. We then use community detection methods to reveal emergent structures in a network of schools defined by the flow of mobile students. We interpret the clusters of schools derived from these analyses as the set of schools likely to be considered by families in similar social and geographic positions.

We find that our network-derived emergent consideration sets are able to pick up structural and geographic divisions in the region that are based on more than just student demographics and straight-line distance. For example, two schools on opposite sides of a major road are very unlikely to be in the same emergent consideration set regardless of their student composition or how far apart they are. Given the historic patterns of residential segregation in the Baltimore region, these structural boundaries tend to produce relatively racially homogeneous subsets of schools.

Next, we examine the role of these social divisions in the logic of student mobility flows between schools. Here, we find that models are better at predicting flows of students when they take these emergent consideration sets into account than naïve models that include every possible school. Finally, within emergent consideration sets, students flow away from schools with relatively higher levels of free- and reduced-price meal eligibility and towards schools with higher proficiency levels. School racial composition, however, plays a much smaller role in predicting movement between schools, in part because schools in the same consideration sets are likely to be quite similar on this measure.

These findings have important implications for our understanding of school segregation processes and the policies used to combat them. First, the findings highlight the intractability and self-reproducing nature of segregation and the role of sometimes subtle structural boundaries in that sorting process. Specifically, building on work by Bruch and Swait (2019) and Krysan and Crowder (2017), this study finds that patterns of consideration set formation are heavily dependent on the social and geographic distribution of racial groups in the region. Since parents of different races often do not interact with each other in the same social or geographic circles, they do not end up considering the same places and, therefore, have little chance of ever enrolling in the same schools, even when looking for the same objective characteristics. The specific geography and demographics of emergent consideration sets estimated in this study may be particular to the Baltimore region, but the theory applies everywhere. Past experiences and existing residential segregation patterns lead families to very different schools even without a clear preference for same-race peers. Reducing school segregation is going to take more than widening access to schools through open enrollment policies, we also have to remove the geographic and social boundaries that shape patterns of interaction and geographic exposure.

These findings can help explain the persistent pattern of higher racial segregation between rather than within school districts (Reardon, Yun, and Eitle 2000). District boundaries play a strong role in limiting our emergent consideration sets. In this study, none of the sets of schools that are always in the same cluster include schools in different districts. We interpret this to mean that districts are "deal breakers" (Bruch et al. 2016) in families' decision process, and that families tend to rule out entire districts without actively considering the specifics of schools within excluded districts. While district reputations are certainly associated with their racial composition, this process is different than families actively avoiding specific schools with concentrations of racial minorities. It also means that policies and programs designed to expand families' consideration sets to include different districts could lead to more school integration (Bergman et al. 2019).

Second, this study adds to a growing body of literature that suggests that school consideration sets are quite limited. In other words, parents are not looking carefully at the whole district or region when making enrollment decisions. Instead, they are likely weighing only a few options and choosing the best fit from within that limited set. In this context, the competition for students does not take place at the district or regional level, but at the local level. This is similar to the notion of local maximums versus global maximums that create niches in the ecology literature (Martin and Wainwright 2013). Taking these limited consideration sets into account yields different conclusions about what parents are looking for in a school. Compared to all schools in the area, they may not end up in the highest achieving school, but given their local options, there is a clear preference for higher proficiency rates. In other words, parents' stated preferences for high quality schools are true, but their definition of high quality is often relative, not absolute.

The fact that students often attend schools close to home has not gone unnoticed in the education policy world, but to date, the focus has been largely on travel time and cost (Blagg, Rosenboom, and Chingos 2018; Glazerman and Dotter 2017). The findings described in this study suggest that the decreased likelihood of attending a school far away is not simply about how hard it is to get there, but also whether it even occurs to parents to consider going there. This can be seen in the fact that adding measures of logged distance and other structural boundaries do not substantially improve the prediction of mobility flows within emergent consideration sets. Moreover, the geographic size of the different consideration set clusters is quite variable. In some parts of the region, the clusters are very small and divide relatively small geographic areas. In other parts of the region, schools that are much farther apart are still always considered part of the same cluster. In other words, distance matters to the degree that it predicts social divisions. It is not that families thought about a school and decided it was too far away. Instead, their own social and spatial experiences likely kept them from considering the school at all. Therefore, improving transit to schools is unlikely to do much on its own. If we really want to improve access to integrated schools, we also need to improve social integration and generate positive experiences across a larger number of areas in the city.

Finally, the methods and findings of this study do not apply only to school enrollment. The importance of social structure in limiting consideration sets applies to any type of decision, including many that we often think of as based solely on preferences, such as where to eat, where to shop, where to live, or who to date. Therefore, the methods developed in this paper could be used with data from other areas. Our finding that these social divisions are not simply defined by demographics, but also by the social and geographic structure of the city, is likely to hold for other types of sorting. Major roads, for example, are likely to serve as a dividing line for other types of social and spatial behavior. One of the key benefits of using network ties to reveal the structure of consideration sets, however, is that we do not need to measure all of the little things that generate social divisions. We can instead rely on the observed movement of people to reveal those dividing factors.

The increasing availability of population-level administrative data has yet to be fully exploited within sociological analysis. This study demonstrates that these datasets provide the opportunity to illustrate more than what is available in the variable list (which is often thin). By examining the links between units and the emergent network structures that they form, we can learn a lot about the socially structured sorting processes that drive other forms of behavior and inequality.

References

Adair, Alastair S., JN Berry, W. Stanley J. McGreal, B. Murtagh, and C. Paris. 2000. "The Local Housing System in Craigavon, N. Ireland: Ethno-Religious Residential Segregation, Socio-Tenurial Polarisation and Sub-Markets." *Urban Studies* 37(7):1079–92.

Alexander, Karl, Doris Entwisle, and Linda Olson. 2014. *The Long Shadow: Family Background, Disadvantaged Urban Youth, and the Transition to Adulthood*. Russell Sage Foundation.

Altenhofen, Shannon, Mark Berends, and Thomas G. White. 2016. "School Choice Decision Making among Suburban, High-Income Parents." *AERA Open* 2(1):2332858415624098.

Bader, Michael DM, Annette Lareau, and Shani A. Evans. 2019. "Talk on the Playground: The Neighborhood Context of School Choice." *City & Community* 18(2):483–508.

Ball, Stephen J., and Carol Vincent. 1998. "'I Heard It on the Grapevine':'Hot'Knowledge and School Choice." *British Journal of Sociology of Education* 19(3):377–400.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115(4):588–638.

BCPS (Baltimore County Public Schools). 2020. "Student Registration and Moving Tips." https://www.bcps.org/schools/moving_tips.html.

BCPSS (Baltimore City Public Schools). 2018. "Comprehensive Educational Facilities Master Plan (CEFMP)." https://www.baltimorecityschools.org/sites/default/files/2019-01/CEFMP-2018.pdf.

Bell, Courtney. 2009. "All Choices Created Equal? The Role of Choice Sets in the Selection of Schools." *Peabody Journal of Education* 84(2):191–208.

Bell, Courtney A. 2007. "Space and Place: Urban Parents' Geographical Preferences for Schools." *The Urban Review* 39(4):375–404.

Benard, Stephen, and Robb Willer. 2007. "A Wealth and Status-Based Model of Residential Segregation." *Mathematical Sociology* 31(2):149–74.

Benenson, Itzhak, Erez Hatna, and Ehud Or. 2009. "From Schelling to Spatially Explicit Modeling of Urban Ethnic and Economic Residential Dynamics." *Sociological Methods & Research* 37(4):463–97.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F. Katz, and Christopher Palmer. 2019. *Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice*. 0898–2937. National Bureau of Economic Research.

Bifulco, Robert, and Helen F. Ladd. 2007. "School Choice, Racial Segregation, and Test-score Gaps: Evidence from North Carolina's Charter School Program." *Journal of Policy Analysis and Management* 26(1):31–56.

Billingham, Chase M., and Matthew O. Hunt. 2016. "School Racial Composition and Parental Choice: New Evidence on the Preferences of White Parents in the United States." *Sociology of Education* 89(2):99–117.

Blagg, Kristin, Victoria Rosenboom, and Matthew M. Chingos. 2018. "The Extra Mile: Time to School and Student Outcomes in Washington, DC. Research Report." *Urban Institute*.

Blau, Peter M., Terry C. Blum, and Joseph E. Schwartz. 1982. "Heterogeneity and Intermarriage." *American Sociological Review* 47(1):45–62.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. “Fast Unfolding of Communities in Large Networks.” *Journal of Statistical Mechanics: Theory and Experiment* 2008(10):P10008.

Bowe, Richard, Stephen Ball, and Sharon Gewirtz. 1994. “‘Parental Choice’, Consumption and Social Theory: The Operation of Micro-markets in Education.” *British Journal of Educational Studies* 42(1):38–52.

Browning, Christopher R., Catherine A. Calder, Jodi L. Ford, Bethany Boettner, Anna L. Smith, and Dana Haynie. 2017. “Understanding Racial Differences in Exposure to Violent Areas: Integrating Survey, Smartphone, and Administrative Data Resources.” *The Annals of the American Academy of Political and Social Science* 669(1):41–62.

Bruch, Elizabeth, and Jon Atwell. 2015. “Agent-Based Models in Empirical Social Research.” *Sociological Methods & Research* 44(2):186–221.

Bruch, Elizabeth E. 2014. “How Population Structure Shapes Neighborhood Segregation.” *American Journal of Sociology* 119(5):1221–78.

Bruch, Elizabeth E., and Robert D. Mare. 2006. “Neighborhood Choice and Neighborhood Change.” *American Journal of Sociology* 112(3):667–709.

Bruch, Elizabeth E., and Robert D. Mare. 2009. “Preferences and Pathways to Segregation: Reply to van de Rijt, Siegel, and Macy.” *American Journal of Sociology* 114(4):1181–98.

Bruch, Elizabeth E., and MEJ Newman. 2018. “Aspirational Pursuit of Mates in Online Dating Markets.” *Science Advances* 4(8):eaap9815.

Bruch, Elizabeth, Fred Feinberg, and Kee Yeun Lee. 2016. “Extracting Multistage Screening Rules from Online Dating Activity Data.” *Proceedings of the National Academy of Sciences* 113(38):10530–35.

Bruch, Elizabeth, Ross A. Hammond, and Peter M. Todd. 2015. “Coevolution of Decision-Making and Social Environments.” *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* 1–16.

Bruch, Elizabeth, and Joffre Swait. 2019. “Choice Set Formation in Residential Mobility and Its Implications for Segregation Dynamics.” *Demography* 56(5):1665–92.

Buckley, Jack, and Mark Schneider. 2003. “Shopping for Schools: How Do Marginal Consumers Gather Information about Schools?” *Policy Studies Journal* 31(2):121–45.

Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson. 2015. “What Parents Want: School Preferences and School Choice.” *The Economic Journal* 125(587):1262–89.

Cahill, Dennis J. 1994. “A Two-Stage Model of the Search Process for Single-Family Houses: A Research Note.” *Environment and Behavior* 26(1):38–48.

Centola, Damon, and Michael Macy. 2007. “Complex Contagions and the Weakness of Long Ties.” *American Journal of Sociology* 113(3):702–34.

Clark, William AV. 1991. “Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model.” *Demography* 28(1):1–19.

Cowan, Nelson. 2010. “The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why?” *Current Directions in Psychological Science* 19(1):51–57.

Delale-O’Connor, Lori. 2018. “Choosers, Obstructed Choosers, and Nonchoosers: A Framework for Defaulting in Schooling Choices.” *Teachers College Record* 120(4):1–36.

DeLuca, Stefanie, Jennifer Darrah-Okike, and Kiara Millay Nerenberg. 2018. “‘I Can Fill in the Gaps’: Social Class Differences in How Parents Make School Choices.” Presented at the Annual Meeting of the American Sociological Association, August 11, Philadelphia, PA.

Denice, Patrick, and Betheny Gross. 2016. “Choice, Preferences, and Constraints: Evidence from Public School Applications in Denver.” *Sociology of Education* 89(4):300–320.

Duong, Deborah Vakas, and Kevin D. Reilly. 1995. “A System of IAC Neural Networks as the Basis for Self-organization in a Sociological Dynamical System Simulation.” *Behavioral Science* 40(4):275–303.

Fack, Gabrielle, and Julien Grenet. 2010. “When Do Better Schools Raise Housing Prices? Evidence from Paris Public and Private Schools.” *Journal of Public Economics* 94(1–2):59–77.

Flache, Andreas, and Rainer Hegselmann. 2001. “Do Irregular Grids Make a Difference? Relaxing the Spatial Regularity Assumption in Cellular Models of Social Dynamics.” *Journal of Artificial Societies and Social Simulation* 4(4).

Fong, Kelley. 2019. “Subject to Evaluation: How Parents Assess and Mobilize Information from Social Networks in School Choice.” Pp. 158–80 in Vol. 34. Wiley Online Library.

Fossett, Mark. 2006. "Ethnic Preferences, Social Distance Dynamics, and Residential Segregation: Theoretical Explorations Using Simulation Analysis∗." *Journal of Mathematical Sociology* 30(3–4):185–273.

Frisch, Deborah, and Robert T. Clemen. 1994. "Beyond Expected Utility: Rethinking Behavioral Decision Research." *Psychological Bulletin* 116(1):46.

Furnham, Adrian, and Hua Chu Boo. 2011. "A Literature Review of the Anchoring Effect." *The Journal of Socio-Economics* 40(1):35–42.

Gasper, Joseph, Stefanie DeLuca, and Angela Estacion. 2010. "Coming and Going: Explaining the Effects of Residential and School Mobility on Adolescent Delinquency." *Social Science Research* 39(3):459–76.

Gibbons, Stephen, Stephen Machin, and Olmo Silva. 2013. "Valuing School Quality Using Boundary Discontinuities." *Journal of Urban Economics* 75:15–28.

Glazerman, Steven, and Dallas Dotter. 2017. "Market Signals: Evidence on the Determinants and Consequences of School Choice from a Citywide Lottery." *Educational Evaluation and Policy Analysis* 39(4):593–619.

Graif, Corina, Andrew S. Gladfelter, and Stephen A. Matthews. 2014. "Urban Poverty and Neighborhood Effects on Crime: Incorporating Spatial and Network Perspectives." *Sociology Compass* 8(9):1140–55.

Green, Donald P., Dara Z. Strolovitch, and Janelle S. Wong. 1998. “Defended Neighborhoods, Integration, and Racially Motivated Crime.” *American Journal of Sociology* 104(2):372–403.

Hanushek, Eric A., and Steven G. Rivkin. 2012. “The Distribution of Teacher Quality and Implications for Policy.” *Annu. Rev. Econ.* 4(1):131–57.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. “Information, School Choice, and Academic Achievement: Evidence from Two Experiments.” *The Quarterly Journal of Economics* 123(4):1373–1414.

Harris, Douglas N., and Matthew Larsen. 2015. "What schools do families want (and why)." New Orleans, LA: Education Research Alliance for New Orleans.

Hertwig, Ralph, and Stefan M. Herzog. 2009. “Fast and Frugal Heuristics: Tools of Social Rationality.” *Social Cognition* 27(5):661–98.

Holland, John Henry. 1995. *Hidden Order: How Adaptation Builds Complexity*.

Holme, Jennifer Jellison. 2002. “Buying Homes, Buying Schools: School Choice and the Social Construction of School Quality.” *Harvard Educational Review* 72(2):177–206.

Johnson, Rucker C. 2011. *Long-Run Impacts of School Desegregation & School Quality on Adult Attainments*. National Bureau of Economic Research.

Kahneman, Daniel, and Amos Tversky. 1979. “On the Interpretation of Intuitive Probability: A Reply to Jonathan Cohen.”

Kahneman, Daniel, and Amos Tversky. 2013. “Prospect Theory: An Analysis of Decision under Risk.” Pp. 99–127 in *Handbook of the fundamentals of financial decision making: Part I*. World Scientific.

Kauko, Tom. 2004. “A Comparative Perspective on Urban Spatial Housing Market Structure: Some More Evidence of Local Sub-Markets Based on a Neural Network Classification of Amsterdam.” *Urban Studies* 41(13):2555–79.

Kerbow, David. 1996. “Patterns of Urban Student Mobility and Local School Reform.” *Journal of Education for Students Placed at Risk (JESPAR)* 1(2):147–69.

Koretz, Daniel M. 2008. *Measuring up: What Educational Testing Really Tells Us*. Boston: Harvard University Press.

Kotok, Stephen, Erica Frankenberg, Kai A. Schafft, Bryan A. Mann, and Edward J. Fuller. 2017. “School Choice, Racial Segregation, and Poverty Concentration: Evidence from Pennsylvania Charter School Transfers.” *Educational Policy* 31(4):415–47.

Krysan, Maria, and Kyle Crowder. 2017. *Cycle of Segregation: Social Processes and Residential Stratification*. Russell Sage Foundation.

Lareau, Annette, and Kimberly Goyette. 2014. *Choosing Homes, Choosing Schools*. Russell Sage Foundation.

Leishman, Chris, Greg Costello, Steven Rowley, and Craig Watkins. 2013. “The Predictive Performance of Multilevel Models of Housing Sub-Markets: A Comparative Analysis.” *Urban Studies* 50(6):1201–20.

Lincove, Jane A., Joshua M. Cowen, and Jason P. Imbrogno. 2018. "What's in your portfolio? How parents rank traditional public, private, and charter schools in post-Katrina New Orleans' citywide system of school choice." *Education Finance and Policy* 13(2): 194-226.

Logan, John R., and Julia Burdick-Will. 2017. "School Segregation and Disparities in Urban, Suburban, and Rural Areas." *The Annals of the American Academy of Political and Social Science* 674(1):199–216.

Logan, John R., Elisabeta Minca, and Sinem Adar. 2012. "The Geography of Inequality: Why Separate Means Unequal in American Public Schools." *Sociology of Education* 85(3):287–301.

Lomi, Alessandro, and Erik Reimer Larsen. 1998. "Density Delay and Organizational Survival: Computational Models and Empirical Comparisons." *Computational & Mathematical Organization Theory* 3(4):219–47.

Macy, Michael W., and Robert Willer. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28(1):143–66.

Manski, Charles F. 1977. "The Structure of Random Utility Models." *Theory and Decision* 8(3):229–54.

March, James G. 1994. *Primer on Decision Making: How Decisions Happen*. Simon and Schuster.

March, James G., and Hernert A. Simon. 1958. "Organizations John Wiley & Sons." *New York*.

Martin, Christopher H., and Peter C. Wainwright. 2013. “Multiple Fitness Peaks on the Adaptive Landscape Drive Adaptive Radiation in the Wild.” *Science* 339(6116):208–11.

Miller, George A. 1956. “The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information.” *Psychological Review* 63(2):81.

Moody, James. 2001. “Race, School Integration, and Friendship Segregation in America.” *American Journal of Sociology* 107(3):679–716.

MSDE, (Maryland State Department of Education). 2015. “Maryland Report Card.”

NCES (National Center for Education Statistics). 2015. “Public Elementary/Secondary School Universe Survey Data 2014-15.” http://nces.ed.gov/ccd.

Pattillo, Mary E., Lori Delale-O’Connor, and Felicia Butts. 2014. “High-Stakes Choosing: Race and Place in Chicago School Reform.” Pp. 237–67 in *Choosing homes, choosing schools*. Russell Sage Foundation.

Phillippo, Kate L., and Briellen Griffin. 2016. “The Social Geography of Choice: Neighborhoods’ Role in Students’ Navigation of School Choice Policy in Chicago.” *The Urban Review* 48(5):668–95.

Posey-Maddox, Linn, Shelley Kimelberg, and Maia Cucchiara. 2014. “Middle-Class Parents and Urban Public Schools: Current Research and Future Directions.” *Sociology Compass* 8(4):446–56.

Reardon, Sean F., and Glenn Firebaugh. 2002. “Measures of Multigroup Segregation.” *Sociological Methodology* 32(1):33–67.

Reardon, Sean F., Elena Tej Grewal, Demetra Kalogrides, and Erica Greenberg. 2012. "Brown Fades: The End of Court-ordered School Desegregation and the Resegregation of American Public Schools." *Journal of Policy Analysis and Management* 31(4):876–904.

Reardon, Sean F., Demetra Kalogrides, and Kenneth Shores. 2019. "The Geography of Racial/Ethnic Test Score Gaps." *American Journal of Sociology* 124(4):1164–1221.

Reardon, Sean F., and Ann Owens. 2014. "60 Years after Brown: Trends and Consequences of School Segregation." *Annual Review of Sociology* 40:199–218.

Reardon, Sean F., John T. Yun, and Tamela McNulty Eitle. 2000. "The Changing Structure of School Segregation: Measurement and Evidence of Multiracial Metropolitan-Area School Segregation, 1989–1995." *Demography* 37(3):351–64.

Renzulli, Linda A., and Lorraine Evans. 2005. "School Choice, Charter Schools, and White Flight." *Social Problems* 52(3):398–418.

Rhodes, Anna, and Siri Warkentien. 2017. "Unwrapping the Suburban 'Package Deal' Race, Class, and School Access." *American Educational Research Journal* 54(1_suppl):168S-189S.

Roda, Allison, and Amy Stuart Wells. 2013. "School Choice Policies and Racial Segregation: Where White Parents' Good Intentions, Anxiety, and Privilege Collide." *American Journal of Education* 119(2):261–93.

Sampson, Robert J. 2008. "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure." *American Journal of Sociology* 114(1):189–231.

Saporito, Salvatore, and Annette Lareau. 1999. “School Selection as a Process: The Multiple Dimensions of Race in Framing Educational Choice.” *Social Problems* 46(3):418–39.

Saporito, Salvatore, and Deenesh Sohoni. 2006. “Coloring Outside the Lines: Racial Segregation in Public Schools and Their Attendance Boundaries.” *Sociology of Education* 79(2):81–105.

Schelling, Thomas C. 1971. “Dynamic Models of Segregation.” *Journal of Mathematical Sociology* 1(2):143–86.

Schneider, Mark, and Jack Buckley. 2002. “What Do Parents Want from Schools? Evidence from the Internet.” *Educational Evaluation and Policy Analysis* 24(2):133–44.

Schwartz, Barry. 2004. “The Paradox of Choice: Why More Is Less.” Ecco New York.

Sen, Ashish, and Tony E. Smith. 2012. *Gravity Models of Spatial Interaction Behavior*. Springer Science & Business Media.

Simon, Herbert A. 1990. “Bounded Rationality.” Pp. 15–18 in *Utility and probability*. Springer.

Sirer, M. Irmak, Spiro Maroulis, Roger Guimera, Uri Wilensky, and Luís A. Nunes Amaral. 2015. “The Currents beneath the ‘Rising Tide’ of School Choice: An Analysis of Student Enrollment Flows in the Chicago Public Schools.” *Journal of Policy Analysis and Management* 34(2):358–77.

Small, Mario Luis. 2017. *Someone to Talk To*. Oxford University Press.

Smrekar, Claire, and Ellen Goldring. 1999. *School Choice in Urban America: Magnet Schools and the Pursuit of Equity. Critical Issues in Educational Leadership Series.* ERIC.

Stein, Marc L. 2015. “Public School Choice and Racial Sorting: An Examination of Charter Schools in Indianapolis.” *American Journal of Education* 121(4):597–627.

Stein, Marc L., Julia Burdick-Will, and Jeffrey Grigg. 2019. “A Choice Too Far? Transit Difficulty and Early High School Transfer in a System of School Choice.” Presented at the Association for Public Policy and Management Fall Conference, Denver.

Stein, Marc L., and Sarah Nagro. 2015. “The Readability and Complexity of District-Provided School-Choice Information.” *Journal of Education for Students Placed at Risk (JESPAR)* 20(3):199–217.

Suttles, Gerald Dale. 1972. *The Social Construction of Communities*. Vol. 111. University of Chicago Press Chicago.

Swait, Joffre, Neil Brigden, and Richard D. Johnson. 2014. “Categories Shape Preferences: A Model of Taste Heterogeneity Arising from Categorization of Alternatives.” *Journal of Choice Modelling* 13:3–23.

Tatian, Peter, and Serena Lei. 2015. “Washington, DC: Our Changing City.” *Urban Institute.*

Tuljapurkar, Shripad, Elizabeth E. Bruch, and Robert D. Mare. 2008. “Neighborhoods and Individual Preferences: A Markovian Model.”

Vaisey, Stephen, and Lauren Valentino. 2018. “Culture and Choice: Toward Integrating Cultural Sociology with the Judgment and Decision-Making Sciences.” *Poetics* 68:131–43.

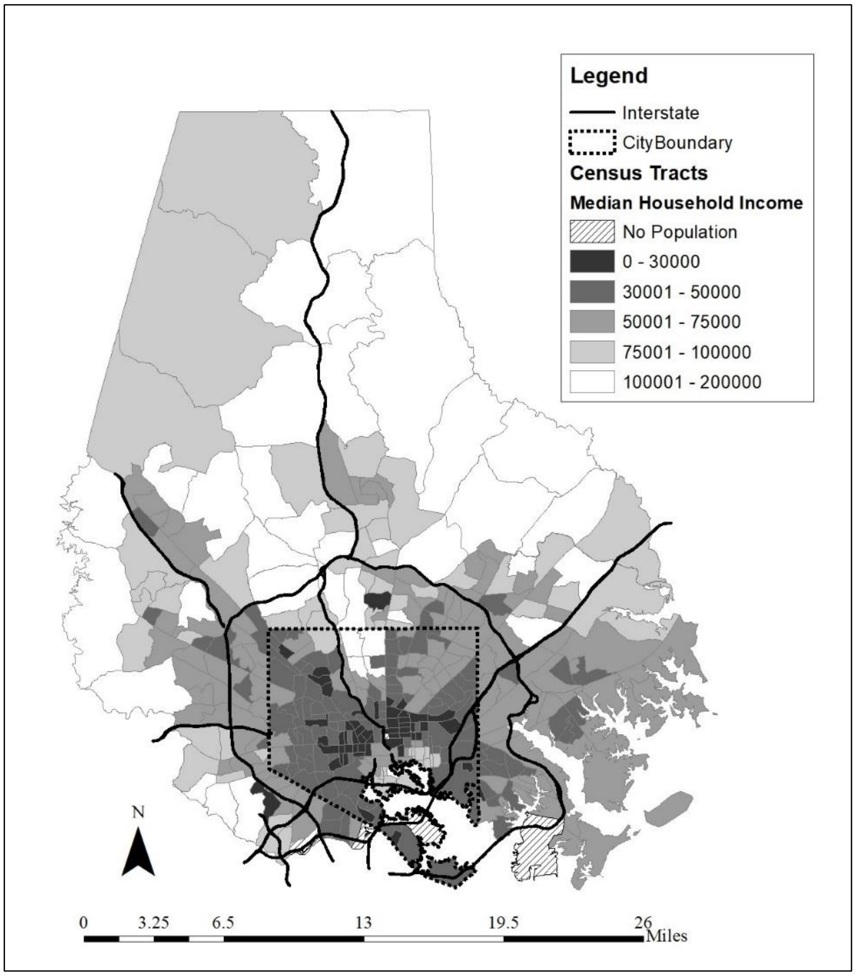
Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson. 2018. "Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities." *Proceedings of the National Academy of Sciences* 115(30):7735–40.

Welsh, Richard O. 2017. "School Hopscotch: A Comprehensive Review of K–12 Student Mobility in the United States." *Review of Educational Research* 87(3):475–511.

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116(2):583–642.

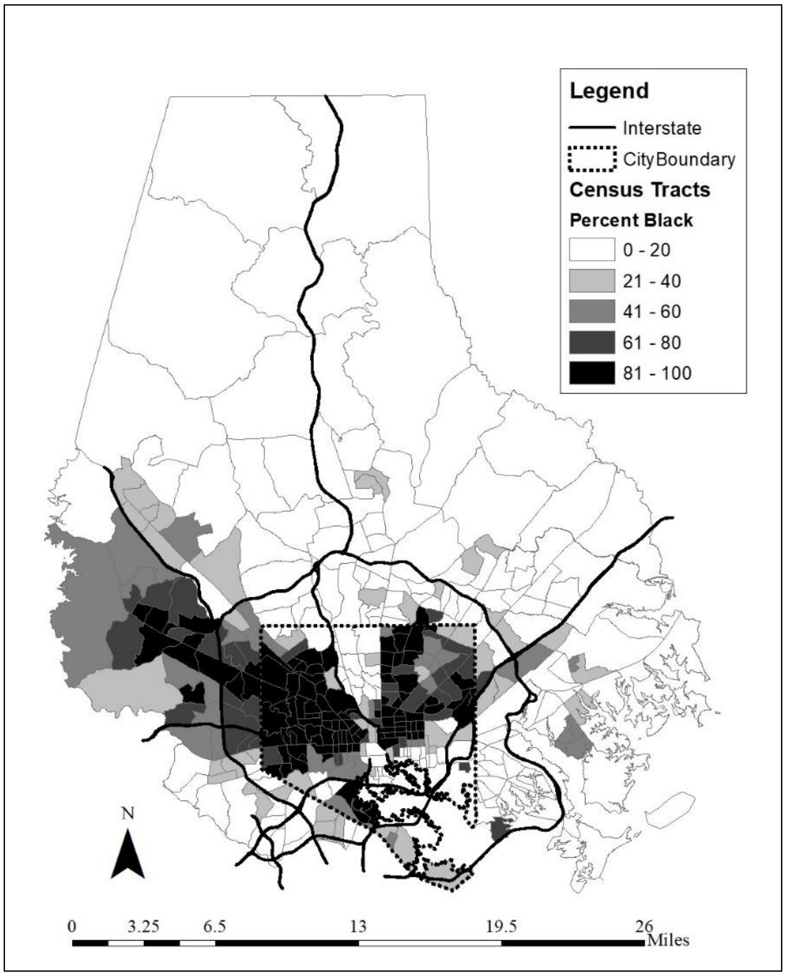
Figures & Tables

Figure 1: Tract Median Household Income



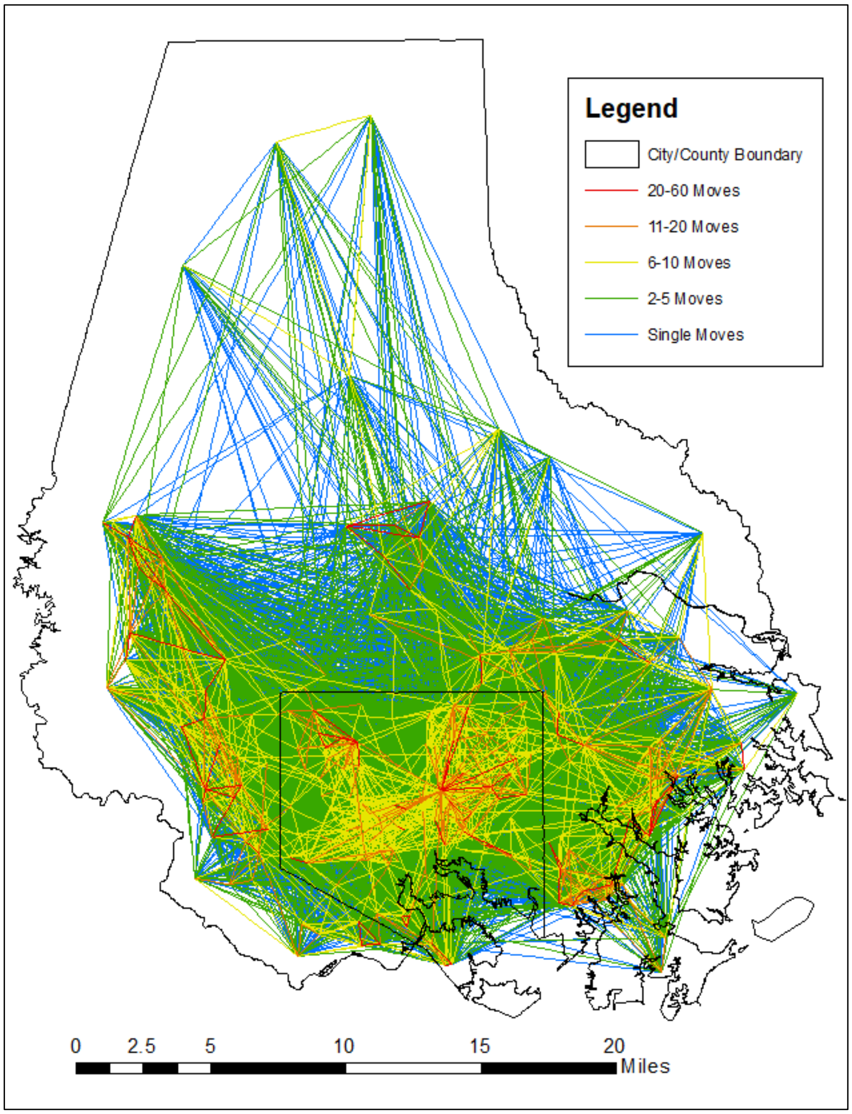
Source: Authors' calculations based on data the 2010-2014 American Community Survey.

Figure 2: Tract Percent Black



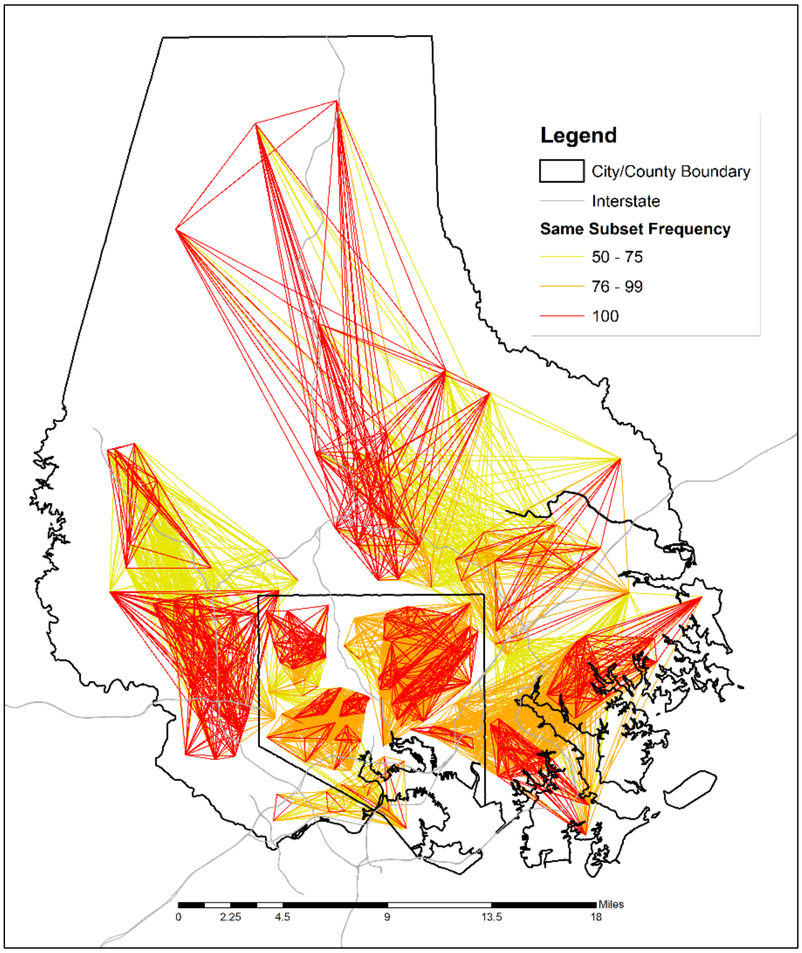
Source: Authors' calculations based on data the 2010-2014 American Community Survey.

Figure 3: Density of Ties between Schools



Source: Authors' calculations based on data from the Baltimore City Public School System and the Baltimore County Public School.

Figure 4: Geographic Distribution of School Mobility Subset



Source: Authors' calculations based on data from the Baltimore City Public School System and the Baltimore County Public School.

Table 1: School Demographics and Achievement in 2014-15 by District

	All	City	County
City	50.8	100.0	0.0
	(50.0)	(0.0)	(0.0)
Black	57.9	78.9	36.1
	(35.1)	(27.0)	(28.5)
White	26.1	9.8	42.9
	(28.6)	(17.0)	(28.3)
Hispanic	8.8	9.1	8.6
	(12.4)	(16.0)	(6.8)
Asian	3.8	1.0	6.8
	(5.5)	(2.0)	(6.4)
Free and Reduced Meals	69.2	87.2	50.6
	(27.3)	(14.8)	(24.7)
Special Education	12.4	13.3	11.4
	(4.0)	(4.3)	(3.6)
English Language Learner	5.9	5.2	6.6
	(8.5)	(10.6)	(5.4)
Attendance	94.9	93.2	96.5
	(2.7)	(2.5)	(1.5)
Proficient - Reading	26.1	14.9	37.5
	(16.3)	(10.0)	(13.3)
Proficient - Math	17.1	11.3	23.0
	(11.9)	(8.1)	(12.2)
Charter	8.7	17.2	0.0
	(28.3)	(37.7)	(0.0)
Number of Students	92361	40802	51559
Number of Schools	236	128	108

Source: Authors' calculations based on data from the Maryland State Department of Education.

Table 2: Characteristics of Ever Movers vs Non-Movers

	Non-Movers	Ever Movers
White	33.2	15.6
Black	49.8	72.1
Hispanic	8.3	6.8
Male	50.9	52.0
School Percent FRM	62.2	76.1
School Proficiency	49.0	49.3
Number of Students	102,725	36,575

Source: Authors' calculations based on data from the Baltimore City Public School System and the Baltimore County Public Schools.
Note: FRM stands for free- and reduced-price meal-eligible students.

Table 3: Comparison of Emergent Clusters and Demographic Similarity

Demographic and Distance Tie	Cluster Tie: Always		Cluster Tie: Never		Total
	No	Yes	No	Yes	
No	25,706	967	6,765	19,908	26,673
Yes	776	281	872	185	1,057
Total	26,482	1,248	7,637	20,093	27,730

Source: Authors' calculations based on data from the Baltimore City Public School System and the Baltimore County Public Schools.

Table 4: Predicted Number of Times Appearing in the Same Emergent Consideration Set

Percent Black Difference	-2.40***	-1.06***
	(0.06)	(0.05)
Percent Hispanic Difference	-1.19***	-0.67***
	(0.14)	(0.15)
Percent FRM Difference	-1.49***	1.45***
	(0.07)	(0.06)
Percent Proficient Difference	-0.14	-0.43***
	(0.12)	(0.10)
Log Distance		-20.75***
		(0.30)
Closest		13.18***
		(2.09)
Different Charter Status		-5.16***
		(0.43)
Different District		-16.20***
		(0.33)
Cross Major Road		-8.16***
		(1.51)
Constant	31.01***	73.29***
	(0.49)	(1.45)
Observations	27,730	27,730
Adjusted R-squared	0.10	0.37

Note: Robust standard errors in parenthesis. *** p<0.001, ** p<0.01, * p<0.05. One unit in each of the difference measures represents 10 percentage points. FRM stands for free- and reduced-price meal-eligible students.
Source: Authors' calculations based on data from the Baltimore City and Baltimore County Public Schools and the Maryland State Department of Education.

Table 5: Theil Indices of Racial, Economic, and Proficiency Segregation Between and Within Subsets

	Race/Ethnicity	Free and Reduced Meal Eligibility	Reading Proficiency
Between Schools	0.35	0.30	0.20
Between Emergent Subsets (average across runs)	0.18 (0.01)	0.16 (0.01)	0.11 (0.01)
Between Emergent Subsets (always same)	0.25	0.23	0.14

Note: Between and within subset are reported as the mean across all 100 runs with the standard deviation across runs in parentheses.
Source: Authors' calculations based on data from the Maryland State Department of Education.

Table 6: Predicted Number of Movers between Sending and Receiving School

	Model 1			Model 2		
	Unweighted	Continuous Weights	Always Weights	Unweighted	Continuous Weights	Always Weights
Percent Black Difference	0.01*	0.02	0.09	0.001	0.003	0.09
	(0.003)	(0.02)	(0.11)	(0.002)	(0.02)	(0.11)
Percent Hispanic Difference	-0.005	0.03	-0.13	0.01	0.05	-0.12
	(0.01)	(0.04)	(0.23)	(0.01)	(0.03)	(0.22)
Percent FRM Difference	-0.01	-0.09**	-0.30*	-0.02***	-0.14***	-0.36**
	(0.01)	(0.03)	(0.13)	(0.01)	(0.03)	(0.12)
Proficiency Difference	0.05***	0.21***	0.40***	0.04***	0.17***	0.33***
	(0.004)	(0.02)	(0.05)	(0.004)	(0.02)	(0.05)
Log Distance				-1.06***	-1.78***	-1.18**
				(0.02)	(0.11)	(0.43)
Closest				5.10***	3.50***	3.71***
				(0.59)	(0.61)	(0.84)
Cross District				-0.38***	-0.18*	
				(0.01)	(0.08)	
Cross Major Road				-1.46***	-1.32***	-1.04*
				(0.18)	(0.27)	(0.48)
Different Charter Status				-0.33***	-0.95***	-1.67***
				(0.02)	(0.13)	(0.40)
Exits (Sending)	0.04***	0.12***	0.20***	0.02***	0.07***	0.16**
	(0.003)	(0.02)	(0.05)	(0.003)	(0.02)	(0.06)
Entries (Receiving)	0.04***	0.14***	0.29***	0.04***	0.12***	0.28***
	(0.005)	(0.03)	(0.07)	(0.005)	(0.03)	(0.08)
Size (Sending)	-0.01	0.37**	1.03*	0.31***	1.00***	1.42***
	(0.03)	(0.14)	(0.43)	(0.03)	(0.14)	(0.41)
Size (Receiving)	-0.02	0.08	-0.11	0.18***	0.49*	0.10
	(0.04)	(0.22)	(0.59)	(0.04)	(0.24)	(0.68)
Constant	-0.66***	-2.87***	-5.41***	2.55***	-0.30	-3.90*
	(0.09)	(0.50)	(1.54)	(0.19)	(0.60)	(1.74)
Observations	55,460	15,274	2,496	55,460	15,274	2,496
R-squared	0.05	0.12	0.16	0.18	0.19	0.19

Source: Authors' calculations based on data from the Baltimore City and Baltimore County Public Schools and the Maryland State Department of Education.
Note: Robust standard errors in parenthesis. *** p<0.001, ** p<0.01, * p<0.05. One unit in each of the difference measures represents 10 percentage points. One unit in school size represents 100 students. FRM stands for free- and reduced-price meal-eligible students.

Appendix A

Limits of Existing Methods

The hard to measure and individually-specific nature of choice sets make them difficult to model on a large scale, which has limited their utility in empirical studies of decision-making and segregation. The few studies that have tried do so use methods that are not well suited to school segregation dynamics. For example, Bruch and Swait (2019) limit consideration sets to only neighborhoods that respondents are likely to be able to afford and that are reasonably close to their current home. Their studies find that affordability and distance constraints lead to racially segregated consideration sets and amplify patterns of residential inequality. School consideration sets are likely similarly limited by distance, but price and affordability are not always as straightforward for public schools as they are for buying and renting homes (Bayer, Ferreira, and McMillan 2007).

Other researchers have gathered data on internet search processes. For example, Buckley and Schneider (2003) examined which boxes a relatively small number of school searchers used to filter their results on a custom-designed website. They found that parents ruled out large numbers of schools with simple search filters and only considered a handful of individual school characteristics in depth. Bruch, Fienberg, and Lee (2016) use similar data to understand how users of online dating websites limit their search criteria. They develop methods to identify deal-breaker screening behavior in online searches and show that the criteria used for evaluation differ during the screening and final selection phases. Unfortunately, no such search data exist on a large scale for the school selection process. Moreover, since research shows that most parents rely heavily on word-of-mouth to gain information about select schools, those who rely only on publicly available websites for school information are likely to be new to the area or otherwise weakly embedded in the local social structure.

One source of information that is sometimes available on a large scale are ranked-choice preferences in a unified enrollment system (Denice and Gross 2016; Harris and Larsen 2015; Glazerman and Dotter 2017; Lincove, Cowen, and Imbrogno 2018). Unfortunately, these data sets are only available for districts that operate with an open enrollment policy, which tend to be urban with large numbers of low achieving schools and high proportions of poor and minority students. The findings are not necessarily applicable to more advantaged, suburban districts that still rely heavily on residential assignment. They also usually only allow for examination of preferences within that particular district. They, therefore, cannot be used to assess the degree that districts themselves serve as a limiting factor for consideration sets. Finally, the schools listed on these application forms only represent explicitly stated preferences at one point in time. Many students do not end up enrolled in schools that were listed on these forms, especially after the period of initial enrollment (Stein, Burdick-Will, and Grigg 2019). This may be because these forms tend to limit the number of schools a family can include or because where families are willing to consider changes over time. Either way, these forms act like interviews. They can certainly tell us a lot about families' stated preferences at a specific point in time, but are less likely to tell us about how actual behavior is structured and constrained.

In contrast, we propose a novel method that can be applied to any large population data or highly-saturated survey. Rather than make assumptions about what families use to limit their consideration sets or rely on active search behavior and stated preferences, we use the observed connections and flow of students between schools to estimate the emergent structure of consideration sets. These structures are based entirely on the actual patterns of flows between schools and are not limited by researchers' assumptions of how families behave or what families say they look for.

Appendix B

Supplemental Tables

Table B1: Comparison of Emergent Clusters and With Different Definitions of Demographic Similarity

	Cluster Tie: Always		Cluster Tie: Never		
	No	Yes	No	Yes	Total
Narrow Definition					
No	26,337	1,159	7,417	20,079	26,337
Yes	145	89	220	14	145
Wide Definition					
No	24,472	756	5,856	19,372	26,673
Yes	2,010	492	1,781	721	1,057
Total	26,482	1,248	7,637	20,093	27,730

Note: In the narrow definition of demographic distance schools are considered demographically similar if they are less than 10 percentage points different in composition and proficiency and less than 2 miles apart. The R-squared predicting total moves with this narrow definition is 0.03. In the wide definition of demographic distance schools are considered demographically similar if they are less than 30 percentage points different in composition and proficiency and less than 4 miles apart. The R-squared predicting total moves with this wide definition is 0.09.
Source: Authors' calculations based on data from the Baltimore City Public School System and the Baltimore County Public Schools.

Table B2: Predicted Log Odds Appearing in the Same Emergent Consideration Set in a Single Run

Percent Black Difference	-0.22***	-0.15***
	(0.01)	(0.01)
Percent Hispanic Difference	-0.04*	0.04*
	(0.01)	(0.02)
Percent FRM Difference	-0.17***	0.19***
	(0.01)	(0.01)
Percent Proficient Difference	-0.01	-0.06***
	(0.01)	(0.01)
Log Distance		-1.72***
		(0.03)
Closest		0.31
		(0.28)
Different Charter Status		-0.72***
		(0.05)
Different District		-2.13***
		(0.06)
Cross Major Road		-0.09
		(0.10)
Constant	-0.53***	1.98***
	(0.03)	(0.10)
Observations	2,773,000	2,773,000
Number of ties	27,730	27,730
Pseudo R-squared	0.10	0.35

Note: The data has been expanded so that each observation represents one of each of the 100 runs for every tie. Standard errors in parenthesis are clustered at the level of the school tie. *** p<0.001, ** p<0.01, * p<0.05. One unit in each of the difference measures represents 10 percentage points. FRM stands for free- and reduced-price meal-eligible students.
Source: Authors’ calculations based on data from the Baltimore City and Baltimore County Public Schools and the Maryland State Department of Education.

Table B3: Marginal Predicted Number of Movers between Sending and Receiving School

	Model 1			Model 2		
	Unweighted	Continuous Weights	Always Weights	Unweighted	Continuous Weights	Always Weights
Percent Black Difference	0.001	-0.009	0.083	-0.005	-0.034	0.060
	(0.004)	(0.025)	(0.095)	(0.003)	(0.020)	(0.082)
Percent Hispanic Difference	-0.014	-0.019	0.048	0.005	-0.020	0.059
	(0.008)	(0.047)	(0.198)	(0.007)	(0.038)	(0.174)
Percent FRM Difference	-0.012*	-0.112**	-0.313**	-0.040***	-0.189***	-0.471***
	(0.006)	(0.039)	(0.140)	(0.005)	(0.032)	(0.125)
Proficiency Difference	0.047***	0.205***	0.395***	0.023***	0.134***	0.291***
	(0.004)	(0.019)	(0.049)	(0.003)	(0.015)	(0.043)
Log Distance				-1.059***	-2.511***	-3.104***
				(0.023)	(0.096)	(0.240)
Closest				-0.123	0.245	1.039*
				(0.067)	(0.207)	(0.443)
Cross District				-0.535***	-1.613***	-2.765***
				(0.025)	(0.138)	(0.420)
Cross Major Road				-0.609***	-1.088***	
				(0.020)	(0.139)	
Different Charter Status				-0.082*	-0.479**	-0.402
				(0.035)	(0.143)	(0.331)
Exits (Sending)	0.041***	0.125***	0.198***	0.019***	0.055***	0.111***
	(0.001)	(0.007)	(0.020)	(0.001)	(0.005)	(0.016)
Entries (Receiving)	0.033***	0.102***	0.193***	0.029***	0.090***	0.190***
	(0.002)	(0.012)	(0.028)	(0.001)	(0.011)	(0.028)
Size (Sending)	-0.016	0.290**	0.695*	0.284***	1.153***	1.791***
	(0.020)	(0.111)	(0.309)	(0.015)	(0.104)	(0.301)
Size (Receiving)	0.052*	0.385**	0.355	0.224***	0.681***	0.460
	(0.021)	(0.111)	(0.279)	(0.014)	(0.104)	(0.277)
Observations	55,460	15,274	2,496	55,460	15,274	2,496
R-squared	0.04	0.06	0.06	0.10	0.13	0.18

Source: Authors' calculations based on data from the Baltimore City and Baltimore County Public Schools and the Maryland State Department of Education.

Note: Marginal predicted number of movers calculated from negative binomial regression results. Robust standard errors in parenthesis. *** $p<0.001$, ** $p<0.01$, * $p<0.05$. One unit in each of the difference measures represents 10 percentage points. One unit in school size represents 100 students. FRM stands for free- and reduced-price meal-eligible students.

Table B4: Predicted Number of Movers between Sending and Receiving School

	Model 1			Model 2		
	Unweighted	Continuous Weights	Always Weights	Unweighted	Continuous Weights	Always Weights
Percent Black Difference	0.007*	0.011	0.062	0.00001	-0.012	0.040
	(0.003)	(0.025)	(0.119)	(0.002)	(0.024)	(0.116)
Percent Hispanic Difference	-0.015	-0.032	-0.318	0.010	0.022	-0.284
	(0.014)	(0.072)	(0.350)	(0.012)	(0.064)	(0.333)
Percent FRM Difference	-0.008	-0.081*	-0.235	-0.022***	-0.107**	-0.266*
	(0.006)	(0.038)	(0.148)	(0.006)	(0.037)	-0.135
Proficiency Difference	0.053***	0.205***	0.383***	0.044***	0.155***	0.310***
	(0.005)	(0.022)	(0.064)	(0.004)	(0.020)	(0.062)
Attendance Rate Difference	-0.004	0.113	0.512	0.023	0.319	0.740
	(0.040)	(0.213)	(0.709)	(0.036)	(0.197)	(0.685)
Percent Special Education Difference	0.030	-0.055	-0.480	-0.020	-0.158	-0.650
	(0.018)	(0.097)	(0.317)	(0.017)	(0.092)	-0.335
Percent ELL Difference	0.018	0.070	0.115	-0.009	-0.003	0.025
	(0.018)	(0.095)	(0.335)	(0.016)	(0.085)	(0.313)
Log Distance				-1.058***	-1.787***	-1.210**
				(0.025)	(0.105)	(0.431)
Closest				5.103***	3.490***	3.718***
				(0.592)	(0.605)	(0.845)
Different Charter Status				-0.335***	-0.961***	-1.715***
				(0.023)	(0.125)	(0.400)
Cross District				-0.380***	-0.192*	
				(0.014)	(0.076)	
Cross Major Road				-1.457***	-1.313***	-1.033*
				(0.178)	(0.272)	(0.475)
Exits (Sending)	0.040***	0.120***	0.194***	0.018***	0.069***	0.154**
	(0.003)	(0.017)	(0.049)	(0.003)	(0.018)	(0.057)
Entries (Receiving)	0.043***	0.145***	0.291***	0.036***	0.123***	0.277***
	(0.005)	(0.026)	(0.072)	(0.005)	(0.026)	(0.078)
Size (Sending)	-0.013	0.384*	1.133*	0.315***	1.034***	1.551***

	(0.026)	(0.149)	(0.451)	(0.027)	(0.140)	(0.415)
Size (Receiving)	-0.021	0.065	-0.206	0.173***	0.464	-0.022
	(0.042)	(0.234)	(0.630)	(0.041)	(0.254)	(0.718)
Constant	-0.659***	-2.868***	-5.363***	2.550***	-0.275	-3.815*
	(0.087)	(0.496)	(1.532)	(0.194)	(0.601)	(1.729)
Observations	55,460	15,274	2,496	55,460	15,274	2,496
R-squared	0.05	0.12	0.16	0.18	0.19	0.19

Source: Authors' calculations based on data from the Baltimore City and Baltimore County Public Schools and the Maryland State Department of Education.
Note: Robust standard errors in parenthesis. *** p<0.001, ** p<0.01, * p<0.05. One unit in each of the difference measures represents 10 percentage points. One unit in school size represents 100 students. FRM stands for free- and reduced-price meal-eligible students.

Endnotes

[i] Ideally we would include all counties in the metropolitan area in our analysis. However, given that our findings suggest that consideration sets are geographically constrained and more likely to fall within rather than across district boundaries we think it is unlikely that including a larger number of districts would dramatically change the overall findings.

[ii] According the National Center for Education Statistics, in 2014 there were 35 charter schools in Baltimore City (18.5 percent of all schools), but none in Baltimore County (NCES 2015). Students must be residents of Baltimore City to attend one of the City charter schools.

[iii] The pattern of flows is very similar between summer and midyear movers. Including midyear moves does not change the substantive findings, but leads to a more disadvantage population of movers.

[iv] Individual student race and gender are available for this project, but not individual test scores or free or reduced-price meals status. We therefore rely on school level characteristics to describe the differences between movers and non-movers.

[v] Stochastic Block Models do similar things, but to our knowledge there are no available statistical packages that can easily produce a probability of group membership for a given school.

[vi] Alternatively, we tried expanding the data so that each of the 100 runs was its own observation. Logistic regressions predicting inclusion in the same set with these expanded data and standard errors clustered at the tie level produce substantively similar results. Margin estimations of predicted probabilities reveal relatively linear relationships between key predictors and probability of inclusion. For the sake of simplicity, we report only the OLS regression results in the main paper, but the logistic regression results can be seen in Appendix Table B2.

[vii] It is possible that the preferences of students who do not move differs from those of movers. Unfortunately, without knowledge of the other schools or neighborhoods that these families consider before enrolling in kindergarten it is impossible to test this directly. Since the focus is on the change across models, rather than the absolute magnitude of the parameters we believe that these analyses still provide important insight into the role of multi-stage decision-making in school segregation dynamics.

[viii] Given the distribution of the outcome, a negative binomial regression better satisfies the normality assumption than a simple OLS regression. However, the coefficient patterns and overall conclusions are very similar in both models (See Appendix Table B3), so we chose to present the OLS regression results to simplify the presentation and interpretation.

[ix] We also ran models that included differences in attendance rates and percentages of special education students and English Language Learners. None of these measures were statistically significant. Nor did including them change the magnitude of the other coefficients or the predictive power of the models. See Appendix Table B4.

[x] Altering the cut offs for these demographic choice sets does not change the findings. Demographics and distance alone are always poor measures of mobility flows. See Appendix Table B1.

[xi] It could be that the coefficients for racial composition are small and not significant because students of different races have different preferences. In other words, white students prefer schools with higher percentages of white students and black student prefer schools with higher proportions of black students. We, therefore, also ran models predicting race specific flows looking at same race preferences rather than percent Black or Hispanic. These models do not fit

the data as well as the aggregate flows with the actual racial composition. Moreover, the coefficients for same race measures are equally small and not statistically significant.